Global Conference on Business & Social Science-2014, GCBSS-2014, 15th & 16th December,
Kuala Lumpur

An examination of online learning effectiveness using data mining

Nurbiha A Shukor[a]*, Zaidatun Tasir[a], Henny Van der Meijden[b]

[a]Faculty of Education, Universiti Teknologi Malaysia, Skudai, 81310, Malaysia
[b]Radboud University, 6525 Nijmegen, The Netherlands

Abstract

Online learning has become increasingly popular due to technology advancement that allows discussion to occur at distance. Most studies report on students' learning achievement as a result of effective online learning while assessment on the learning process is also necessary. It is possible by applying data mining technique where students' online learning experiences can be assessed based on their log files. This study found that students could also perform well by being a silent learner during online learning. However, students have to spend more effort to be a successful silent learner as suggested by the produced predictive model.

1. Introduction

Online learning offers possibility to implement various instructional designs. With technology advancement, online learning comes in handy for collaborative learning implementation because group collaboration can occur at distance and thus favours divergent thinking, and exploration of multiple perspectives as compared to face-to-face collaboration (Swan, 2003). More importantly, online learning keeps track of students' learning activities where assessment on their collaborative learning experiences can be conducted. Macdonald (2003) underlined the importance of assessing online collaborative learning, as it will provide guidelines and ways to improve students' competency. The assessment will

----

* Corresponding author. Tel.: +6017-3745760
  E-mail address: nurbiha@utm.my

help students to display interaction behaviour that contributes to their learning and encourage them to foster participation during collaboration.

Although the advantages of online collaborative learning are well known (Springer et al., 1999; Stacey, 1999), different pedagogical strategies will have different impact on students' learning especially when different learning tasks were integrated within a specific course (Zafeiriou et al., 2001). Most studies on assessing the effectiveness of online collaborative learning emphasize on the product of learning in the form of learning achievement based on the collaborative work (Macdonald, 2003). There are also studies that report on students' perception about online collaborative learning experiences as indicator of successful online collaborative learning (Dewiyanti et al., 2007). These studies gave us information that collaborative learning can results in better knowledge acquisition while some studies reported differently. A more robust investigation has to be carried out to explore the effects of online collaborative learning intervention on students' learning.

Janssen et al (2007) found that investigation on students' participation during learning will enhance the group awareness and promotes coordination and regulation of social activities for collaboration. However, students' participation in online learning is expressed in various forms. In Fung (2004), students' participation in online collaborative learning is defined as students' interaction activity of posting messages and replying messages that relates to tasks while students' activity of 'viewing messages' is not considered as an 'active' participation. Similarly, Lipponen et al. (2003) defined reading messages as only 'lurking' and it is not an indication of participation. However, there are other learning attributes that could serve as useful indicators of students' participation during online collaborative learning.

2. Background of Problem

The growing interest in online collaborative learning has induced several challenges in assessing online collaborative learning experiences. The importance of assessing online collaborative learning experiences are highlighted by Brindley, Blaschke and Walti (2009) where assessment will encourage participation in online collaborative learning that results in better learning outcomes and skills acquisition. Most studies assessed students' online collaboration based on students' learning outcome as a product of collaboration (Macdonald, 2003) while a handful of information can be retrieved from students' log files available in online databases that signify learning processes have occurred (Fung, 2004; Cocea & Weibelzahl, 2007; Hung & Zhang, 2008).

In Cocea and Weibelzahl (2007), students' log files were retrieved from online learning databases to identify students' engagement. They defined 'engagement' based on number of pages that the students read, time spent reading the pages, and time spent on quizzes. Similarly, log files from online learning databases were used to understand students' motivation in learning based on their participation in the provided learning activities such as drills, games and self-test (Ben-zadok et al., 2011). Hung and Zhang (2008) evaluate students' participation in online learning based on students' log files and then identify the important parameters of participation that can predict better learning outcome. These and many more studies evaluate students' online learning experiences based on the available online learning log files.

