Onliners versus On-grounders in Computer and Information Systems courses in Higher Education: A Two-Step Cluster Analysis

Alan Peslak
arp14@psu.edu
Information Sciences and Technology
Penn State University
Dunmore, PA 18512 USA

Lisa Kovalchick
kovalchick@pennwest.edu
Computer Science and Information Systems
Pennsylvania Western University
California, PA 15419 USA

Wenli Wang
wangw@rmu.edu

Paul Kovacs
kovacs@rmu.edu

Computer and Information Systems
Robert Morris University
Moon Township, PA 15108 USA

Abstract

Are students who prefer online education different from those who prefer on-ground education, and how? This is an important question because educational institutions need to better understand student segmentations. This research examined 251 survey responses from students enrolled in Computer Information Systems courses at three universities over five years (2016-2021) and reviewed student attitudes, perceived skills, and their sociological characteristics. Through two-step cluster analysis unsupervised machine learning, two distinct clusters of students emerged, namely Onliners and On-grounders. The top nine out of the eleven group characteristics for Onliners are: select more online courses, regard online instruction as effective, work better without supervision, rely less on classroom interactions in learning, value convenience, can prioritize, are better organized, better prepared, and older. By understanding these group characteristics, educational institutions can make better decisions in policy making, resources allocation, and student recruitment/retention.

Keywords: Online education, on-ground education, cluster analysis, machine learning, persona, cluster

1. INTRODUCTION

As the onset of the COVID-19 pandemic brought about an abrupt transition from traditional on-ground education to online learning, online education has become a focal point of research for educators. Although online education existed pre-pandemic, the pandemic has made it much more prevalent. Now with the shift to more relaxed COVID-19 restrictions and, hopefully, the end of the pandemic, decision-makers in higher education face challenges in making the right decisions concerning online versus on-ground education such as policy making, resource allocation, and student recruitment/retention.

These decision-making challenges concerning online versus on-ground education have resemblance to those faced by CEOs concerning remote versus in-office work. Some firms such as Yelp, AirBnB, 3M, Lyft, and Spotify have gone fully-remote (Lufkin, 2022). PayPal posts both "fully remote" and "opt for remote" jobs. SAP allows employees to choose from remote, in-office, or hybrid work (Smith, 2022). JPMorgan Chase's CEO, Jamie Dimon, on the contrary, has a long-held preference of in-office work (Shevlin, 2022).

Despite some CEO's preference, many firms have based their decisions of employee work locations on meeting employees' expectations. Microsoft surveyed over 31,000 employees in 31 countries in 2022, and 52% of them were willing to switch to fully-remote or hybrid jobs (Microsoft, 2022). Employees are now more likely to prioritize their health and wellbeing over work; this is especially true of employees who are parents and/or women (Microsoft, 2022).

Although in-office work may help strengthen culture, improve collaboration, and reinforce purpose (Markman, 2021) and remote work may reduce costs and offer flexibility, the long-term impact of remote work is inconclusive. Similarly, Peslak, Kovalchick, Wang, and Kovacs (2021) showed mixed-results pertaining to online education. Just as employees in different clusters have varied expectations and performances, students in different clusters have varied attitudes and learning effectiveness (Peslak, Kovalchick, Wang, & Kovacs, 2021). For instance, Bishop (2022) studied employees at different age groups and concluded that 81% of under-35-year-olds fear loneliness from long-term homeworking. Similarly, do younger students fear more adverse effects of online education?

There is limited literature that has examined student clusters and investigated whether there are distinct characteristics of students' preferring online education versus on-ground education. This research aims to fill in the knowledge gap and answer the following research questions.

RQ1: What are the subgroup models that emerge from a multi-year, multi-university student learning preferences survey?

RQ2: What are the characteristics of the student subgroups that emerge from a multi-year, multi-university student learning preferences survey?

The first research question was modeled after Stewart, Miller, Audo, and Stewart (2012) who used cluster analysis to identify patterns in student responses. The second research question aims to provide more insights of student segmentation.

This research employs cluster analysis in order to better understand online versus on-ground education and the groups of individuals that may have significantly different views toward these educational modes. Cluster analysis allows us to identify distinct student subgroups and their characteristics. Providing understanding on student segmentation, this research provides insights into the types of students who prefer online versus on-ground education. It provides decision support regarding policy making, resource allocation, and how to better market to, recruit, and retain students within potentially distinct subgroups.

2. LITERATURE REVIEW

Online versus On-ground Education

In 2020, as the arrival of the COVID-19 pandemic closed university campuses, a record number of students were forced into online classes. In the early months of the pandemic, estimates surfaced that at least 14 million students in the United States moved to online learning (Hess, 2020).

The expansion of online courses is not merely a product of the pandemic. There has been an increasing number of higher education students enrolled in online courses in the United States since the early 2000s. A comprehensive report by the Babson Survey Research Group in 2016 indicated that more than six million students were enrolled in at least one online course. This accounted for 31.6% of all college students (Seaman, Allen, & Seaman, 2018). In 2018, more than one third (35%) of college students in the United States took at least one online course and

17% were fully enrolled online (De Brey, Snyder, Zhang, & Dillow, 2021).

The Babson study also reported that the percentage of academic leaders who rated online education as good as or better than on-ground instruction was 57.2% in 2003. This outlook of the quality of online education has shown a pattern of steady improvement from 2003 until 2012, where 77.0% of the administrators in higher education rated online as good or better. Results since then, however, have been less positive, with the results for 2015 showing only 71.4% of respondents rating online as good or better (Seaman et al., 2018).

