ARTIGO ORIGINAL*

Two-stage cluster analysis in distance learning: A way to reduce gaps in the scientific literature on open and distance education

Análise de agrupamento (*Clusters Analysis*) em duas etapas no ensino à distância: Uma forma de reduzir as lacunas na literatura científica no ensino à distância

Guillaume Desjardins (1)

Cathia Papi (2)

Serge Gérin-Lajoie (2)

Louise Sauvé (2)

(1) *Université du Québec en Outaouais, Gatineau, Canada*

(2) *Université TÉLUQ, Québec, Canada*

Check for updates

Abstract

Background: Dropout rates are often very high in distance education. A plethora of research has been conducted to identify the contributing factors; however, the majority of the findings are inconclusive and point to the fact that it is difficult to isolate a single explanatory factor. While frequently examined factors are personal and environmental, there is less research on the relationship between course design and retention or dropout. **Method:** This paper presents a study involving two-stage cluster analysis of 623 variables from 19 university courses at one open and distance education (ODE) institution. To this end, the current study grouped the courses into five types based on 22 variables. **Results**: The results indicate that certain sociodemographic variables become a risk factor for course dropout depending on their distribution in the standard courses. **Conclusions**: This result highlights the importance of instructional design in the ODE retention and dropout equation and helps explain, in part, why previous studies have not reached a consensus on which variables should be considered to explain dropout rates.

Keywords: Online Course; Distance education; Higher education; Perseverance; Cluster analysis.

Guillaume Desjardins
Université du Québec en Outaouais, 283, boulevard Alexandre-Taché, Local C-3821 Gatineau (Québec) J8X 3X7, Canada
E-mail: guillaume.desjardins@uqo.ca

Resumo

Contexto: Embora as taxas de abandono escolar sejam frequentemente muito elevadas no ensino à distância, tem sido realizada muita investigação para identificar os fatores que influenciam o abandono escolar ou a persistência neste modo de aprendizagem. As conclusões destes estudos nem sempre convergem e salientam que é difícil isolar um único fator explicativo. Embora a maioria dos fatores sejam pessoais e ambientais, há menos investigação sobre a relação entre a conceção e a retenção ou desistência do curso. **Método:** Este estudo apresenta uma metodologia que envolve uma análise em duas fases de 623 variáveis de 19 cursos universitários de uma instituição de ensino à distância (EAD). Este estudo agrupou os cursos em cinco tipos de cursos com base em 22 variáveis. **Resultados**: Os resultados indicaram que certas variáveis sociodemográficas se tornam um fator de risco de desistência dos cursos, dependendo da sua distribuição nos cursos padrão. **Conclusão**: Esta metodologia sublinha a importância da conceção instrucional na equação de retenção e desistência da EAD e ajuda a explicar, em parte, porque é que estudos anteriores não chegaram a um consenso sobre quais as variáveis que devem ser utilizadas para explicar a desistência.

Palavras-Chave: Curso online; Educação à distância; Educação superior; Perseverança; Análise de cluster.

Introduction

Dropout rates are often very high in distance education. A raft of research has been conducted over the last few decades to identify the contributing factors influencing dropout rates in this type of education. However, the results are inconclusive and point out that it is difficult to isolate a single explanatory factor. For example, while Bennett (2003) indicates that learners' financial difficulties are the main reason students drop out of school, Willging and Johnson (2009) suggest that this risk factor does not act alone. According to the authors, personal, work-related, program-related, and technology-related factors are also at play. In their meta-analysis, Lee and Choi (2011) identify 69 different factors for dropping an online course in a distance education setting. Such magnitude in the number of variables renders grouping an onerous task (Arifin, 2018). The takeaway of these studies is that sociodemographic variables do not converge from one study to another. Therefore, it is difficult for scholars to pinpoint the unavoidable variables to include in their research protocol.

While the literature shows most factors are personal (sociodemographic characteristics, motivation, etc.) and environmental (including work-family-study balance and support received or lack thereof), there is a dearth of research on the relationship between course design and student retention or dropout. Very few studies on this matter focus primarily on assessing instructional structure (Garratt-Reed et al., 2016; Hammond & Shoemaker, 2014). Yet, Bernard et al. (2004) indicate that course instructional design would affect persistence. Despite this finding, no study to date has been able to establish a typology of online courses.

