Relations between Student Online Learning Behavior and Academic Achievement in Higher Education: A Learning Analytics Approach

Il-Hyun Jo[1], Taeho Yu[2], Hyeyun Lee[1,*] and Yeonjoo Kim[1],

[1] Department of Educational Technology, Ewha Womans University, Seoul
ijo@ewha.ac.kr, {hyeyun521,aoao1992}@naver.com
[2] Purdue University, IN, USA
yu134@purdue.edu

Abstract. The purpose of this study is to suggest more meaningful components for learning analytics in order to help learners to improve their learning achievement continuously through an educational technology approach. 41 undergraduate students in a women's university in South Korea participated in this study. The seven-predictor model was able to account for 99.3% of the variance in the final grade, $F_{(8, 32)} = 547.424$, $p < .001$, $R^2 = .993$. Total login frequency in LMS, (ir)regularity of learning interval in LMS, and total assignments and assessment composites had a significant ($p < .05$) correlation with final grades. However, total studying time in LMS (ß=.038, t=.868, p > .05), interactions with content (ß=-.004, t=-.240, p > .05), interactions with peers (ß=.015, t=.766, p > .05), and interactions with instructor (ß=.009, t=.354, p > .05) did not predict final grades. The results provide a rationale for the treatment for student time management effort.

Keywords: Learning analytics, Educational technology, Higher education, E-learning.

1 Introduction

Learning analytics have received significant attention from educators and researchers in higher education since its inception [6, 14, 33]. The main concept of learning analytics is quite attractive in that by means of this concept instructors can predict their students' learning outcomes in advance through use of big data mining technology. To produce a fast and precise prediction, the pioneers of learning analytics, such as Baylor University, the University of Alabama, Northern Arizona University, and Purdue University, have considered diverse exogenous variables such as SAT scores, cumulative GPA, and high school GPA as a significant component of their learning analytics models [6]. In fact, it has

G. Chen et al. (eds.), *Emerging Issues in Smart Learning*,
Lecture Notes in Educational Technology, DOI 10.1007/978-3-662-44188-6_38

been verified that these variables are significant factors when predicting students' learning outcomes from the previous research. However, these variables cannot be improved by students' current efforts because these are past-oriented variables. For this reason, it is important to provide precautionary interventions with present-oriented controllable variables, which can be improved by each learner's effort, for individual learners to support their learning process [19].

In the process of integrating education and technology, a new learning environment has been created and a variety of controllable variables can be collected. For instance, students login into a Learning Management System (LMS) to take online courses or to download course materials. Whenever students utilize the Internet, computers, or LMS, many log files are recorded [4, 15]. We can understand the current status of students' learning and even predict their possible learning achievements in a course by analyzing those log data which they leave within the database. In other words, those log data are the core source to generate controllable variables such as: a) total studying time in LMS, b) total login frequency in LMS, c) (ir)regularity of learning interval in LMS, d) interactions with content, instructor, and peers, e) assignments and assessment composite, and f) discussion composite.

Since the concept of learning analytics was derived from business analytics, prediction of learning outcomes became a major area of learning analytics [14, 33]. However, the main purposes should be different between learning analytics and business analytics. Learning analytics should pay greater attention to students and the improvement of their processes toward learning achievement, whereas the main focus of business analytics is to maximize a profit through the prediction of customer's behaviors and patterns [14]. Therefore, this study determines which controllable components need to be included in learning analytics in order to help learners to continuously improve their learning achievements through an educational technology approach.

2 Review of the Literature

2.1 Learning Analytics

Learning analytics has been defined by different researchers since its emergence in education. Elias [14] defines learning analytics as a process of data gathering, information processing, knowledge application, and sharing for the improvement of teaching and learning. Johnson et al. [15] defines learning analytics as "the interpretation of a wide range of data produced by and gathered on behalf of students in order to assess academic progress, predict future performance, and spot potential issues" (p. 28). In addition, Brown [4] describes it as collecting and analyzing "the usage data associated with student learning" (p. 1) to provide actionable interventions for students by observing and understanding learning behaviors. Learning analytics also refers to a student success management system with warranted interventions that operates by collecting and analyzing data from the Learning Management System (LMS) and Student Information System (SIS) [2]. In addition, learning analytics refers to "the use of predictive modeling and other advanced analytic techniques to help target instructional, curricular and support resources to support the achievement of specific learning goals (p. 2)" [3].