Data mining technique is the technique that discovers potentially useful information from the bulk amount of data retrieved from online databases (Han, Kamber & Pei, 2011). It is a data analysis technique commonly used in business and management but becomes increasingly popular in educational research. In some studies, data mining technique is used to predict students' dropout cases (Dekker, Pechenizkiy and Vleeshouwers, 2009; Kotsiantis, 2009) and also predict students' performance based on the current available data about the students (Kovačić, 2010; Bhardwaj & Pal, 2012). Consequently, using the similar technique, students' participation in online collaborative learning can also be assessed vigorously other than primary assessment on students' learning outcome. Dringus and Ellis (2005) stated that the indicators of participation in online learning can be based on students' interaction (such as responding to messages), sharing resources and including lurking (spending time to read messages, viewing resources).

Assessing students' online collaborative learning experiences using data mining technique will provide a useful solution for teachers to understand the influence of students' participation during online collaboration on students' learning achievement. Based on the results, teachers will be informed about the important parameters for students' success (Hung and Zhang, 2008). The obtained information will be a very significant measure of the online

collaborative learning experiences and the results will be promising for advising students at the start and early identification of effective and poor practices, in time for remediation.

Accordingly, this study is carried out to answer the following questions:

- What are the students' performances in an online collaborative learning environment?
- How participation in online collaborative learning environment influences students' learning achievement?
- What are the most important predictors of learning achievement based on participation in online collaborative learning environment?

3. Research Methodology

3.1. Participants

Participants in this study included 20 undergraduate students enrolling the course Web-based Multimedia Development. The students are computer-literate and have been using online learning as part of their courses requirement.

3.2. Research procedure

In this study, pre-performance tests were given before online collaborative learning took place. Students were then divided into groups of five students in each group. They were given five collaborative tasks to be solved in the given time frame (Week 1 until Week 10). Finally, students' log files were collected from online databases for data mining analysis and post-performance test was carried out. The summary of the research procedure is shown in Fig. 1.
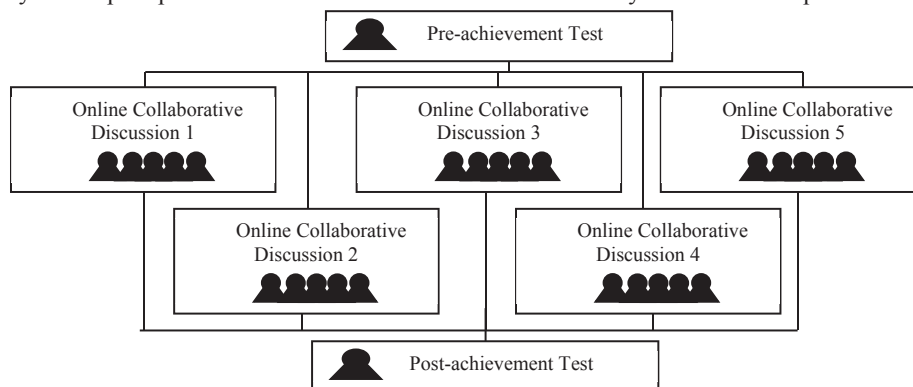

Fig. 1. Research procedure.

3.3. Data analysis

The collected data were analyzed where students' log files were mined using data mining technique.

3.3.1. Analysis of students performance in online collaborative learning

The measurement of students' performance in online collaborative discussions was based on their pre and post-achievement test scores. Their scores were graded into categories based on their total score in achievement tests. This is done to enable observation and monitoring on students' performance. The grading system used is as in Table 1.

Table 1. Students' category according to achievement test scores

Scores	Category
26-30	Outstanding, O
21-25	High-achiever, H-A
16-20	Medium-achiever, M-A
11-15	Low-achiever, L-A
6-10	Weak, W
0-5	Poor, P

Students' scores were categorized twice to clearly examine how much the students improve. Firstly, the scores from pre and post- tests were categorized according to categories in Table 1. Next, the scores were compared and then were categorized into categories as in Table 2.

