

Using learner profiling technique to predict college students' tendency to choose elearning courses: A two-step cluster analysis

Cheng-Chang “Sam” Pan ^{1, a} and Francisco García ^{2, b}

^{1, 2} The University of Texas at Brownsville, ^a Sam.Pan@utrgv.edu,
^b Francisco.Garcia@utrgv.edu

Abstract

Profiling elearning students is becoming a common practice in the field. In this phase of the investigation, we plan to (a) follow up on the recommendation for further research we stated in an earlier study on learner preference in types of elearning courses and (b) explore plausible patterns (profiles) based on two learner characteristics/behaviors (i.e., perceived distance between social life and school life and perceived affinity for technology) and their relationship with choices of learning environments where students learn most. Results suggested that (a) the probability of a student in favor of elearning was 1.29 times more likely when the student was LDAA, as opposed to HDLA, and (b) the probability of a student in favor of elearning was 1.26 times more likely when the student was HDHA, as opposed to HDLA. Implications of the results are discussed.

Keywords: College students; distance education enterprise, elearning; affinity for technology; separation of school life and personal life

Introduction

Categorizing, or arguably profiling, elearning students is becoming a common practice in the field (Archer, Chetty, & Prinsloo, 2014; Baxter, 2012; Yukselturk & Top, 2013), despite that profiling carries a negative connotation to some (Jones, 2012). Using a Web survey, an increasing number of learner characteristics and demographics can be studied in a form of data. Given this easy access to the collected data, researchers have attempted to take into account multiple profiling variables at once, instead of dealing with one variable at a time. This attempt makes the design of their studies more sophisticated and more versatile. It also assists the researchers in finding hidden patterns of the learners and their behaviors (Shih, Jheng, & Lai, 2010). Most importantly, their research results enable the top management team to make informed decisions. One major advantage of two-step cluster analysis adopted in this study is that it allows researchers to consider both continuous/numerical and categorical/nominal variables at a time as other clustering techniques, such as K-Means Cluster and Hierarchical Cluster in SPSS, are limited respectively, as Şchiopu (2010) pointed out.

Review of Literature

Literature shows that profiling students as a predicting technique can provide insights in who appears more college-ready than others and how successful college students strive to progress, against all odds, in their academic endeavors. In an attempt to ascertain student success, Purnell, McCarthy, and McLeod (2010) reported the promise and efficacy of an early warning system that identifies students at-risk prior to these

students' enrollment in any of the college classes at an Australian regional university.

They also argued that the timing of the institutional intervention is critical. As soon as the students at-risk are profiled, the intervention measures must be put in place. The sooner the personalized intervention is taken, the more effective those measures (e.g., receiving direct, immediate assistance with studies, and setting personal, realistic goals) are found.

To enhance open distance learning or ODL students' success, Subotzky and Prinsloo (2011) proposed a hypothetical model intended to predict the student success by taking into account the extent to which *both* the students and the university fit. Subotzky and Prinsloo argued that their broadly defined student success, in the dynamic context of ODL, is attributed to the degree of fittedness between the two agents (i.e., students and the university), of which the processes are constantly interactive and inherently transformative. Given the fluid nature of the context that characterizes those processes, the two "situated agents" (p. 184), as the two researchers of South Africa called it, there appears to be also unforeseeable consequences or uncertain events that are deemed less predictable than others.

Based on the proposed model by Subotzky and Prinsloo (2011), Archer, Chetty, and Prinsloo (2014) conducted a pilot program to profile successful students and students at-risk in terms of habits and behaviors (e.g., an inclination to change and own), using a commercial instrument. In spite of the resistance of the academics and their concerns over the legitimacy and trustworthiness of the questionnaire, which was initially developed for corporate use, the pilot program was found positive by the majority of student participants.