The Noel-Levitz National Online Learners Priorities Report (2012) found that the top three priorities of the eleven enrollment factors for both undergraduate and graduate students' choosing an online course were: convenience, flexible programming, and the ability to fit education into their work schedule. The survey results were based on data from 123,594 students at 109 institutions from the fall of 2009 through the spring of 2012. This report also found that sixty-five percent (65%) of online learners perceived their experiences exceeded their expectations while twenty-four percent (24%) of them perceived their experiences met with their expectations. Seventy-three percent (73%) of the online learners were satisfied or very satisfied with their experience. In the 2017 follow-up study, the original three priorities (convenience, work schedule, and flexible pacing for completing a program) still matter and seventy-four percent (74%) of online learners were satisfied with their online programs (The Noel-Levitz National Online Learners Priorities Report, 2017).

Ortega-Maldonado, Llorens, Acosta, and Coo (2017) applied the Analysis of Variance (ANOVA) method to build student profiles of those preferring face-to-face (i.e., on-ground) versus online education. Their results indicate online Master's students were older than on-ground counterparts and were living in different cities and even countries. Unlike the on-ground students, online students did not fit a 'recent graduates' profile. Most online students had a full-time job and tended to be practitioners without too much time to spend on sustained long activities. Unlike their study's focus on Master's students in Organizations Psychology and Human Resources in one university in Spain, our research focuses on Computer Information Systems (CIS) students at both the undergraduate and graduate levels across three universities in the United States.

The research of Vidanagama (2016) involved 209 undergraduate students enrolled in computer-related degrees and focused on the role of technology. He applied the Technology Acceptance Model to determine if factors associated with online learning (e.g., perceived attitude, perceived enjoyment, and perceived usefulness) are affected by technology. The study shows that students in computing degrees are more satisfied with online learning when the technological environment (Learning Management System, software used in courses, etc.) performs adequately and is easy to use. It can be inferred from this study that students in computing fields are critical – probably more than students in other degree programs – of the technological environment involved in online learning. Our research emphasizes the understanding of CIS students to further examine their characteristics.

Cluster Analysis

Cluster analysis is a statistical process wherein data are placed into groups (i.e., clusters) based on how closely each item relates to a given set of characteristics. Classification is considered the most common use of cluster analysis; subjects are separated into groups such that each subject is more similar to other subjects within its group than to subjects outside of the group (Qualtrics, 2022). The success of clustering lies in the distinctness of the clusters that result from its application; the goal is to increase the similarity of items within a group (i.e., cluster) and to increase the difference between groups (Tan, Steinbach, Karpatne, & Kumar, 2019).

Although it is getting a renewed interest within the emerging field of data science, cluster analysis is not a new concept, as it is often used to identify groups. As Scoltock (1982) noted, cluster analysis was first developed to study the fields of biology and zoology; within these fields, clusters were used to group plants and animals and to create taxonomies for the resulting groups. Since its emergence, cluster analysis has been used in a number of other industries to distinguish attributes of a large population of subjects; over the years, it has been commonly used in a variety of areas including: biology, psychology, social sciences, medicine, etc. (Tan et al., 2019). Recently, a few researchers have used cluster analysis to study pedagogy.

Handoyo, Mukhibad, Tusyanah, and Ekaningsih (2021) utilized K-Means Clustering to group the performance of lecturers based on online pedagogical practices. This study surveyed 278 lecturers at the Universitas Negeri Saemarang and used six variables, which included course content, teaching design, video quality, teaching service, teaching evaluation, and learning effect, to measure the performance of lecturers in online learning practices during the recent COVID-19 pandemic. The research resulted in the creation of three distinct clusters relating to lecturer performance, including low, moderate, and high performance. Based on these findings, the researchers were able to make recommendations to improve lecturers' performance in online learning which included: improving competence in operating technology-based media, adjusting the learning design to the learning conditions to make the learning process more interactive and efficient, making good and interesting learning videos, establishing intense communication with students, utilizing technology-based media when conducting learning evaluations, and motivating students to become more active (Handoyo, Mukhibad, Tusyanah, & Ekaningsih, 2021).

Koh and Chai (2014) administered a pre-course survey to teachers participating in Information and Communication Technology (ICT) lesson design professional development activities. They used the results of this survey to perform a cluster analysis to categorize teachers into groups based on their self-reported technological, pedagogical, and content knowledge (TPACK). The cluster analysis resulted in two categories of pre-service and in-service teachers, respectively. From these clusters, the researchers were able to determine that the initial TPACK differences observed in teachers lead to different effects on their perceptions of TPACK development at the end of their ICT lesson design professional development session (Koh & Chai, 2014).

Mulenga and Marbán (2020) studied the "online mathematics behaviors in the context of social media applications via online learning in mathematics activities" of high school student teachers and used cluster analysis which resulted in three clusters of students. These clusters were formed using variables relating to the extent that the student teachers use the Internet (chat, Google, wikis, etc.) when completing mathematics assignments. Although there were significant mean differences in the clustering, and the student teachers within the clusters exhibited different levels of online participation with mathematics activities, the authors concluded that these prospective mathematics teachers expressed "positive attitudes toward online learning behaviors and are likely to adopt e-learning during the coronavirus outbreak" (Mulenga & Marbán, 2020).

Mehanna (2004) examined pedagogic techniques with the goal of establishing effective e-learning practices in higher education. Effectiveness of the pedagogic techniques was determined by examining the students' outcomes on the courses that were reviewed. The results of this study revealed that seven clusters of pedagogies correlated with students' grades and an educational significance for all seven of the clusters was determined. Utilizing these pedagogies in online learning may lead to the enhancement of student learning (Mehanna, 2004).