Table 2. Students' final categorization

Category in Pre-test (i)	Category in Post-test (f)	Final Category
P_i	W_f	+1 Improvement, P1
W_i	$L\text{-}A_f$	+1 Improvement, P1
$L\text{-}A_i$	$M\text{-}A_f$	+1 Improvement, P1
$M\text{-}A_i$	$H\text{-}A_f$	+1 Improvement, P1
$H\text{-}A_i$	O_f	+1 Improvement, P1
P_i	P_f	Static, S
W_i	W_f	Static, S
$L\text{-}A_i$	$L\text{-}A_f$	Static, S
$M\text{-}A_i$	$M\text{-}A_f$	Static, S
$H\text{-}A_i$	$H\text{-}A_f$	Static, S
O_i	O_f	Static, S
P_i	$L\text{-}A_f / M\text{-}A_f / H\text{-}A_f / O_f$	+2/ +3/ +4/ +5 Improvement, P2/P3/P4/P5
W_i	$M\text{-}A_f / H\text{-}A_f / O_f$	+2/ +3/ +4 Improvement
$L\text{-}A_i$	$H\text{-}A_f / O_f$	+2/ +3 Improvement, P2/ P3
$M\text{-}A_i$	O_f	+2 Improvement, P2

3.3.2. Analysis of the influence of participation in online collaborative learning on students' learning achievement

Students' participation in online collaborative learning is based on their activity log files. The participation attributes used in data mining to predict their learning achievement are login frequency (log_), viewing discussions (disc_v), viewing supporting learning materials (res_v), posting comment during collaboration (post_c) and learning achievement (achieve_). Students' log files were analysed using Weka data mining software by applying the C4.5 algorithm to produce a decision tree that shows how online collaborative learning influence students' achievement.

4. Results and Findings

A paired-samples *t*-test was conducted to compare students' performance before and after online collaborative learning intervention. Results in Table 3 show that there was a significant difference in scores for before online collaborative learning (M=11.342, SD=2.592) and after online collaborative learning intervention (M=16.711, SD =3.302); $t(19)=-8.693$, $p=0.000$. Specifically, results suggest that students were able to improve their learning achievement after learning in online collaborative learning took place.

Table 3. Students' category according to achievement test scores

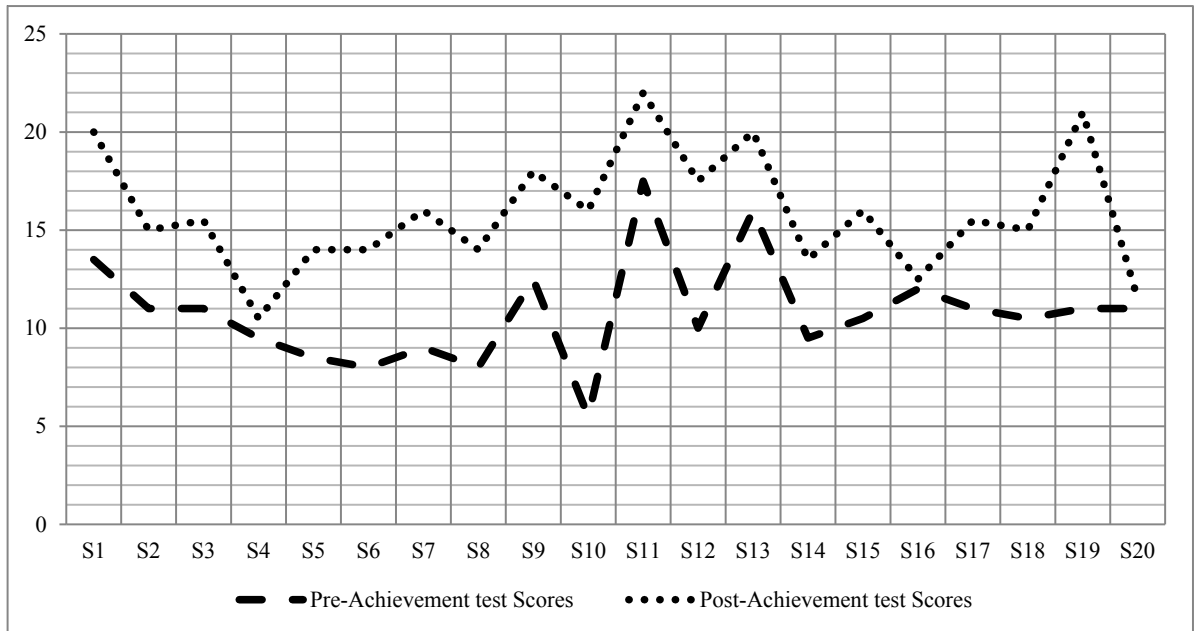| Source	N	M	SD	*t*	*df*	*p*
Pre-achievement test	20	10.775	2.717	-8.693	19	0.000*
Post-achievement test	20	15.875	3.099			