To profile distance learning students for student success, Onyancha (2010) turned to learner demographics for answers and categorized the learners by their “geographical location (country of residence), gender, occupation, age, and home language” (p. 159). Similarly, in examining learner profiling as a means to predict student behaviors in the online classroom, Yukselturk and Top (2013) studied online learners’ entry characteristics in a hope to explain these learners’ classroom behaviors between three distinct learner groups, clustered on gender and work status: male worker group, female-dominated group (with over 50% working), and male non-worker group. In conducting a pairwise comparison, the two Turkish researchers found a significant difference between female-dominated group and male non-worker group in the two classroom behaviors. According to the results of their follow-up comparison analysis, the female-dominated group participated significantly more in synchronous text-based chat sessions and more in asynchronous text-based discussion list (by posting more messages) than the male non-worker group. Though, the academic achievement (i.e., end-of-class grade) was not found significantly different between the three clusters in their study.

Also interested in learner demographics, Jelfs and Richardson (2013) investigated the age factor in testing the common assumption that digital natives outperform digital immigrants in the use of digital technologies. Their findings revealed that older age groups tend to be deep (with a defined goal to fully understand the content) and strategic (with a distinctive goal to score as most points as possible) learners rather than shallow and less strategic ones, when compared with the younger age groups. Though, when controlling for age, gender, and response mode (online vs. postal), the two British

researchers claimed that a positive attitude toward digital technologies can equally predict the deployment of deep and strategic approaches to learning.

Baxter (2012) found in a qualitative study that successful distance education students can form an identity of a student through communications with fellow students and the tutors (moderators) on the online discussion board. She asserted that at certain points of their school life, distance education students would have discontinued with the school were it not for the tutors' interventions that had helped with the transition phases. In a sense, Baxter's assertion is similar to what Purnell, McCarthy, and McLeod (2010) argued, which we previously mentioned. Both studies seemed to have suggested the importance of timing of the intervention. In Baxter's view, the personal and social identity of being a student is one key factor that explains why students remain motivated and resilient and continue to make due effort in order to succeed in the virtual learning environment. Though, Baxter and Haycock (2014) later noted using online discussion boards to enhance student motivation and foster student identity can be paradoxical within the context of community of practice. The social benefits of such use are more evident and more positive when the students are more engaged (a) in the academic or content aspect than the social aspect of the online forum use, (b) in an inviting and encouraging fashion, and (c) in structured, moderated discussion forums (Baxter & Haycock, 2014).

In this phase of the investigation, we planned to (a) follow up on the recommendation for further research we stated in an earlier study on learner preference in types of elearning courses (Pan, Sivo, García, Goldsmith, & Cornell, 2014) and (b) explore plausible patterns (profiles) based on two learner characteristics/behaviors (i.e.,

perceived distance between social life and school life and perceived affinity for technology) and their relationship with choices of learning environments where students learn most. In this context, we decided to conduct a two step cluster analysis to profile our students, a profiling technique suggested by Yukselturk and Top (2013). There is few research in the literature that closely resembles what we intended to do. Research questions we studied are as follows:

- Q1. To what degree are students clustered on two variables, perceived distance between social life and school life and perceived affinity for technology?
- Q2a. Is there any significant difference between student clusters in their propensity to choose elearning classes (as opposed to face-to-face classes) as the learning environment where they learn most?
- Q2b. To what degree do student clusters differ in their probability of favoring elearning when compared with each other?

Method

The present study was designed to continue with our series of research on how college students perceive their existing use and future needs of campus technology. Broadly speaking, the ultimate goal of this research project is to explain and predict the trends of college students’ use of information communication and technology, and eventually influence the trends to optimize student success.

This survey research was centered around student success with a sole emphasis on the quantitative nature of the inquiry. The data were initially collected online in collaboration with EDUCAUSE Center for Applied Research (ECAR) in 2013. These

secondary or archival data with a sample size of approximately 2,000 undergraduate students from a southern state university were analyzed for the quantitative research. The university is classified as a Hispanic-Serving Institution or HSI by the U.S. Department of Education. Below are selected demographics of the collected data.