Aggarwal and Sharma (2019) studied the performance of first year students in a Masters in Computer Application (MCA) post graduate program. The authors performed k-means clustering on the university exam data of these students and arrived at five clusters. Analyzing these clusters, the authors learned that females had better academic performance than males in the first year of the program (Aggarwal & Sharma, 2019).

Perrotta and Williamson (2018) examined the relevance of cluster analysis in categorizing and measuring online education, specifically focusing on algorithms used in learning analytics. Their focus was more on the introduction of the cluster analysis - a social science methodology - to education, rather than on the profiling of online students.

Cluster analysis, as an often-deployed methodology of studying market segmentations, is applied in this research to identify student segmentations. Since the field of online education is relatively new in comparison to the field of, say, biology, there is a lack of research in group characteristics in online education using cluster analysis. This study aims to fill this gap.

3. METHODOLOGY

An online survey regarding students' perceptions of the effectiveness of various course delivery methods was administered between the years of 2016 and 2021 at three universities: one private, one state-owned, and one state-related. The survey was IRB approved at each of the three universities and QuestionPro online survey software was used to administer the survey to students enrolled in CIS courses, regardless of major.

Two-step cluster analysis was employed to answer the two research questions regarding: whether there are specific groups of students who shared similar characteristics with regard to their attitudes toward online education, and, if so, identifying these characteristics.

Cluster analysis, or clustering, is an unsupervised machine learning task. It involves automatically discovering natural groupings in data. Unlike supervised learning (like predictive modeling), clustering algorithms in unsupervised learning only interpret the input data and find natural groups or clusters in feature space (Wilson, 2020).

The clustering algorithm is based on a distance measure that gives the best results if all variables are independent, continuous variables that have a normal distribution, and categorical variables that have a multinomial distribution. (IBM Statistics 19).

The two-step cluster analysis is a hybrid approach which first uses a distance measure to separate groups and then a probabilistic approach (similar to latent class analysis) to choose the optimal subgroup model. (Gelbard, Goldman, & Spiegler, 2007).

Silhouette score (i.e., silhouette coefficient) is a typical measure of the success of a clustering technique. It ranges from -1 to 1. A silhouette score of 1 means that the clusters are very dense and nicely separated; whereas, a silhouette score of 0 means that clusters are overlapping. A silhouette score of less than 0 means that data belonging to clusters may be incorrect.

Cluster results are considered appropriate when the silhouette score is greater than 0.2. Though 0.2 is regarded as a fair score (Boos, Wang, Karst, Hymel, & Pediatric Brain Injury Research Network, 2021), the goal in this study was to obtain a minimum silhouette score of at least 0.3 as an indication of more robust clustering. To achieve this outcome, an iterative process of eliminating variables was deployed.

SPSS 27 was used to perform a two-step cluster analysis on the data set. Rundle-Thiele, Kubacki, Tkaczynski, and Parkinson (2015) explained that two-step cluster analysis in SPSS uses the log-likelihood measure to reveal natural groupings in a data set. It forms clusters based on both continuous and categorical data (Chiu, Fang, Chen, Wang, & Jeris, 2001; Norusis, 2008). Data transformation prior to analysis is also unnecessary.

4. RESULTS

The overall survey response at three universities in years between 2016 and 2021 was nearly 700. However, some of the respondents did not complete the entire survey; therefore, the actual number of responses to each survey question varied by question. For this two-step cluster analysis unsupervised machine learning methodology, our research focused on the student respondents that answered all 34 survey questions. Since there were already over 250 survey responses with all 34 survey questions answered, the dataset for this research contains these 251 responses. Due to the lack of time to examine each and every survey question, this research did not apply statistical methods, such as imputation, to replace missing values in order to increase the size of the data set. Regardless of this caveat, the sample size of 251 still represents a robust group for valid research.

Among the 251 students who completed the survey, 36% were female and 64% were male. The majority of these survey respondents (59%) were in the age group of 18-21. The percentage of respondents in other age groups decreased, as the ages increased, as shown in Figure 1. This basic demographic information demonstrates that the sample is representative of the population of students enrolled in CIS courses at all three universities between the years of 2016 and 2021.

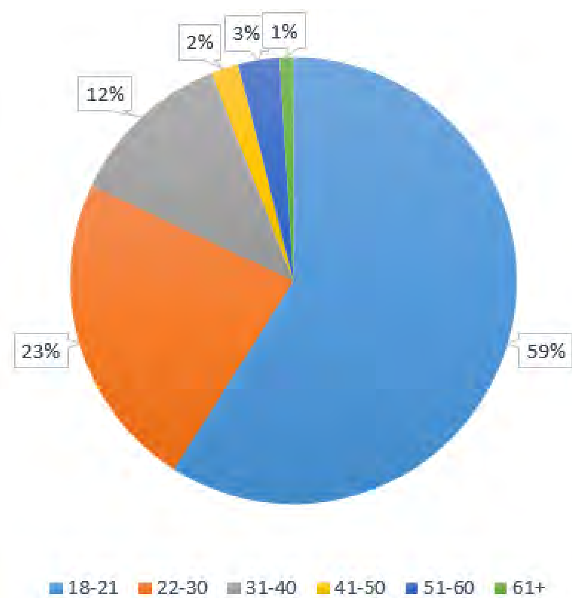


Figure 1: Percentage of Survey Respondents Grouped by Age

Research Question 1

What are the subgroup models that emerge from a multi-year, multi-university student learning preferences survey?

To answer this first research question, the following steps were taken in the data analysis.