Fig. 2. Students' score distribution in pre and post-achievement tests

Although generally all students show learning improvement upon learning in online collaborative learning, there are also students who did not make progressive improvement during post-test (Student 4, Student 16 and Student 20) as indicated in Fig. 2. These students continued to be 'Medium-achiever, M-A' with least improvement by being at the 'Static' level. Student that shows obvious improvement is Student 19 who scored from being a 'Medium-Achiever' towards being the 'High-Achiever' with P2 Improvement. Student 7 and Student 10 also show similar improvement. Investigation on their online activities log files provides more insight on the contribution factors that contribute to these differences.

Fig. 3 shows the produced decision tree after data mining process took place using the available data in online databases based on the attributes (log_), (disc_v), (res_v), (post_c) and (achieve_) as the predictor. The res_v has the maximum gain ratio, which made it the starting node and the most effective attribute. Other attributes participated in the decision tree were disc_v and log_. The tree shows that whenever a student achieved P2 Improvement, P1 Improvement or Static, the probable cause would be their activities of viewing resources, viewing discussions as well as login into the online collaborative learning space. These attributes are the predictors of students' achievement in online collaborative learning environment.

Based on the decision tree, it can also be concluded that students took several pathways to be at P2 Improvement and P1 Improvement and they only took 1 pathway to be at Static, S. The summary of the pathways is shown in Table 4. It also indicates that although a total of 5 attributes were tested, the important predictors for students' achievement in online collaborative learning are only consist of 'viewing learning resources', 'login to the learning sessions' and 'viewing discussions'.
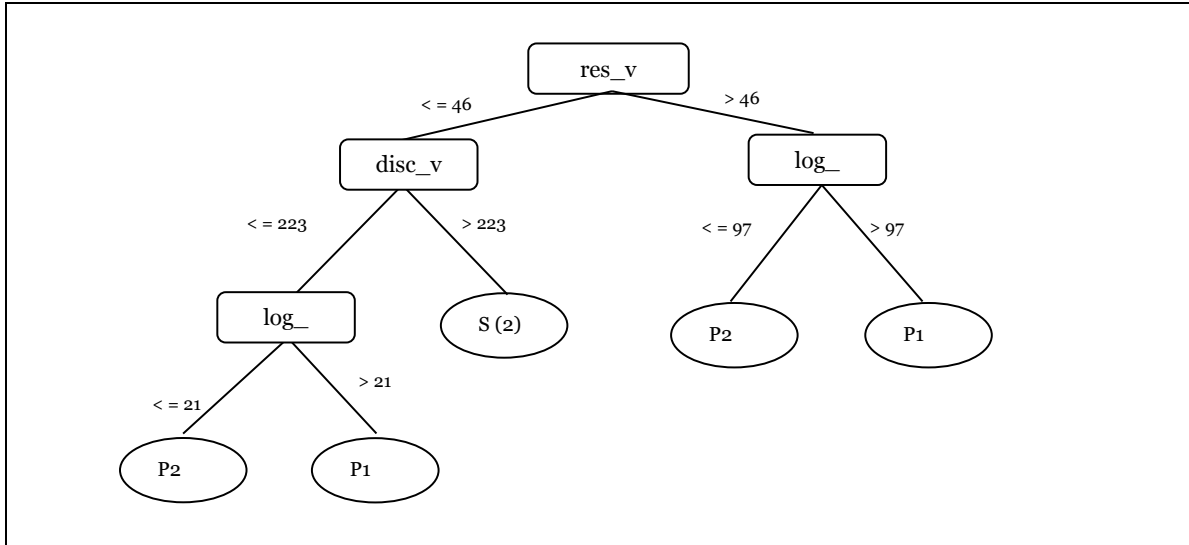
Fig. 3. Summary of Students' Activities during Online Collaborative Learning

Table 4. Summary of students' pathways during online collaborative learning

Performance Types	Pathway
P2 Type I	res_v → log_
P2 Type II	res_v → disc_v → log_
P1 Type I	res_v → log_
P1 Type II	res_v → disc_v → log_
S	res_v → disc_v