There are other attempts to define learning analytics through the educational technology approach. Jo, Kang, Yoon, and Kang [17] define learning analytics as a "systematic understanding of each learner's educational needs and prepared customized instructional strategy and contents by collecting, analyzing, and systematizing learner's data especially from LMS" (p. 3). Jo and J. Kim [18] have pointed out that learning analytics is an emerging field that applies the prediction model identified in educational systems. Furthermore, Jo [16] introduces the Learning Analytics for Prediction and Action (LAPA) model as shown in Figure 1. In his paper, he indicates that it is possible to provide a prompt and personalized educational opportunity to both the student and instructor in accordance with their level and needs though learning analytics with an educational technology approach.

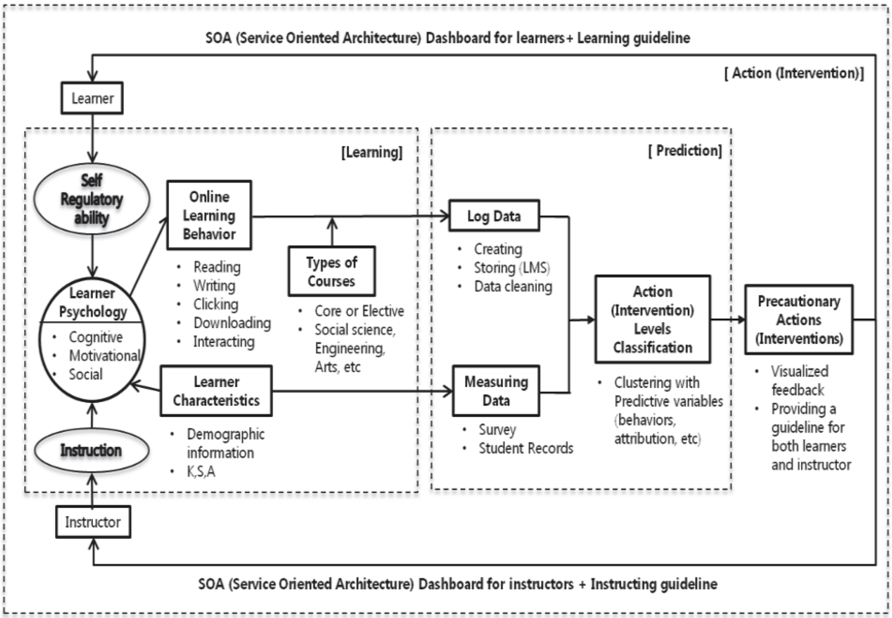

Fig. 1. LAPA (Learning Analytics for Prediction & Action) Model (Jo, 2012).

According to Jo [16], LAPA consists of three segments, identified as learning, prediction, and action (intervention). The first segment presents the learning process with six specific components, i.e. the learner's self-regulatory ability, learner psychology, instruction, online learning behavior, learner characteristics, and types of courses. In the second segment, predicting student's learning achievements and classifying action (intervention) levels are implemented by analyzing log data and measuring data. Finally, precautionary actions (interventions) are provided through the Service Oriented Architecture (SOA) dashboard and guideline for both learners and instructors.

Although learning analytics has various definitions created by different researchers, there is some common ground in these definitions, such as big-data mining, predicting future performance, providing interventions, and improving teaching and learning. Thus, in this study, learning analytics refers to the process of predicting academic performance and providing meaningful interventions with educational big-data mining for students to improve their learning achievement continuously.

2.2 Controllable Variables for Learning Analytics with an Educational Technology Approach

Non-Controllable Variables for Learning Analytics. Numerous non-controllable variables have been determined to be influential factors in terms of learning achievement. Much research has indicated the existence of highly positive correlations between high school GPA and SAT scores and learning outcomes in both face-to-face and online learning environments [1, 5, 7, 22, 25, 31]. Individual student characteristics, such as age, residency, gender, or race, are also determined to be significant factors that are positively related to academic achievement [1, 5, 7]. Other research has investigated the effect of socio-economic variables on students' academic outcomes. Allen [1] argued that the parents' financial and affective support impacts students' learning achievements. Campbell [5] found that a negative relationship existed between the amount of aid and learning outcomes. However, these variables are not changeable despite the learners' efforts and educators' interventions.

Controllable Variables for Learning Analytics. Based on the literature review, this study pays greater attention to the variables that can be improved in an educational technology approach, which are as follows: a) total studying time in LMS, b) total login frequency in LMS, c) (ir)regularity in the learning interval in LMS, d) interactions with instructor, e) interactions with peers, f) interactions with content, g) assignments and assessment composite, and h) discussion composite.