The first step was to include all possible relevant variables in the cluster analysis. This pass included 26 relevant variables (out of 34 total variables) and did not produce any clustering, resulting in a lack of differentiation of any distinct groups or differentiated clusters.

Next, a review began to eliminate variables that were non-relevant or non-independent. This reduced the number of variables from 26 to 20. The questions where these 20 variables were extracted are shown in Appendix A. When cluster analysis was performed on these variables, two clusters were identified, as shown in Figure 2.

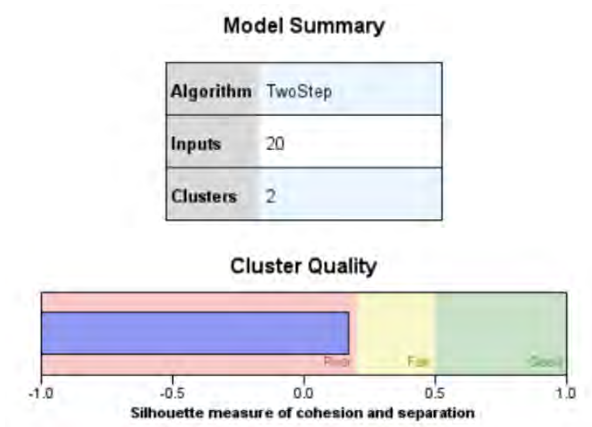


Figure 2: Clusters Obtained by Analyzing 20 Variables

The output of the 20-variable analysis depicts two clusters that were obtained from the two-step analysis. Details of the output are not illustrated in this paper due to the limitation of space. Variables were ordered based on importance, with the most important variable listed at the top. In this iteration, the most important variable was the expectation of online effort required. The first cluster had a highest selection of "less effort expected for online courses" with 68.5% expecting less effort; while the second cluster had a highest selection of "same effort expected for online and on-ground courses" with 53.2% expecting the same amount of effort. Similarly, each variable from the output can be interpreted in this fashion. The second most important variable was perception of online course effectiveness. Moving down through the list of variables, the importance of each variable becomes less in each cluster and the last two variables have no effectiveness. Since this second pass resulted in only an acceptable silhouette average of 0.2 and many variables had low or no impact, these low or no impact variables were regarded as less relevant and hence eliminated iteratively in subsequent passes in order to create more robust clusters, indicated by achieving a higher average silhouette such as 0.3.

In the next iteration, after eliminating the last two variables of the second pass that showed no effectiveness, the silhouette results remained at 0.2. Thus, variable eliminations were iteratively performed until achieving a silhouette of 0.3. This occurred when 11 variables remained as predictors.

The model summary graphic from SPSS is shown in Figure 3. The silhouette, though fair, has achieved the 0.3 goal.

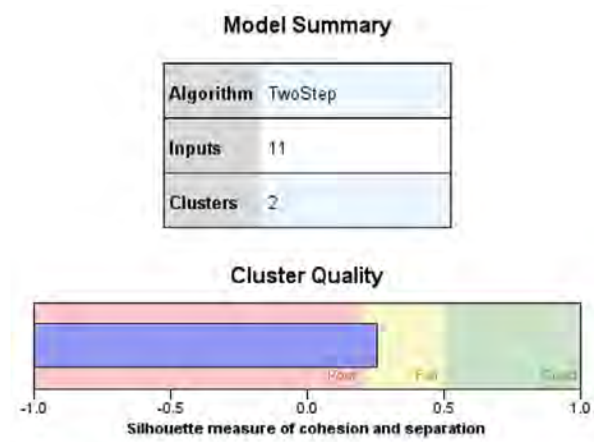
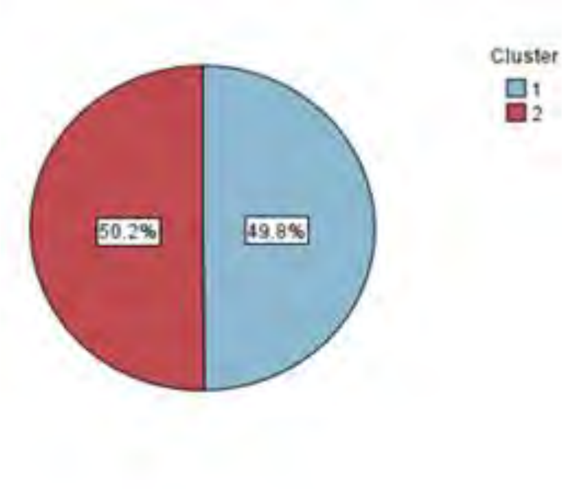


Figure 3: Clusters Obtained by Analyzing Remaining 11 Variables

As shown in Figure 4, there are two clusters of nearly equal size. These two clusters demonstrate that there are two distinct groups of students; those who prefer online education (Onliners) and those who prefer on-ground education On-grounders).



Size of Smallest Cluster	125 (49.8%)
Size of Largest Cluster	126 (50.2%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	1.01

Figure 4: Cluster Sizes

Onliners are students taking a course in which the format involves active instruction, testing, assignments, and discussion conducted over the Internet through a learning management system, in which the delivery of the course content is 100% online with no on-ground or in class meetings. Ongrounders are students taking a course on ground in the traditional brick-and-mortar classroom. Although an on-ground course might contain additional online resources such as assignments, videos, examinations, and podcasts the use of these additional resources are to enhance the class but the course is still on-ground. Finally, if the delivery format occurs when 25% - 50% of instruction, assignments, and discussions, take place online (hybrid), this online material is simply viewed as an alternative to in-person material with the intent to create a flexible learning experience.

Cluster 1 includes 49.8% of the survey respondents and represents Onliners and Cluster 2 includes 50.2% of the survey respondents and represents On-grounders.