5. Discussion

The online collaborative learning environment was found to have significant influence upon students' learning achievements. The majority of the students improved in performance tests by being able to step up a level above the previous stage. In fact, some of the students performed very well in performance tests by being able to move directly from the low-achiever level to the high-achiever level. This finding confirms the hypothesis that students can perform significantly better upon learning in online collaborative learning environment, and is in line with other previous findings (Thomas & MacGregor, 2005; Zhu, 2012). Based on the produced decision tree using data mining technique, this study is able to reveal valuable information about the influence of participation to predict students' achievement. Using data mining, this study shows that learner-learner interaction does not necessarily predicts students' achievement in online collaborative learning environment as frequently reported by other studies (Jung et al., 2002; McDonough, 2004). This is because, this study shows that interacting with content which is by 'viewing resources' and 'viewing discussion' are also predictors of learning achievement. It is in line with findings by Davies and Graff (2005) who found that online participation and interaction do not necessarily translate into higher grades. This study shows that being a silent reader by viewing discussion is equally important to predict learning achievement.

According to Beaudoin (2002), although students are 'less visible' during online learning, it does not necessarily indicates that the students are not learning. In fact, the students are spending effort to read the learning materials, covering the learning-related tasks and including logging on. However, data mining results show that students have to spend a significant amount of time and frequencies on these activities for them to be successful silent learners during online collaborative learning (for example viewing or reading learning resources). In fact, Strachota (2003) found that students prefer learner-content interaction as the priority source of satisfaction in online learning

experiences. These students have to be given scaffolding and encouragement to promote participation of 'learner-learner interaction' during online collaborative learning as Bernard et al., (2004) explain that students have to have the desire for interaction in online learning.

Garrison and Cleveland-Innes (2010) stated that the design or the structure of the online learning as well as the leadership also play important role in online learning success. In this study, the design of collaborative learning environment facilitates students to carry out collaboration, as they have to solve the given problems in the given time frame with their group members. Students being positioned in different learning groups are the sources of different social grounding that results in different level of knowledge acquisition. Different individual learning styles may hinder or foster group collaboration that can both favours better learning outcome or alternatively (Brindley, Blaschke, & Walti, 2009). Every student gained various social supports within the group they work in which in turn affects their learning performance. Weinberger et al. (2005) stipulated that scripting the students' role in collaborative learning could promote participation as well as encourage knowledge construction. By receiving the necessary support that the students need, students can then focussed towards participating in discussions and thus being able to generate more ideas as a result of collaboration.

6. Conclusion

This paper has concentrated on assessing the learning effectiveness of online collaborative learning environment based on students' log files as well as student learning achievement test scores as the primary indicator. Using data mining technique this study shows the predominant attributes that predict students' learning outcomes. Interaction plays an important role in collaborative learning however this study reveals that being a silent learner can equally benefit online learners. Activities such as reading the provided learning resources and reading the discussion threads could also aid students' learning. However, this study is limited to the available data in online databases while factors such as students' positioning in the collaborative group and the structure of the collaborative tasks could also contribute to students' learning achievement. Nevertheless, the proposed predictors generated using data mining technique can serve as a valuable information resource to advice students on learning effectively in an online collaborative learning environment.

Acknowledgements

References

Beaudoin, M. F. (2002). Learning or lurking?: Tracking the "invisible" online student. *The Internet and Higher Education*, *5*(2), 147-155.

Ben-Zadok, G., Leiba, M., & Nachmias, R. (2011). Drills, Games or Tests? Evaluating Students' Motivation in Different Online Learning Activities, Using Log File Analysis. *Interdisciplinary Journal of E-Learning and Learning Objects*, *7*(1), 235-248.

Bhardwaj, B. K., & Pal, S. (2012). Data Mining: A prediction for performance improvement using classification. *International Journal of Computer Science and Information Security*, *9*(4), 136-140.

Brindley, J., Blaschke, L., & Walti, C. (2009). Creating Effective Collaborative Learning Groups in an Online Environment. *The International Review Of Research In Open And Distance Learning*, *10*(3). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/675/1271

Cocea, M., & Weibelzahl, S. (2007). Eliciting motivation knowledge from log files towards motivation diagnosis for Adaptive Systems. In *User Modeling 2007* (pp. 197-206). Springer Berlin Heidelberg.