For total studying time in LMS, Rau and Durand [29] and Thurmond el al. [32] determined that total studying time was a significant predictor for GPA. In both studies, the total amount of time spent between login and logout in LMS was considered as the total studying time in LMS. In addition, Piccoli, Ahmad, & Ives [28] argued that when learners login frequently into LMS, they become more satisfied with online learning. In addition, Kang, Kim, and Park [23] confirmed that total login frequency in LMS is positively related to learning performance and attendance rate. To generate the variable of total login frequency in LMS, these researchers totaled the amount of each student's login time. With respect to the (ir)regularity of learning interval in LMS, Jung, Jo, and Lim [20] introduced the (ir)regularity of learning interval in LMS as a significant factor that is positively related to a distance learner's learning outcomes with online courses. In their study, they used the standard deviation of average login time into the LMS to calculate the (ir)regularity of learning interval. Moore [26] also confirmed that the (ir)regularity of learning interval positively correlated with learning achievement even in the traditional face-to-face class setting.

In addition, Swan [30] found positive relationships existent between interactions with content, instructor, and peers and student satisfaction and perceived learning. Jung et al. [21] also insisted that academic, collaborative, and social interaction have an effect on learning satisfaction, participation, and attitude towards online learning. The positive correlations between attendance rate and learning outcomes have been supported by a number of studies [9, 10, 12, 13, 26, 27]. These studies verified that attendance rate has a significant effect on learning outcomes in both an online learning environment and the traditional face-to-face class setting. Last, but not least, assignments, assessment, and discussion composites are determined to be important components in the prediction of learners' academic achievements [5, 7, 32]. Since final grades tend to consist of a variety of assignments, assessment, and discussion composites, it is natural that these variables are considered as a significant factor with learning outcomes.

In sum, this study suggests eight controllable variables for learning analytics, as shown in Table 1. These controllable variables are more meaningful in terms of educational technology because these are actionable and changeable based on a learner's effort.

Table 1. Eight suggested controllable variables for learning analytics on the ground of the educational technology approach

Number	Suggested variables	Relations with learning achievement
1	Total studying time in LMS	Positive
2	Total login frequency in LMS	Positive
3	(Ir)regularity of learning interval in LMS	Positive
4	Interactions with instructor	Positive
5	Interactions with peers	Positive
6	Interactions with content	Positive
7	Assignments and assessment composite	Positive
8	Discussion composite	Positive

3 Research Questions

The purpose of this study is to suggest more meaningful components for learning analytics by which to assist learners improve their learning achievement continuously within an educational technology approach. The specific research questions that will be addressed in the study are:

1. What are the correlations among the eight suggested independent variables and learners' academic achievements?

2. Do the eight suggested independent variables (IV1: Total studying time in LMS, IV2: Total login frequency in LMS, IV3: (Ir)regularity of learning interval in LMS, IV4: Interactions with instructor, IV5: Interactions with peers, IV6: Interactions with content, IV7: Assignments and assessment composite, and IV8: Discussion composite) predict learners' academic achievements?

4 Methods

4.1 Research Context

The participants in this study were 41 undergraduate students who were participants in a face-to-face course entitled 'Organizing Behavior and Leadership' respectively. This course had the following features: a) it was a three credit core course for undergraduate students offered by the department of Science of Public Administration, b) the instructor taught the course during the spring semester 2013 for 16 weeks, c) 20% of the final grade was assigned for online discussion participation in the Learning Management System (LMS), and d) the students used LMS to download course materials, including the syllabus or assigned readings. All of these participants are female students since this is a women's university.

4.2 Data Collection

Seven independent variables among the eight suggested controllable variables. The entirety of the web-log data was collected by means of the Moodle-based Learning Management System (LMS), and the independent variables for this study, as shown in Table 2, were computed by an automatic data collection module embedded in the LMS. First, total login frequency in LMS was calculated by collecting the amount of each student's login time into the LMS. Second, total studying time in LMS was computed by calculating the total amount of time spent between login and logout. Third, the (ir)regularity of learning interval in LMS was computed by calculating the standard deviation of average login time into the LMS, which calculated through use of the same method that Kim [24] executed in her research. Fourth, interactions with content were computed by adding up the number of downloaded course materials. Fifth and sixth, interactions with the instructor and peers were calculated separately by counting the total number of each student's postings in response to the instructor and peers. Seventh, total assignments and assessments composite were computed by adding up all of the scores for assignments and assessment in the course, excluding the attendance rate. Last, to avoid using duplicated variables, discussion composites were also removed from the independent variables because the level of online discussion participation was computed by counting the total number of students' response postings to the instructor and peers.