Regarding the methodology of studies in education, the lack of research that crosses student sociodemographic data with course instructional design may be explained by the methods used in the field of education. According to Tan and Shao (2015), education research primarily employs four types of methods: 1) data collection through questionnaires and analysis of correlations between variables (No & Hirakawa, 2012; Yasmin, 2013); 2) identification of trends in the distribution of demographic variables in relation to dropout variables, using data acquired through questionnaires or transcribed information systems (Garrison 1987; Rekkedal, 1985); 3) identification of reasons for dropout through analysis of transcripts of interviews conducted with relevant parties (Bakker et al., 2019; Li et al., 2010; Ramsdal et al., 2018), or 4) a combination of these three methods (Arifin, 2018; Li et al., 2012).

This research aimed to explain why previous studies have not reached a consensus on which sociodemographic variables should be considered to explain dropout rates in distance learning. To achieve this objective, we documented the correlation of students' sociodemographic variables with their retention in a post-secondary online learning context, using course design as a moderator. This study has five sections. First, the literature review highlights elements in the pedagogical design that have been linked to student's dropout. Second, the methodology is presented. This section stands out by the inclusion of clusters analysis as part of the methodology. Third, results — including the identification of model courses and student's sociodemographic variables are presented. Fourth, a discussion will emphasize the main findings of the study by presenting a student's dropout model using pedagogical design as a moderator. Fifth, a conclusion will be provided.

Review of the Literature

Pedagogical design in ODE

Most literature on distance learning environment have been not very articulate around pedagogical design. Despite the growing interest in the design, conception, or pedagogical organization (terms that will be considered synonymous here) of ODE courses, there is not much work on the relationship between course design and student retention or dropout. The literature on the topic highlights that students value courses with a logical content structure which enables them to easily find what they are looking for in the course material (Hammond & Shoemaker, 2014). Students also appreciate having clear instructions, guidance, and interactive learning materials (Garratt-Reed et al., 2016). Similarly, Burns (2013) indicates that some students appreciate the fact that distance courses are highly organized and structured so that there is a guided learning process, interventions by the instructor, a pace to follow over the weeks, and synchronous meetings, when necessary. In other words, students do not have to decide how to organize their learning experience. Rather, they only need to experience what is offered to them. Thus, distance students are aware that the pedagogical strategies deployed in their courses create conditions that may encourage them to persevere or not (Jézégou, 2010; Nichols, 2010).

In addition to research on students' appreciation and perceptions of the pedagogical design of their distance education courses, some research findings point to the effects of course design on distance students' retention and dropout. For instance, Pittenger and Doering (2010) postulate that scaffolding processes in course design help to make courses engaging, interesting, and motivating to students. In the same vein, the systematic review results by Monteiro et al. (2017) highlight some significant findings. As an instance, the authors draw on Kalet's et al. (2013) work that indicates that the importance (weighting) given to exams is correlated with the course dropout rate. Moreover, Flynn (2015) demonstrated that repeating weekly activities and having courses with an established form and structure promotes persistence, as does incorporating student support materials into course materials (Robinia, 2012). However, other elements, such as quizzes about course syllabi or formulating contracts with students, do not influence withdrawal (Leeds et al., 2013).

Methodology

The study was conducted in two phases. In Phase I, the pedagogical structure of 19 courses from a university distance learning institute was systematically analyzed. This analysis aimed to identify the possible link between the design of the courses and the propensity of students to drop out. In phase II, data was gathered from the

students at the same university through a questionnaire, which aimed to examine the relationship between students' sociodemographic variables and course withdrawal.

Stratified sampling was used in this project to reach the student population as well as their courses. Three factors were considered when selecting courses for analysis, namely, the department, the number of students per course, and the variability of failure and dropout rates in courses. The courses selected were from at least four different departments (language department, social sciences, education science, and administration science). According to the data obtained from the registrar for the year 2016, courses have an average of 312 enrollments per year. Therefore, only courses with at least 300 annual enrollments were selected to be included in the study. Regarding the third consideration, the variability of failure and dropout rates in courses, 19 courses with failure rates ranging from 4.4% to 26.5% and dropout rates between 4.3% and 26.4% were selected in order to account for diversity and ensure sample representativeness.