It should be noted that during the iterative variable elimination process, some variables regarding demographics were eliminated such as gender, employment status, full-time versus part-time student status, etc. These eliminations indicated the non-significant impact of these demographics in cluster identification. The only impactful demographic variable in the remaining 11 variables was age, which will be discussed briefly later in the paper.

Research Question 2
What are the characteristics of the student subgroups that emerge from a multi-year, multi-university student learning preferences survey?

The tables displayed in Appendix B and Appendix C roughly demonstrate the clusters and the variables used to identify the characteristics in the clusters. These characteristics are listed in order of importance, with the most important variable listed first. Appendix B depicts some descriptive statistics; whereas, Appendix C provides some rough graphical presentations. Precise data analysis is provided in detail in later tables and figures.

Next, key variables were reviewed in detail, in terms of predictor importance and how they define clusters, with regard to online education. Figure 5 displays a graphical presentation of the predictor importance of each variable.

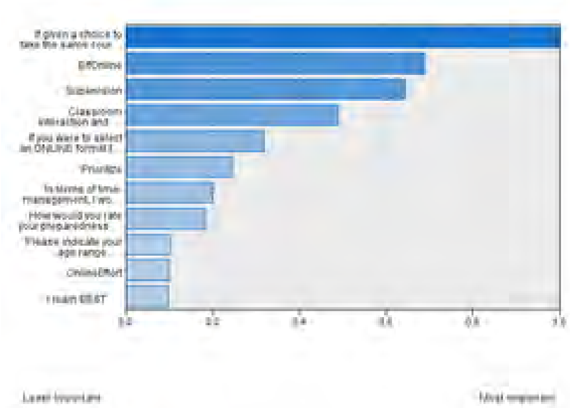


Figure 5: Predictor Importance for Each Variable

The variable with the highest predictor importance, shown in Table 1, is the question regarding selection of the online versus on-ground course format. Here, there is a clear dichotomy with 91% of the respondents in Cluster 1 preferring online and only 18% of the respondents in Cluster 2 preferring online. Hence, this reinforces the clarity of the conclusion that two discrete clusters exist in the data set, one that prefers online (i.e., Cluster 1, the Onliners) and another that prefers on-ground (i.e., Cluster 2, the On-grounders).

If given a choice to take the same course in an ONLINE format or an ONGROUND format, would you select the ONLINE format?				
	Yes		No	
Cluster	Freq.	%	Freq.	%
1: Onliners	100	91%	25	18%
2: On-grounders	10	9%	116	82%

Table 1: Frequency and Percentage Results, **by Cluster, for "Select ONLINE Format"**

The second most important predictor is the rating of the effectiveness of online instruction. As shown in Table 2, Cluster 1 (the Onliners) rated online education to be effective with an average rating of 2.27; while Cluster 2's (the On-grounder's) rating of online education leaned more towards somewhat ineffective with an average rating of 3.61.

	Effective Online	
Cluster	Mean	Std.
1: Onliners	2.27	.928
2: On-grounders	3.61	1.103
Combined	2.94	1.219

Table 2: Rating Results, by Cluster, for **"Effectiveness of Online Instruction"**

The ability to work with or without direct supervision was the next most important predictor. As shown in Table 3, 80% of the respondents in Cluster 1 (the Onliners) indicated they work best without direct supervision compared to those in Cluster 2 (the On-grounders), of which 78% work better with direct supervision.

Work better				
	Without Supervision		With Supervision	
Cluster	Freq.	%	Freq.	%
1: Onliners	96	80%	29	22%
2: On-grounders	24	20%	102	78%

Table 3: Frequency and Percentage Results, **by Cluster, for "The Ability to Work without Direct Supervision"**

As shown in Table 4, classroom interaction and discussion are not essential in learning for 86% of the respondents in Cluster 1 (the Onliners) and somewhat helpful for 65% of them; whereas, classroom interaction and discussion are always helpful for 83.5% of the respondents in Cluster 2 (the On-grounders).

Classroom interaction and discussion helpfulness in learning						
	Not		Sometimes		Always	
Cluster	Freq	%	Freq	%	Freq	%
1: Onliners	24	86%	86	65%	15	16%
2: On-grounders	4	14%	46	35%	76	84%

Table 4: Frequency and Percentage Results, **by Cluster, for "The Helpfulness of** Classroom Interaction and Discussion in **Learning"**

The ability to prioritize also distinguished Cluster 1 from Cluster 2. As shown in Table 5, 63% of the respondents in Cluster 1 (the Onliners) can prioritize well; whereas, 73% of the respondents in Cluster 2 (the On-grounders) lack the ability to prioritize.

Prioritize				
	Can		Can't	
Cluster	Freq.	%	Freq.	%
1: Onliners	100	63%	25	27%
2: On-grounders	59	37%	67	73%

Table 5: Frequency and Percentage Results, **by Cluster, for "The Ability to Prioritize"**

As shown in Table 6, time management skills are much more honed for Cluster 1 (the Onliners) than Cluster 2 (the On-grounders). Fifty-seven percent of the respondents in Cluster 1 consider themselves well organized, when it comes to time management skills; whereas, over 83% of the respondents in Cluster 2 indicated that they have difficulty completing assignments and/or projects.

Time-management				
	Well-organized		Not organized	
Cluster	Freq.	%	Freq.	%
1: Onliners	117	57%	8	17%
2: On-grounders	88	43%	38	83%

Table 6: Frequency and Percentage Results, **by Cluster, for "The Time Management Skills"**

As shown in Appendix B and Figure 5, age is the only impactful demographic predictor for cluster identification; however, its impact was less important than the predictors discussed above in detail.