Table 2. Data collecting methods for each independent variable

Number	Suggested independent variables	Data collecting methods
1	Total login frequency in LMS	Adding up the number of individual student's login time into the LMS
2	Total studying time in LMS	Calculating the total amount of time spent between login and logout
3	(Ir)regularity of learning interval in LMS	Calculating the standard deviation of average login time into the LMS
4	Interactions with content	Adding up the numbers of course materials downloaded
5	Interactions with peers	Counting the total number of student's postings when responding to peers
6	Interactions with instructor	Counting the total number of student's postings when responding to instructor
7	Total assignments and assessment composite	Adding up all scores for assignments and assessment in the course

Dependent variable. Final grades were collected as a dependent variable in this study. A data matching process between independent variables and final grades was executed automatically in the database system. The student's final grade was considered to be the same as that learners' academic achievement because it was synthetically computed from several course components, such as midterm and final exam scores, attendance score, case study report score, and participation score in online discussions.

4.3 Data Analysis

A multiple linear regression analysis was conducted by using Statistical Package for the Social Sciences (SPSS, version 21). Modeling the relations between the explanatory variables and response variables is the main purpose of conducting multiple linear regression [8]. In other words, multiple linear regression analysis is an appropriate statistical method to determine which factors influence changes to the dependent variable [11]. For this reason, this study implemented a multiple linear regression analysis with the seven independent variables, as described above, and with final grades as a dependent variable.

5 Results

5.1 Descriptive Statistics

Table 3 illustrates the descriptive statistics (i.e. the minimums, maximums, means, and standard deviations) for the six independent variables and the dependent variable in this study. It reveals that participating students logged into the LMS 252 times and studied for 34 hours in LMS on average in a semester. There were huge variations in the total login frequency in LMS (M=252.93, SD=110.99, Min=70, Max=604) and total studying time in LMS (M=123,776, SD=70,218, Min=33,389, Max=373.938) among students. The minimum of (ir)regularity of learning interval in LMS was 5.82, whereas the maximum was 72.84. The average number of downloads and the number of postings responding to peers was 36.41 and 2.51 respectively. The students' final grades were distributed between 43.97 and 89.37, and the mean was 81.95.

Table 3. Descriptive statistics of seven independent variables and a dependent variable

	Minimum	Maximum	Mean	S.D.
IV1: Total login frequency in LMS	70	604	252.93	110.99
IV2: Total studying time in LMS	33,389	373,938	123,776	70,218
IV3: (Ir)regularity of learning interval in LMS	5.82	72.84	18.27	10.53
IV4: Interactions with content	5	60	36.41	11.73
IV5: Interactions with peers	0	15	2.51	2.87
IV:6 Interactions with instructor	0	64	14.90	13.98
IV7: Total assignments and assessment composite	36.97	79.37	66.57	10.45
DV: Final grade	43.47	89.37	75.52	11.83

5.2 Multiple Linear Regression Analysis

A multiple linear regression analysis was conducted to develop a model for predicting students' academic achievement based on their total studying time in LMS, total login frequency in LMS, (ir)regularity of learning interval in LMS, interactions with instructor, interactions with peers, interactions with content, and total assignments and assessment composite. Results of this multiple linear regression analysis are illustrated in Table 4. The seven-predictor model provided justification for 99.3% of the variance in the final grade, $F(8, 32) = 547.424$, $p < .01$, $R2 = .993$. Total login frequency in LMS, (ir)regularity of learning interval in LMS, and total assignments and assessment composite had a significant ($p < .05$) correlation with final grades. However, total studying time in LMS (ß=.038, t=.868, $p > .05$), interactions with content (ß=-.004, t=-.240, $p > .05$), interactions with peers (ß=.015, t=.766, $p > .05$), and interactions with instructor (ß=.009, t=.354, $p > .05$) were not essential in the final grade prediction.