A course analysis grid was developed and used to obtain data concerning the pedagogical organization of 19 courses in a distance education institution in Quebec, Canada. The cluster analysis of the variables selected to process was carried out in two steps. First, a matrix of variables with the codifications of the design grid was developed in SPSS v21. This matrix had 623 variables, which encompassed all the data collected per courses. Afterward, the course data were systematically entered into the statistical software.

The second step was to measure the relative weight of dropout by course. To this end, the registrar data for the cohort of students who took the courses for the fall 2017 semester were examined. The researchers paid special attention to ensure that no changes in the instructional design of the courses occurred between fall 2017 and the course analysis date. Subsequently, each student who had taken one of the designated courses was entered into an entry line in the statistical software, and their course progress was specified (dropout = 1, persistence = 0) as the last variable. For example, if a course had 100 students enrolled in the fall 2017 semester with ten dropouts, 100 entries of the course (and its associated variables) were formed in the statistical software, with ten of those entries having "1" as the last variable in order to indicate the dropouts.

This choice to weigh course enrollment in the statistical analysis served two purposes. First, it took into consideration the weight of course enrollments as a value during the analysis. Second, weighting the enrollment rate by course analysis provided a clearer picture of the training offered at the institution. For example, of 20,671 enrollments in the fall 2017 semester, the status of 11,078 students was analyzed, which represented 54.0% of the total enrollment for the period covered.

The first version of the two-step cluster analysis was based on the entire course data. Then, iteration by iteration, the researchers removed variables deemed insignificant through the analysis until the cohesion and separation index was equal to or greater than 0.5. Finally, five clusters that were statistically distinct from 22 variables were identified. This procedure resulted in the formation of course-types or archetypes courses which not only functioned as theoretical constructs but statistically allowed to determine whether course design could influence student retention and success.

In parallel with the instructional design analysis, sociodemographic data from the students was obtained via a questionnaire. The questionnaire consisted of eight sections and covered a variety of topics, including sociodemographic data, use of learning strategies, satisfaction with the guidance received, and the digital tools offered in the course. In addition, other variables such as the degree of satisfaction of students' interactions with their peers and their family were included.

The sociodemographic and academic data section of the questionnaire consisted of 20 questions. Excluding the student number and the school and the email address, each variable in this section came from the scientific literature, indicating a relationship between the sociodemographic variables and perseverance in studies (Sauvé et al., 2014). Table 1 represents all the sociodemographic variables measured by the survey. The construction of the questionnaire was carried out following the three stages proposed in the literature on this subject (Cooper & Schindler, 2011; Franklin et al., 2010), namely the pretest, the pilot, and the administration.

Table 1

Student's Sociodemographic Variables Measured in the Survey

Student's status	Father's education	Current session	Numbers of hours worked per week
First language	Family status	Total credit earned in the current program	Presence of disability
Gender	Distance of the campus	Total credits currently enrolled in this session	Judgment of financial situation
Age	Student's education	Type of program	Number of credits
Marital Status	Attendance of international education	Scholarship financing	Mother's education

The questionnaire was administered online using the Limesurvey platform (limesurvey.org) in accordance with the rules and restrictions indicated by the ethics committee. Questionnaires were available between summer 2018 and summer 2019. Participants who received an invitation from the researchers had to click on a URL link to have access to the questionnaire webpage. The questionnaire was available only in French and could be completed using a computer, tablet, or smartphone.

At the end of the semester, the institution's registrar provided the research team with the final course grades and enrollment status information (enrolled or dropped the course). As a prerequisite to accessing the questionnaire, participants were required to read an agreement to disclose their information.

Results

The results of the study will be presented in three parts. The first part presents the cluster analysis conducted to identify the five archetypes. The second part reports on further analysis based on the data collected with regards to the crossing variables specific to the design of the courses and the sociodemographic variables of the students and their dropout from the courses in question. In the end, a model will be proposed to help understand the results obtained.