Dekker, G. W., Pechenizkiy, M., & Vleeshouwers, J. M. (2009). Predicting Students Drop Out: A Case Study. *Proceedings of the 2nd International Conference on Educational Data Mining* (EDM'09), 41-50.

Dewiyanti, S., Brand-Gruwel, S., Jochems, W., & Broers, N. J. (2007). Students' experiences with collaborative learning in asynchronous computer-supported collaborative learning environments. *Computers in Human Behavior*, *23*(1), 496-514.

Dringus, L. P., & Ellis, T. (2005). Using data mining as a strategy for assessing asynchronous discussion forums. *Computers & Education*, *45*(1), 141-160.

Fung, Y. Y. (2004). Collaborative online learning: Interaction patterns and limiting factors. Open Learning, *The Journal of Open, Distance and e-Learning*, *19*(2), 135-149.

Garrison, D. R., & Cleveland-Innes, M. F. (2010). Foundations of distance education. In M. F. Cleveland-Innes & D. R. Garrison (Eds.), *An introduction to distance education: Understanding teaching and learning in a new era* (pp. 13-25). London: Routledge.

Han, J. Kamber, M. and Pei. J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.), USA: Morgan Kaufmann.

Janssen, J., Erkens, G., Kanselaar, G., & Jaspers, J. (2007). Visualization of participation: Does it contribute to successful computer-supported collaborative learning?. *Computers & Education*, *49*(4), 1037-1065.

Jung, I., Choi, S., Lim, C., & Leem, J. (2002). Effects of different types of interaction on learning achievement, satisfaction and participation in web-based instruction. *Innovations in Education and Teaching International*, *39*(2), 153-162.

Kotsiantis, S. (2009). Educational data mining: a case study for predicting dropout-prone students. *International Journal of Knowledge Engineering and Soft Data Paradigms*, 1(2), 101-111.

Lipponen, L., Rahikainen, M., Lallimo, J., & Hakkarainen, K. (2003). Patterns of participation and discourse in elementary students' computer-supported collaborative learning. *Learning and instruction*, *13*(5), 487-509.

Macdonald, J. (2003). Assessing online collaborative learning: process and product. *Computers & Education*, *40*(4), 377-391.

McDonough, K. (2004). Learner-learner interaction during pair and small group activities in a Thai EFL context. *System*, *32*(2), 207-224.

Minaei-Bidgoli, B., Kashy, D. A., Kortmeyer, G., & Punch, W. F. (2003, November). Predicting student performance: an application of data mining methods with an educational web-based system. In *Frontiers in Education, 2003*. FIE 2003 33rd Annual, 1, T2A-13.

Springer, L., Stanne, M.E., & Donovan, S.S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, 69(1), 21-51.

Stacey, E. (1999). Collaborative learning in anonline environment. *Journal of Distance Education*, *14*(2). Retrieved from: http://web.mit.edu/acs/faq/Online-collaboration/collab-learning_files/stacey.htm

Strachota, E. M. (2003). Student satisfaction in online courses: An analysis of the impact of learner-content, learner-instructor, learner-learner and learner-teacher interaction. *Dissertation Abstracts International*, *64*(08), 2746.

Swan, K. (2003). Learning effectiveness online: What the research tells us. *Elements of quality online education, practice and direction*, 4, 13-47.

Thomas, W. R., & MacGregor, S. K. (2005). Online project-based learning: How collaborative strategies and problem solving processes impact performance. *Journal of Interactive Learning Research*, *16*(1), 83-107.

Weinberger, A. Ertl,B. Fischer, F. & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science*, 33(1), pp.1-30.

Yvonne Y.H. Fung (2004) Collaborative online learning: interaction patterns and limiting factors, *Open Learning: The Journal of Open, Distance and e-Learning*, 19(2), 135-149.

Zafeiriou, G., Nunes, J. M. B., & Ford, N. (2001). Using students' perceptions of participation in collaborative learning activities in the design of online learning environments. *Education for Information*, *19*(2), 83-106.

Zhu, C. (2012). Student Satisfaction, Performance, and Knowledge Construction in Online Collaborative Learning. *Educational Technology & Society*, 15 (1), 127–136.

Zlatko J. Kovačić, Z. J. (2010). Early Prediction of Student Success: Mining Students Enrolment Data. Paper presented at *Informing Science & IT Education Conference* (InSITE) 2010, Cassino, Italy.