The distinctive characteristics of all 11 variables for Cluster 1 and Cluster 2 are summarized in Table 7.

Variables	Cluster 1: Onliners	Cluster 2: On-grounders
Choose online	Yes (80%)	No (92%)
Effectiveness of Online	Effective (57%)	Somewhat effective (40.5%)
Need Supervision	No (77%)	Yes (81%)
Classroom interaction importance	Sometimes helpful (69%)	Always helpful (60%)
Reason for Online	Convenience (72%)	Scheduling (56%)
Able to Prioritize	Yes (80%)	No (53%)
Time management	Well-organized (94%)	Well-organized (70%)
Preparedness for Online	Extremely prepared (29%)	Extremely prepared (6%)
Age range	18-21 (50%)	18-21 (67.5%)
Effort required for Online	Same effort (44%)	Less effort (64%)
Learn best by	Hands-on (51%)	Hands-on (74%)

Table 7: Summary of Distinctive Characteristics, by Cluster, in Descending Order of Importance, for All 11 Variables

Viewing Table 7, we find that overall, the members of Cluster 1 (the Onliners) are better organized, able to prioritize, more self-reliant, and see classroom interaction as "not essential" and only "somewhat helpful." They also tend to be slightly older, slightly less inclined to learn using hands-on methods, and believe that online courses require the same effort as on-ground courses. The Onliners view online education as effective and choose online courses for convenience; over a quarter of them feel that they are extremely prepared for online learning.

In a similar fashion, using the data in Table 7, a profile can also be built to describe Cluster 2 (the On-grounders), who view online learning as only somewhat effective. The On-grounders need supervision and consider classroom interaction important. They are less able to prioritize and less organized than the Onliners. They also tend to be younger, more inclined to learn using hands-on methods, and believe that online courses require less effort than on-ground courses. These students often choose online learning due to its ease of scheduling; however, very few of them consider themselves extremely prepared for online learning.

The above results of student segmentation regarding online education are somewhat in alignment with the employee segmentation regarding remote work. For instance, younger people have more difficulty embracing fully online education or the remote work modality.

Regarding the age demographic characteristic, the result in this research echoes previous research conducted by the authors which depicts that for those students choosing online education due to scheduling, age rather than gender, plays a significant role in choosing the online modality (Wang, Peslak, Kovacs, & Kovalchick, 2019). Deeper investigation regarding other demographics such as different age groups and generations like those applied in the Microsoft (2022) study would provide further insights.

5. CONCLUSIONS

This research begins to fill the gap in the lack of studies utilizing cluster analysis to obtain group characteristics relating to online education. Through this research the authors were able to examine student clusters and investigate distinct characteristics of students preferring online versus on-ground education.

Utilizing an iterative process of performing two-step cluster analysis of their survey data and eliminating non-relevant variables, the authors were able to arrive at two distinct student clusters in the context of online education -- Onliners and

On-grounders. The 11 variables used to create the clusters indicate the characteristics of students within each cluster.

The 11 characteristics, found in this study, can be used to build a profile of a typical online student and that of a typical on-ground student. These profiles can be used by decision makers in higher education when making policies and allocating resources. For instance, this research suggests post-graduate programs embrace more online education than undergraduate programs.

These profiles can also be used in strategic planning with regard to how to market, recruit, and retain students for both online and on-ground educational programs. For instance, online post-graduate programs can be better marketed to employees who have already adopted fully-remote work.

The identification of specific online versus on-ground clusters and their identifying characteristics provides important insights to better understand students and also to better assist them in improving their acceptance and performance of online education, when necessary. Who knows what the future will bring – another pandemic or a climate change disaster could move education 100% online again. It is better to be prepared for the unknown.

This survey was limited in its audience to only those students enrolled in a CIS course at one of three universities. Therefore, one may conjecture that the majority of survey respondents were computing majors or students with some computing background. Surveying students enrolled in a variety of general education courses (humanities, fine arts, social sciences, etc.) would likely result in a more well-rounded characterization of students who prefer online versus on-ground education.

Now that profiles have been built to describe students enrolled in CIS courses who prefer online versus on-ground education, the authors can continue this study by making changes to the marketing strategies within their departments based on these profiles. Future research may examine the effects of these marketing changes by surveying students recently enrolled in these online and on-ground programs.

6. REFERENCES

Aggarwal, D. & Sharma, D. (2019). Application of Clustering for Student Result Analysis. *International Journal of Recent Technology and Engineering (IJRTE)*, 7(6C), 50-53.

Bishop, K. (2022). Is remote work worse for wellbeing than people think? *BBC*. June 17, 2022. Retrieved August 29, 2022 from https://www.bbc.com/worklife/article/20220616-is-remote-work-worse-for-wellbeing-than-people-think

Boos, S. C., Wang, M., Karst, W. A., Hymel, K. P., & Pediatric Brain Injury Research Network (PediBIRN) Investigators. (2022). Traumatic head injury and the diagnosis of abuse: a cluster analysis. *Pediatrics*, 149(1).

Chiu, T., Fang, D.-P., Chen, J., Wang, Y., & Jeris, C. (2001). A robust and scalable clustering algorithm for mixed type attributes in large database environment. *Proceedings of the 7th ACM SIGKDDD International Conference on Knowledge Discovery and Data Mining*, ACM SIGKDDD, San Francisco, CA, 263-268.

De Brey C., Snyder T. D., Zhang A., Dillow S. A. (2021). *Digest of Education Statistics 2019* (NCES 2021-009). National Center for Education Statistics.