Majority (87.7%) of the respondents were Hispanic. Female students took about 63% of the total respondents; 65% were between 18 and 24 years old; 55.7% were freshmen or sophomore; 94.1% lived off campus; 70.4% were full-time students. Table 1 shows descriptive statistics of studied variables in the present investigation.

Table 1

Descriptive statistics of three studied variables

Variable	Descriptive Statistics		
	Mean	Std. Deviation	N
Technology Affinity	39.97	9.41	1707
Distance Betw. Lives	3.76	1.29	1830
Environment Learn Most	2.17	.860	1830

Table 2 indicates the Pearson correlations between perceived affinity for technology, perceived distance or separation of social life and school life, and preferred learning environments where students learn most.

Table 2

Correlations between studied variables

		Technology	Distance	Environment
		Affinity	Betw. Lives	Learn Most
Technology Affinity	Pearson	1	.079**	.018
	Correlation			
	Sig. (2-tailed)		.001	.447
	N	1707	1700	1701
Distance Betw. Lives	Pearson	.079**	1	-.034
	Correlation			
	Sig. (2-tailed)	.001		.153
	N	1700	1830	1821
Environment Learn Most	Pearson	.018	-.034	1
	Correlation			
	Sig. (2-tailed)	.447	.153	
	N	1701	1821	1830

** . Correlation is significant at the 0.01 level (2-tailed).

As stated above, the data were gathered through an online survey in 2013, targeting the undergraduate students at the participating university. There were three studied factors. The affinity for technology variable is a latent factor, explained by 12 manifest variables

with each measured on a five-point Likert scale (e.g., strongly disagree and strongly agree at the ends of the continuum). A “Don’t Know” option was given. The internal consistency, Cronbach's alpha, was .89. Likewise, the distance between social life and school life factor was manifested by one variable, measured on a five-point Likert scale (e.g., strongly disagree and strongly agree). A “Don’t Know” option was also made available. Besides the two ordinal scales, the third factor dealt with in this study was the type of learning environments students tend to learn more. There were four options for survey participants to choose one form. Three of the options were also measured on an ordinal scale. The remaining one option was for students with no preference.

Results

To answer Q1, the collected data were analyzed using two step cluster analysis in SPSS. Three viable learner groups/profiles were identified, High Distance High Affinity (or HDHA), High Distance Low Affinity (or HDLA), and Low Distance Average Affinity (or LDAA). This profiling was based on average Silhouette = .5, which is considered fair, with the ratio of largest cluster to smallest cluster at 2.26 (<3). Cluster sizes vary. Respectively, they were 47.1%, 20.9 and 32%, % ($N=1694$). See Figure 1 and Figure 2 below for more information.

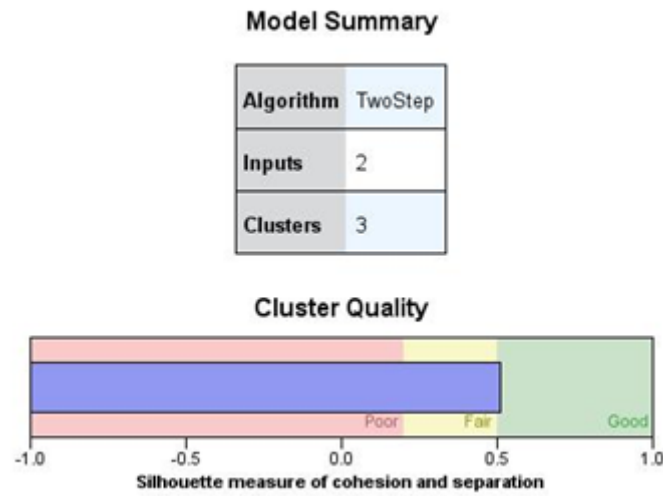


Figure 1. The model summary and quality as a result of the two step cluster analysis.