Table 4. Results of multiple linear regression analysis

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
(Constant)	11.741	1.787		6.572	.000
IV1	-.007	.002	-.064	-2.791	.009
IV2	.000	.000	.038	.868	.392
IV3	-.156	.025	-.139	-6.335	.000
IV4	-.004	.016	-.004	-.240	.812
IV5	.060	.079	.015	.766	.449
IV6	-.007	.021	.009	.354	.726
IV7	1.034	.023	.955	44.487	.000
R^2 (adj. R^2) =.335(.283), F=6.457, p = .000					

a. Dependent Variable: Final grade

6 Discussion

Based on the multiple linear regression analysis, total login frequency in LMS, (ir)regularity of learning interval in LMS, and total assignments and assessment composite were determined to be significant factors for students' academic achievement in an online learning environment. As the course selected for this study is a face-to-face course that used the Learning Management System for limited purposes, such as participating in online discussions or downloading course materials, these results present a positive potential for the variables to predict the students' academic achievements consistently. Moreover, the significant negative correlation between (ir)regularity of learning interval in LMS and final grade explains that students who study more regularly are likely to achieve higher learning outcomes. Since (ir)regularity of learning interval in LMS is computed by calculating the standard deviation of average login time into LMS, then a lower (ir)regularity of learning intervals in LMS indicates an increased regularity in the study pattern.

The findings of this study also confirm that total login frequency in LMS and tendencies toward regularly studying have a greater significant effect on students' academic achievement than does the total studying time. In addition, there is an interesting finding, which is that interactions with instructor, peers, and content have no significant correlation with the students' final grades. Student-instructor interaction has been empirically evidenced and theoretically supported by previous studies. However, since it was a student-centered discussion, and thus there were significantly low numbers of interaction researched between instructor and students, statistical analysis has failed to detect the potential effect in this study. Finally, the assignments and assessments, such as the mid-term and final exam, writing essays, and online discussion, compose the main components of final grades in this course. Thus, the highly significant relationships between the assignments and assessment composites and the students' final grades were natural.

7 Conclusion

The purpose of this study is to suggest more meaningful components for learning analytics to help learners improve their learning achievement continuously in terms of an educational technology approach. The regression model with only controllable variables that can be affected by learners' efforts was able to account for 99.3% of the variance in students' academic achievement. The main focus of learning analytics tends to focus on the prediction of the future learning outcome by adding geographical, demographical, or characteristic factors, such as high school GPA, SAT score, age, or residency. However, these factors are not controllable because they were fixed in the past and given to the instructional setting. For this reason, this study tested seven controllable variables for our learning analytics model and confirmed that three of them were significantly correlated with the learning final grade. Moreover, these three variables not only predict learning outcomes significantly but also can be improved if learners put more effort into the educational process. The advantage of learning analytics using big-data mining is to predict students' future performance. Yet since the subject is students and not financial profit, educators should pay more attention to improving the process of learners' achievement rather than predicting achievement itself. However, this study was merely conducted with a single face-to-face course within a women's university in South Korea. In addition, a discussion composite was not added as an independent variable for this study because of the research context. Thus, more research should be implemented using various course subjects, different learning environments, and diverse participants with different school settings, ages, sex, nationalities, and level of student-instructor interactions.

References

[1] Allen, D. 1999. Desire to finish college: an empirical link between motivation and persistence. *Research in Higher Education, 40*(4), 461-485.

[2] Arnold, K. E. 2010. Signals: Applying Academic Analytics. *EDUCAUSE Quarterly, 33*(1).

[3] Bach, C. 2010. *Learning Analytics: Targeting Instruction, Curricula and Student Support*. Office of the Provost, Drexel University.

[4] Brown, M. 2011. Learning Analytics: The Coming Third Wave. *EDUCAUSE Learning Initiative Brief*, Retrieved from http://www.educause.edu/Resources/LearningAnalyticsTheComingThir/227287.

[5] Campbell, J. P. 2007. *Utilizing student data within the course management system to determine undergraduate student academic success: An exploratory study*. Unpublished doctoral dissertation: Purdue University.

[6] Campbell, J. P., DeBlois, P. B., and Oblinger, D. G. 2007. Academic analytics: A new tool for a new era. *EDUCAUSE Review, 42*(4), 40-42.

[7] Campbell, and Oblinger, D. 2007. *Academic analytics*. Washington, DC: EDUCAUSE Center for Applied Re-search.

[8] Cohen, J., Cohen P., West, S.G., and Aiken, L.S. 2003. *Applied multiple regression/correlation analysis for the behavioral sciences*. (2nd ed.) Hillsdale, NJ: Lawrence Erlbaum Associates.

[9] deJung, J. E., and Duckworth, K.1986. Measuring student absences in the high schools. Paper presented at *The Annual Meeting of the American Educational Research Association*, San Francisco, CA.

[10] Donathan, D. A.2003. The correlation between attendance, grades, and the nontraditional student, *Business Education Forum, 58*(1), 45-47.