Gelbard, R., Goldman, O., & Spiegler, I. (2007). Investigating diversity of clustering methods: An empirical comparison. Data & Knowledge Engineering, 63(1), 155-166.

Handoyo, E., Mukhibad, H., Tusyanah, T. & Ekaningsih, L. (2021). Lecturers' Performance in Pandemic Era Based on Online Pedagogical Practices in Universitas Negeri Semarang (UNNES), Indonesia: A Cluster Analysis-Based Approach. *Education Sciences: Theory & Practice (JESTP)*, 21(4), 138-154.

Hess A. (2020, March 26). *How coronavirus dramatically changed college for over 14 million students*. CNBC.com. https://www.cnbc.com/2020/03/26/how-coronavirus-changed-college-for-over-14-million-students.html

IBM SPSS Statistics 19 Statistical Procedures Companion

Koh, J. H. L., & Chai, C. S. (2014). Teacher clusters and their perceptions of technological pedagogical content knowledge (TPACK) development through ICT lesson design. *Computers & Education*, 70, 222-232.

Lufkin, B. (2022). The companies doubling down remote work. *BBC.* July 24, 2022. Retrieved August 29, 2022 from https://www.bbc.com/worklife/article/20220

722-the-companies-doubling-down-on-remote-work

Markman, A. (2021). Why you may actually want to go back to the office. *Harvard Business Review*. July 1, 2021. Retrieved August 29, 2022 from https://hbr.org/2021/07/why-you-may-actually-want-to-go-back-to-the-office

Mehanna, W. N. (2004). e-Pedagogy: the pedagogies of e-learning. *Research in Learning Technology*, 12(3), 279-293.

Microsoft. (2022). Great expectations: making hybrid work. Retrieved August 29, 2022 from https://www.microsoft.com/en-us/worklab/work-trend-index/great-expectations-making-hybrid-work-work

Mulenga, E. M. & Marbán, J. M. (2020). Prospective Teachers' Online Learning Mathematics Activities in the Age of COVID-19: A Cluster Analysis Approach. *EURASIA Journal of Mathematics, Science and Technology Education*, 16(9).

Noel-Levitz, (2012). Noel-Levitz national online learners priorities report. Retrieved from https://www.ruffalonl.com/upload/Papers_and_Research/2012/2012_Online_Leaners_Report.pdf.

Noel-Levitz, (2017). Noel-Levitz national student satisfaction and priority report. Retrieved from https://learn.ruffalonl.com/rs/395-EOG-977/images/2017_National_Student_Satisfaction_Report_1.0.pdf

Norusis, M.J. (2008), SPSS 16.0 Guide to Data Analysis, 2nd ed., Prentice Hall.

Ortega-Maldonado, A., Llorens, S., Acosta, H. & Coo, C. (2017). Face-to-face vs on-line: an analysis of profile, learning, performance and satisfaction among post graduate students. *Universal Journal of Educational Research*, 5(10), 1701-1706.

Peslak, A., Kovalchick, L., Wang, W. & Kovacs, P. (2021). Effectiveness of educational delivery modes: a study in computer information systems. *Journal of Information Systems Education*, 32(4), 253-261.

Perrotta, C. & Williamson, B. (2018). The social life of learning analytics: cluster analysis and the 'performance' of algorithmic education. *Learning Media and Technology*, 43(1), 3-16.

Qualtrics. (2022). What is cluster analysis? When should you use it for your survey results? (2022). Retrieved August 29, 2022 from https://www.qualtrics.com/experience-management/research/cluster-analysis/

Rundle-Thiele, S., Kubacki, K., Tkaczynski, A., & Parkinson, J. (2015). Using two-step cluster analysis to identify homogeneous physical activity groups. *Marketing Intelligence & Planning* 33, 522-537.

Scoltock, J. (1982). A survey of the literature of cluster analysis. *The Computer Journal*, 25(1), 130-134.

Seaman, J. E., Allen, I., E., & Seaman, J. (2018). *Grade Increase: Tracking Distance Education in the United States*. (ED580852). ERIC. https://files.eric.ed.gov/fulltext/ED580852.pdf

Shevlin, R. (2022). JPMorgan Chase gets "work from home" all wrong. *Forbes*. August 22, 2022. Retrieved August 29, 2022 from https://www.forbes.com/sites/ronshevlin/2022/08/22/jpmorgan-chase-gets-work-from-home-all-wrong/?sh=1badf24a7cb0

Smith, M. (2022). Twitter, Reddit and 8 other companies offering permanent remote or hybrid work—and hiring right now. *CNBC*. April 13, 2022. Retrieved August 29, 2022 from https://www.cnbc.com/2022/04/13/10-companies-that-switched-to-permanent-hybrid-or-remote-work-and-hiring-right-now.html

Stewart, J., Miller, M., Audo, C., & Stewart, G. (2012). Using cluster analysis to identify patterns in students' responses to contextually different conceptual problems. *Physical Review Special Topics-Physics Education Research*, 8(2), 020112.

Tan, P.N., Steinbach, M., Karpatne, A. & Kumar, V. (2019). Cluster Analysis: Basic Concepts and Algorithms. *Introduction to data mining* (2, pp. 525-612). New York, NY: Pearson. Retrieved June 15, 2022 from https://www-users.cse.umn.edu/~kumar001/dmbook/ch7_clustering.pdf

Vidanagama, D. U. (2016). Acceptance of E-learning among undergraduates of computing degrees in Sri Lanka. *International Journal of Modern Education and Computer Science,* 8(4), 25-32.

Wang, W., Peslak, A., Kovacs, P., & Kovalchick, L. (2019). What really matters in online education, *Issues of Information Systems*, 20(1), 40-48.