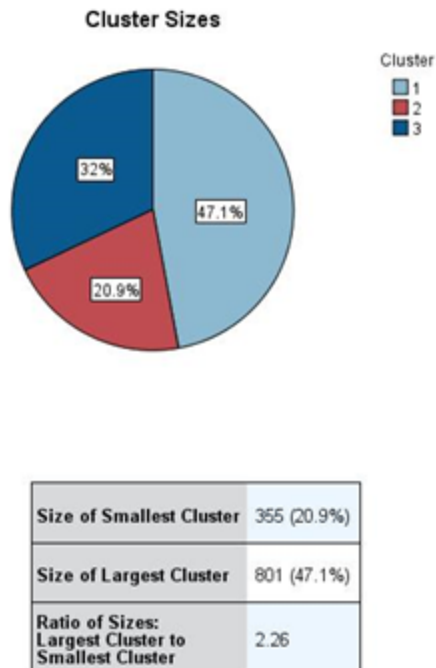


Figure 2. The cluster sizes and their ratio (largest cluster/smallest cluster) as a result of the two step cluster analysis.

In answering Q2a, with student's preference of learning environments regarded as a categorical or nominal variable, we were able to use SPSS to cross tabulate three student clusters and four choices of the preferred learning environment and found there is a scientifically significant difference in the way students in three different profiles perceive different modality as the environment where they learn most, Pearson χ^2 (6, $N=1694$) = 76.04, $p < .001$, Cramer's $V = .15$.

For the purpose of the present study, we proceeded with the investigation with a focus on the issue of students' tendency to sign up for an elearning class, as opposed to a non-elearning class (i.e., face-to-face class). To do so, we removed the no preference group and collapsed the remaining three groups into two: elearning group and non-elearning group. In so doing, we merged the group of courses with some online instruction and the group of courses completely online into one big group, named elearning group. The remaining group, courses without any online instruction, stayed the same and was considered non-elearning group. Afterwards, we re-ran the chi-Square procedure using SPSS and cross-tabulated three learner profiles (i.e., HDHA, HDLA, and LDAA) and two preferred learning environments (i.e., elearning and non-elearning). Further results are as follows.

With a two-way contingency table analysis using crosstabs, we evaluated whether students in any of the three clusters/profiles perceive elearning more as a learning environment to learn most. We found there is a scientifically significant difference in the way students in three different profiles perceive elearning as the environment where they learn most, Pearson χ^2 (2, $N=1481$) = 48.27, $p < .001$, Cramér's $V = .18$, suggesting

students’ tendency to register for elearning classes is correlated with their learner profiles. The proportions of students who perceived elearning is the environment that they learn most across three learner profile groups: HDHA, HDLA, and LDAA were .83, .66, and .85, respectively. Figure 3 below shows the frequency of both elearning and non-elearning counts within the three learner groups.