The Identification of Model Courses

In what follows, the two-step cluster analysis of the 19 courses is presented, and course-types are elaborated upon. Two courses formed Course-type 1; three courses formed Standard Course 2; eleven courses formed Standard Course 3; one course formed Standard Course 4, and two courses formed Standard Course 5. The variance in the distribution of students across the course-types is acceptable since

the difference between the largest cluster does not exceed 3.5 times the size of the smallest (Dolnicar, 2003; Tan et al., 2013).

Course-type 1

Course-type 1 is the one during which (15 sessions/modules[1]) students are called upon to carry out reading comprehension activities in books or texts, as well as activities that aim at the appropriation of new knowledge in almost half of the sessions (7 sessions), along with practical exercises (6 sessions). The archetype offers almost no discovery activities or visits to external sites. In addition, students are required to complete summative assessment activities in more than one-third of the sessions or modules (6 sessions), although formative assessment activities are offered in few sessions.

Course-type 2

In this archetype, students are asked to perform new knowledge appropriation activities (13 sessions) and visit external Websites (12 sessions) in almost all 15 modules of the semester. In more than half of the sessions, students are required to read texts or books (9). There are few sessions for discovery activities (2), writing a visit or research report (1), or doing practical exercises. In addition, students are required to complete formative assessment activities such as quizzes or exercises in two-thirds of the sessions (10 sessions) in which no summative assessment is given.

Course-type 3

Course-type 3 encompasses the one in which, in almost all 15 modules, students need to do activities that aim to acquire new knowledge (14 sessions) and to do readings in several sessions (12 sessions) and practical exercises (11 sessions). In addition, students are required to visit Websites external to the course in almost half of the sessions (6 sessions). Very few sessions are dedicated to discovery activities (1 session), and the writing of a visit or research report is absent. With regards to assessment, students are required to complete formative assessment activities in five of the sessions. However, there is no quiz or questionnaire to be completed as part of a summative assessment activity.

Course-type 4

In this type of course, students are required to do practical exercises (14 sessions), activities for the appropriation of new knowledge (12 sessions), and, very rarely, reading activities (4 sessions). Other activities such as visits to external Websites, discovery activities, and the writing about a visit or research reports are absent. As for the assessment, students are required to complete formative assessment activities in 13 of the sessions, but they are not given a quiz or questionnaire to complete as part of a summative assessment activity.

Course-type 5

The last type of course is the one in which students are asked in almost all 15 modules to perform activities for the appropriation of new knowledge (13 sessions) and discovery activities (13 sessions). Several sessions are dedicated to reading passages from books or other texts (12 sessions), while only one session

---

[1] The terms sessions and modules used in this section have no distinction in meaning.

is assigned to practical exercises. In this sample course, students are required to complete formative assessment activities in eleven of the sessions or modules, but they do not need to complete a quiz or questionnaire as part of a summative assessment activity.

Factor Analysis of Sociodemographic Variables with Course Dropout

The first statistical analysis cross-referenced the sociodemographic variables from the questionnaire with the data on course dropout for the students in our sample to see if certain sociodemographic factors were related to dropout in each course studied. To do this, a factor analysis was conducted. Although factor analysis is not recommended for categorical variables, some authors (Baglin, 2014; Flora & Curran, 2004) maintain that to the extent that variables can be classified ordinally, factor analysis is considered appropriate in that it can provide insights in an exploratory study. In fact, factor analysis, without indicating the direction of an association, can allow us to keep the sociodemographic variables that have an influence on the students' propensity for dropping out of a given course. Moreover, factor analysis offers a holistic view of the variables and their co-dependency. In other words, this procedure allows us to confirm that a variable is not found in the analysis model only because of its strong correlation with a second sociodemographic variable. Using the Eigenvalue (Marcus & Minc, 1988) provided by factor analysis, it is possible to estimate the variance in course dropout. Table 2 shows the results of the factor analysis.