Wilson, J. (2020). How are clustering algorithms different from supervised learning? *Technical-QA.COM*. Retrieved June 15, 2022 from https://it-qa.com/how-are-clustering-algorithms-different-from-supervised-learning/

APPENDIX A

Survey Questions Used for 20 Variable Cluster Analysis

Note: the number of the question refers to the number in the original survey which contains 34 variables

2) If given a choice to take the same course in an ONLINE format or an ONGROUND format, would you select the ONLINE format?

- ☐ Yes
- ☐ No

3) If you did select an ONLINE format for a course, what would be the main reason?

- ☐ Convenience
- ☐ Scheduling
- ☐ Delivery Method
- ☐ To take a particular professor
- ☐ Other (please specify)

If you selected other, please specify________

4) I have taken (or am currently taking) a course that is <u>completely</u> online or is <u>partially</u> online.

- ☐ Yes
- ☐ No

6) What type of formal training did you receive to prepare you to take an online course?

- ☐ No formal training received
- ☐ Training and documentation provided by my school
- ☐ Self-trained
- ☐ Training from course instructor or other faculty member
- ☐ Training from another student
- ☐ Other (please specify)

If you selected other, please specify________

7) How would you rate your preparedness (to take an online course) prior to taking your online course?

- ☐ Extremely unprepared
- ☐ Somewhat unprepared
- ☐ Neither unprepared nor prepared
- ☐ Somewhat prepared
- ☐ Extremely prepared

8) Do you perceive the OVERALL effectiveness of courses that are offered <u>COMPLETELY</u> online as...

- ☐ Very effective
- ☐ Effective
- ☐ Somewhat effective
- ☐ Somewhat ineffective
- ☐ Ineffective
- ☐ Very ineffective

9) Do you perceive the OVERALL effectiveness of courses that are offered PARTIALLY online and

PARTIALLY onground (i.e., Hybrid) as...

- ☐ Very effective
- ☐ Effective
- ☐ Somewhat effective
- ☐ Somewhat ineffective
- ☐ Ineffective
- ☐ Very ineffective

10) Do you perceive the OVERALL effectiveness of courses that are offered ONGROUND but have an ONLINE SUPPLEMENT (i.e., online materials provided on BlackBoard or on an instructor's website) as...

- ☐ Very effective
- ☐ Effective
- ☐ Somewhat effective
- ☐ Somewhat ineffective
- ☐ Ineffective
- ☐ Very ineffective

15) Select one of the following choices

- ☐ I work better without direct supervision
- ☐ I work better when someone is there to keep me focused

16) Select one of the following choices

- ☐ I can prioritize my own workload
- ☐ I tend to put work off until later

17) Select one of the following choices

- ☐ I would allocate as much time and effort for an online course as I would for an on-ground course
- ☐ I feel that LESS time and effort is required for an online course (as compared to an on-ground course)
- ☐ I feel that MORE time and effort is required for an online course (as compared to an on-ground course)

18) In terms of time-management, I would describe myself as...

- ☐ Well organized
- ☐ Having difficulty completing assignments and/or projects

19) Classroom interaction and discussion is...

- ☐ Not essential for me to learn/understand
- ☐ Sometimes helpful for me to learn/understand
- ☐ Always helpful for me to learn/understand

20) Which of the following aspects could influence my decision to take an online course...

- ☐ Instructor teaching the course
- ☐ Design of the course
- ☐ Subject matter of the course
- ☐ Other (please specify)
  If you selected other, please specify________

22) I learn BEST...

- ☐ By seeing (visually)
- ☐ By listening
- ☐ By reading
- ☐ By doing (hands-on)

26) Are you enrolled as a ...

- ☐ Full-time student
- ☐ Part-time student

27) Which of the following best describes your living arrangement...

- ☐ Resident student (live on campus)
- ☐ Commuter student (live off campus)

32) Are you currently employed as a...

- ☐ Full-time employee (>40 hours/week)
- ☐ Part-time employee (
- ☐ Not currently employed

33) Please indicate your sex...

- ☐ Male
- ☐ Female

34) Please indicate your age range...

- ☐ 18 - 21
- ☐ 22 - 30
- ☐ 31 - 40
- ☐ 41 - 50
- ☐ 51 - 60
- ☐ 61 or older

APPENDIX B

Variables Identifying the Characteristics of Each Cluster, Listed in Order of Importance with Some Descriptive Statistics

Input (Predictor) Importance

1.0 0.8 0.6 0.4 0.2 0.0

Cluster	1	2
Label		
Size	49.8% (125)	50.2% (126)
Inputs	If given a choice to take the same course in an ONLINE format or an	If given a choice to take the same course in an ONLINE format or an
	EffOnline 2.27	EffOnline 3.61
	Supervision 1 (76.8%)	Supervision 2 (81.0%)
	Classroom interaction and discussion is	Classroom interaction and discussion is
	If you were to select an ONLINE format for a course, what would be the main reason?	If you were to select an ONLINE format for a course, what would be the main reason?
	Prioritize 1 (80.0%)	Prioritize 2 (53.2%)
	In terms of time-management, I would describe myself as .	In terms of time-management, I would describe myself as .
	How would you rate your preparedness (to take an online course) prior to ...	How would you rate your preparedness (to take an online course) prior to ...
	Please indicate your age range ...	Please indicate your age range ...
	OnlineEffort 1.85	OnlineEffort 2.15
	I learn BEST . . . 4 (52.8%)	I learn BEST . . . 4 (73.8%)

APPENDIX C
Relative Frequencies of Responses to Each Question/Characteristic in Graphical Form