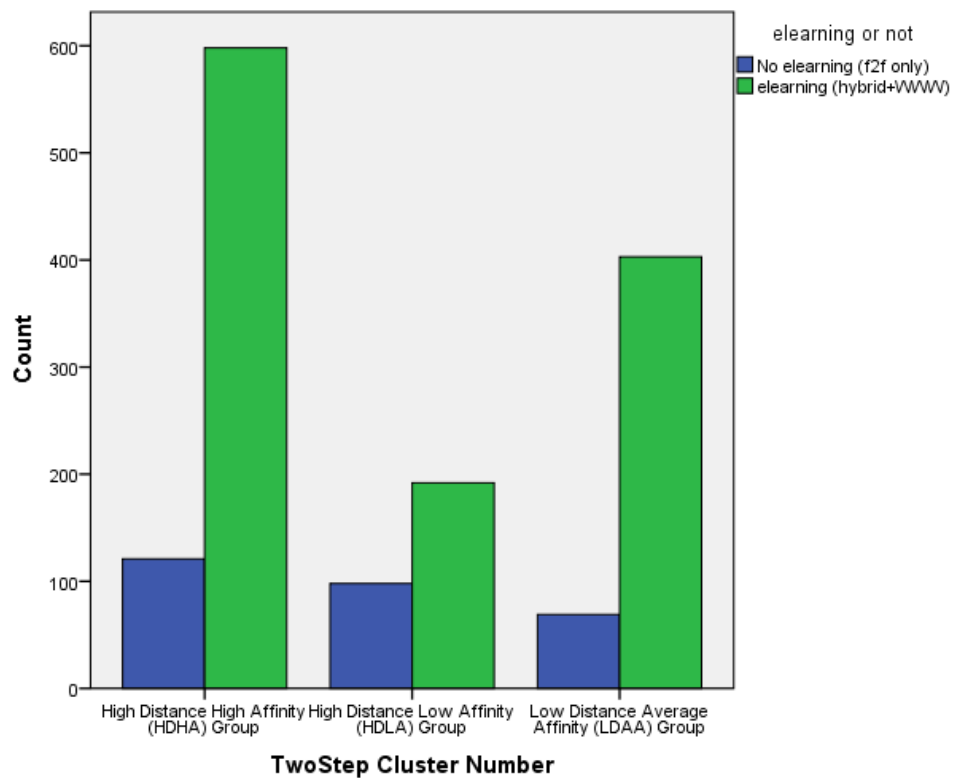


Figure 3. A cluster bar chart of elearning student counts within the student profile categories.

To answer Q2b, three follow-up pairwise comparisons were conducted to evaluate the difference among the proportions aforementioned. Table 3 below shows the results of the analyses.

Table 3

Results for the pairwise comparisons using the Holm's Sequential Bonferroni Method

Comparison	Pearson chi-square	<i>p</i> value (Alpha)	Cramér's V
HDLA vs LDAA	38.59*	<.001(.017)	.23
HDHA vs HDLA	35*	<.001(.025)	.19
HDHA vs LDAA	1.04	.308 (.050)	.03

**p* value \leq alpha

Results indicated that a scientifically significant difference was found between HDLA and LDAA, Pearson χ^2 (1, N=762) = 38.59, $p < .001$, Cramer's V = .23, and also between HDHA and HDLA, Pearson χ^2 (1, N=1009) = 35, $p < .001$, Cramer's V = .19. However, the pairwise comparison between HDHA and LDAA was not found significantly different, Pearson χ^2 (1, N=1191) = 1.04, $p = .308$, Cramer's V = .03. The probability of a student in favor of elearning was 1.29 (.85/.66) times more likely when the student is profiled as a LDAA as opposed to a HDLA; the probability of a student in favor or elearning was 1.26 (.83/.66) times more likely when the student is profiled as a HDHA, as opposed to a HDLA.

Conclusions

In this survey research, we anticipated to focus our investigation on student profiling in hopes to identify plausible student clusters and to explain how the three

clusters differ in their propensity to choose elearning courses, as opposed to completely face-to-face courses. The design of the study was quantitative in nature with a Hispanic-Serving Institution as its setting. The data were collected online through EDUCAUSE ECAR in 2013.

Three questions were studied and answered. *Q1. To what degree are students clustered on two variables, perceived distance between social life and school life and perceived affinity for technology?* We found three student clusters on the two dimensions: affinity for technology and distance between social life and school life. Three clusters were named: High Distance High Affinity or HDHA, High Distance Low Affinity or HDLA, and Low Distance Average Affinity or LDAA.

Q2a. Is there any significant difference between student clusters in their propensity to choose elearning classes (as opposed to face-to-face classes) as the learning environment where they learn most? Through a two-way contingency table analysis using crosstabs, we discovered that there is a scientifically significant difference between the three clusters in their tendency to choose elearning as the learning environments they learn the most. The proportion of the students within each cluster in favor of elearning was computed. The follow-up pairwise comparisons were then conducted to answer *Q2b. To what degree do student clusters differ in their probability of favoring elearning when compared with each other?* Three pairs of cluster comparisons were examined. The results suggested that (a) in comparison with students in High Distance Low Affinity or HDLA cluster, the chance of a student choosing elearning was 1.29 times greater when the student was of Low Distance Average Affinity or LDAA,

and (b) in comparison with students in High Distance Low Affinity or HDLA cluster, the chance of a student choosing elearning was 1.26 times larger when the student was of High Distance High Affinity or HDHA.