Table 2

Factor Analysis of the Sociodemographic Variables using Maximum Likelihood Extraction Method

	Eigenvalue			Extraction of squared sums			SSR
Factors	Total	*POV*	*PC*	Total	*POV*	*PC*	Total
1	2.12	15.15	15.15	1.39	9.92	9.92	1.45
2	1.97	14.04	29.19	1.49	10.62	20.54	1.53
3	1.43	10.24	39.43	1.19	8.52	29.06	1.14
4	1.23	8.76	48.19	0.95	6.75	35.81	0.95
5	1.19	8.51	56.70	0.92	6.57	42.38	1.23
6	1.13	8.05	64.75	0.49	3.53	45.91	0.56

Note. Extraction method: Maximum Likelihood (only the cases where yes or no = 1 were used in the analysis phase). *POV* = percentage of variance; *PC* = percentage cumulative; *SSR* = Sum of squared rotations.

Thus, as shown in Table 3, from among the 20 variables in the questionnaire, ten variables were divided into six factors that were considered significant for our sample.

Table 3

Distribution of the Sociodemographic Variables According to the Factor Analysis

Factors	Variable(s)
Factor 1 (9.9% of total variance)	Student status
	The presence of a previous degree from abroad
Factor 2 (10.6% of total variance)	Family and marital status
	Judgment of financial situation
Factor 3 (8.5% of total variance)	Student's first language
	Type of program
Factor 4 (6.8% of total variance)	Gender
Factor 5 (6.6% of total variance)	Parents' education (mother and father)
Factor 6 (3.5% of total variance)	Presence of disability (according to the registrar's data)
	Number of paid hours worked per week

Discussion and the Proposal of a Model

Following the statistical analyses presented in this chapter, the following model can be proposed to explain dropout rates in distance education courses based on the characteristics of their pedagogical design.

This model, illustrated in Figure 1, leads to three conclusions. First, it is possible to group courses according to certain variables with the two-step cluster analysis. Of the 623 variables identified in the course analysis grid, 22 variables can be grouped to form clusters or course-types that are homogeneous within groups and relatively different from one another. Our sample of 19 courses yielded five consistent course types. When the sociodemographic variables were controlled to test the propensity for students to drop the course-by-course type, no significant differences were found. In other words, the clustering of the course's instructional design alone could not explain whether or not a student dropped a course.

Second, of the 20 sociodemographic variables identified in the questionnaire, ten variables were deemed significant in explaining the high rate of students' dropout (up to 45.9%). These ten variables were further grouped into six factors.

Third, the triangulation of the standard courses with the relationship between student sociodemographic variables and course withdrawal demonstrates that certain archetypes play a protective role (moderating role) for students' dropout. For instance, sociodemographic variables that are statistically significant in their relationship with course withdrawal when crossed alone with course withdrawal become non-significant when they are in standard Courses 1 and 5. In contrast, for Course-types 2, 3, and 4, the degree of the interdependence between certain sociodemographic variables and dropout increases. As an example, the correlation of marital status and family situation for Archetype 2 and 4 and financial status judgment for standard Course 3 for explaining students' dropout is augmented.