[11] Draper, N.R., & Smith, H. 1998. *Applied Regression Analysis* (3rd ed.). John Wiley. ISBN 0-471-17082-8.

[12] Druger, M.2003. Being there: A perspective on class attendance, *Journal of College Science Teaching, 32*(5), 350-351.

[13] Friedman, P., Rodriguez. F., and McComb, J. 2001. Why students do and do not attend classes: myths and realities. *College Teaching, 49*(4), 124-133.

[14] Elias, T. 2011, January. *Learning Analytics: Definitions, Processes, and Potential. Creative Commons*. Retrieved from http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf.

[15] Johnson, L., Smith, R., Willis, H., Levine, A., and Haywood, K. 2011. *The 2011 Horizon Report*. Austin, Texas: The New Media Consortium. Retrieved from http://www.nmc.org/pdf/2011-Horizon-Report.pdf.

[16] Jo, I. 2012. On the LAPA (Learning Analytics for Prediction & Action) Model suggested. *Future Research Seminar*. Korea Society of Knowledge Management. Seoul.

[17] Jo, I., Kang, Y., Yoon, M., and Kang, M. 2012, Fall. Development of cluster-specific learning prediction models: A learning analytics approach, Paper presented at *the HYCU International Conference*, Seoul, Korea.

[18] Jo, I., and Kim, J. H. 2013. Investigation of Statistically Significant Period for Achievement Prediction Model in e-Learning. *The Journal of Educational Information and Media, 29*(2), 285-306.

[19] Jo, I. and Kim, Y. 2013. Impact of Learner's Time Management Strategies on Achievement in an e-learning Environment: A Learning Analytics Approach. *The Journal of Educational Information and Media, 29*(2), 285-306.

[20] Jung, N., Jo, I., and Lim, K. Y. 2003. Study of Influential Factors on Student's Learning Management System (LMS) Usage and Achievement: Focused on Role of Student's Self-regulated Learning Ability and Reality of Assignments. Proceeding at *The Korea Society of Management Information Systems Conference*, Seoul: Korea.

[21] Jung, I., Choi, S., Lim, C., and 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*(20), 153-162.

[22] Kahn, J. H. and Nauta, M. M. 2001. Social-cognitive Predictors of First-year College Persistence: The Importance of Proximal Assessment. *Research in Higher Education, 42*(6), 633-652.

[23] Kang, M., Kim, J., and Park, I. 2009. The examination of the variables related to the students' e-learning participation that have an effect on learning achievement in e-learning environment of cyber university. *Journal of Korean Society for Internet Information, 10*(5), 135-143.

[24] Kim, Y. 2011. *Learning time management variables impact on academic achievement in corporate e-learning environment*. Master's thesis. Ewha Womans University, Seoul.

[25] McGrath, P. A., and Braunstein, A. 1997. The prediction of freshmen attrition: An examination of the importance of certain demographic, academic, financial, and social factors. *College Student Journal, 31*, 396-408.

[26] Moore, R. 2003. Attendance and performance: How Important is It for Students to Attend Class? *Journal of College Science Teaching, 32*(6), 367-371.

[27] Moore, R., Jensen, M., Hatch, J, Duranczyk, I., Staats, S., and Koch, L. 2003. Showing up: The importance of class attendance for academic success in introductory science courses. *American Biology Teacher, 65*(5), 325-329.

[28] Piccoli, G., Ahmad, R., and Ives, B. 2001. Web-based virtual learning environments: A research framework and a preliminary assessment of effectiveness in basic IT skill training. *MIS Quarterly, 25*(4), 401.426.

[29] Rau, W., and Durand, A. 2000. The academic ethic and college grades: Does hard work help students to "make the grade"? *Sociology of Education*, 19-38.

[30] Swan, K. 2001. Virtual interaction: Design factors affecting student satisfaction and perceived learning in asynchronous online courses. *Distance Education, 22*(2), 306-331.

[31] Thompson, J. M. 1998. Developmental students in higher education: Path analysis of a national sample. *College Student Journal, 32*, 499-510.

[32] Thurmond, V. A., Wambach, K., and Connors, H. R. 2002. Evaluation of student satisfaction: Determining the impact of a web-based environment by controlling for student characteristics. *The American Journal of Distance Education, 16*(3), 169.189.

[33] Van Barneveld, A., Arnold, K. E., and Campbell, J.P. 2012. *Analytics in Higher Education: Establishing a Common Language* (white paper). Boulder, CO: EDUCAUSE Learning Initiative.