Evidently, students with high affinity for technology had a tendency to choose elearning for their preferred learning environments as the students believe they learn the most in those non-face-to-face classes. So far as the level of perceived distance between social life and school life was concerned, in both comparisons the level of perceived distance seemed irrelevant.

The institution of higher education, if committed to a distance education enterprise and eager to profile students to whom elearning classes are more appealing or less so, may strategize and devote its limited resources to identifying and targeting those students who were low in their affinity for technology. Students low in affinity for technology, according to the 2013 survey results, tended to feel less connected to the school life and professors, and less prepared in transferring majors and applying to graduate school, among others, via technology. Cautions must be taken before the findings are generalized to a different setting. Further research recommended may deal with profiling students who are attracted to face-to-face courses, include a longitudinal study on the predictability of student affinity for technology in student success, and model causal relationships between affinity for technology and its viable antecedents in the context of elearning.

Acknowledgements

This research was encouraged and supported by Dr. Clair Goldsmith, Vice President for Information Technology and Chief Information Officer of The University of Texas at Brownsville. The authors also wish to extend their appreciation to the two anonymous reviewers for their suggestions and comments on an earlier copy of this article.

References

- Archer, E., Chetty, Y. B., & Prinsloo, P. (2014). Benchmarking the habits and behaviours of successful students: A case study of academic-business collaboration. *The International Review of Research in Open and Distance Learning*, 15(1), 62-83.
- Baxter, J. (2012). Who am I and what keeps me going? Profiling the distance learning student in higher education. *The International Review of Research in Open and Distance Learning*, 13(4), 107-129.
- Baxter, J. A., & Haycock, J. (2014). Roles and student identities in online large course forums: Implications for practice. *The International Review of Research in Open and Distance Learning*, 15(1), 20-40.
- Jelfs, A., & Richardson, J. T. E. (2013). The use of digital technologies across the adult life span in distance education. *British Journal of Educational Technology*, 44(2), 338-351.
- Jones, S. J. (2012). Technology review: The possibilities of learning analytics to improve learner-centered decision making. *Community College Enterprise*, 18(1), 89-92.
- Onyancha, O. B. (2010). Profiling students using an institutional information portal: A descriptive study of the Bachelor of Arts degree students, University of South Africa. *South African Journal of Libraries & Information Science*, 76(2), 153-167.
- Pan, C., Sivo, S., García, F., Goldsmith, C., & Cornell, R. A. (2014, October). *Technology and me--what do students think?* Paper presented at the 64th International Council for Educational Media (ICEM 2014) Conference, Eger, Hungary.

- Purnell, K., McCarthy, R., & McLeod, M. (2010). Student success at university: Using early profiling and interventions to support learning. *Studies in Learning, Evaluation, Innovation & Development*, 7(3), 77-86.
- Şchiopu, D. (2010). Applying two step cluster analysis for identifying bank customers' profile. *BULETINUL*, 62(3), 66-75.
- Shih, M.-Y., Jheng, J.-W., & Lai, L.-L. (2010). A two-step method for clustering mixed categorical and numeric data. *Tamkang Journal of Science and Engineering*, 13(1), 11-19.
- Subotzky, G., & Prinsloo, P. (2011). Turning the tide: A socio-critical model and framework for improving student success in open distance learning at the University of South Africa. *Distance Education*, 32(2), 177-193.
- Yukselturk, E., & Top, E. (2013). Exploring the link among entry characteristics, participation behaviors and course outcomes of online learners: An examination of learner profile using cluster analysis. *British Journal of Educational Technology*, 44(5), 716-728.