Figure 1

Distance Online Learning Student's Dropout Model Using Pedagogical Design as a Moderator
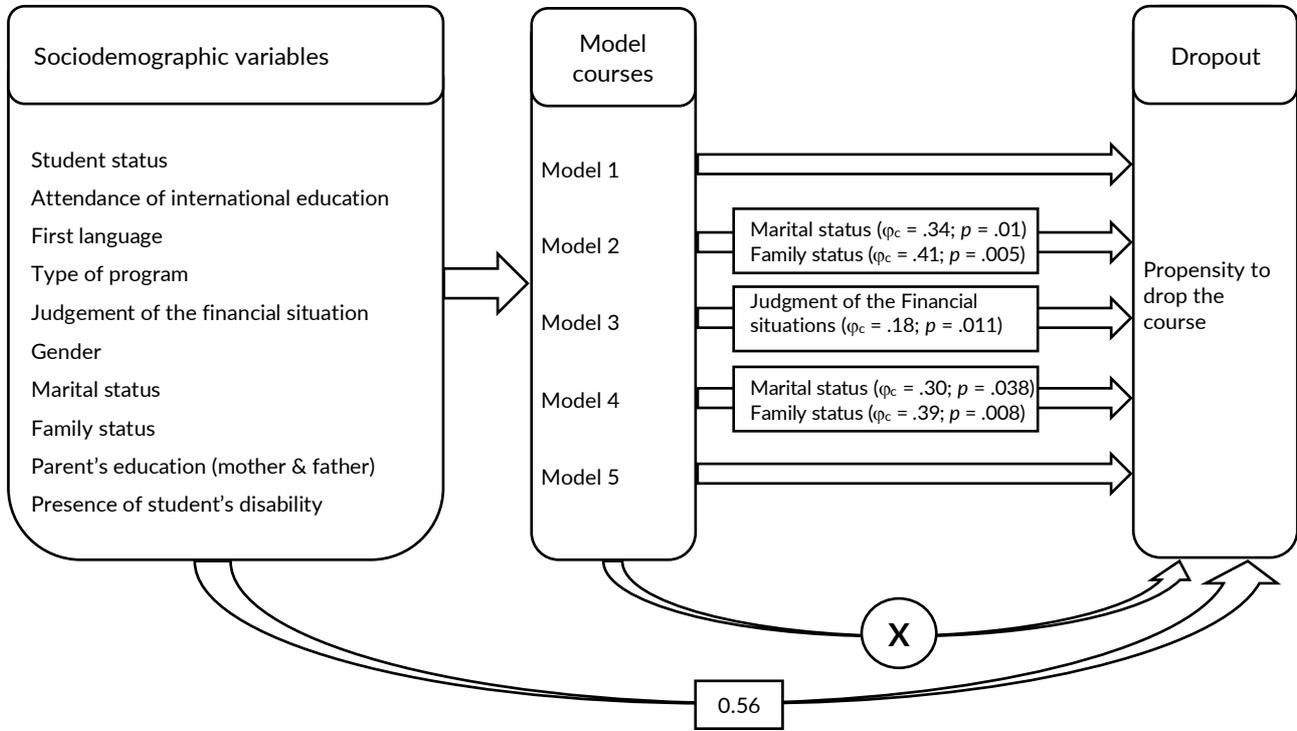
Moreover, these results also indicate that specific course design choices may increase the risk of dropouts among certain students. For example, civil and marital status play a prominent role in explaining the student's dropout, especially in Course-types 2 and 4. Specifically, students living with their spouses with or without children are correlated to a higher dropout rate. A fine analysis of those course-types reveals that both use more often quizzes and formative activities in the session when compared to Course-types 1 and 3. Thus, it would be possible to make a hypothesis that students that have a compressed schedule – when compared to, let's say, a student that is single and living with both parents — may contemplate the work needed to complete to course as too high and, therefore, drop out of the course.

Based on the results of this research, the model presented in Figure 1 helps explain some of the inconsistencies in the literature regarding the influence of students' sociodemographic variables on their propensity to drop courses. While several studies demonstrate a relationship between certain student-specific variables and course withdrawal (Jun, 2005; Packham et al., 2004), other studies do not reach the same conclusion (Kember et al., 1992, 1994). The proposed model can potentially explain the non-convergent conclusions of the previous studies. For instance, it can be argued that studies in which no statistically significant relationship between student sociodemographics and course dropout was found were predominantly from a type of course that offered protection against this risk.

Conclusion

The purpose of this article was to demonstrate that the instructional design of distance education courses plays a moderating role in the relationship between students' sociodemographic variables and course dropout rate. While there is ample research evidence regarding this relationship between the two, authors have not yet reached a consensus as to the specific sociodemographic variables that play a determining role in the relationship. The contribution of this research is that the instructional design of online courses can be grouped into five typologies. These course-types, when other variables are controlled, do not have a significant direct relationship with dropout propensity. However, the model proposed in this study indicates that certain student characteristics (e.g., family status or judgment of financial status) may become a risk factor in some instructional designs. It is important to note that the current study does not claim to have detected all of the pedagogical groupings of distance education courses, and in order to reach a more generalizable conclusion, a larger sample is needed. Nevertheless, this study offers an interesting avenue for explaining the inconsistency of sociodemographic factors in dropout studies.

References

Arifin, M. H. (2018). The role of student support services in enhancing student persistence in the Open University context: Lesson from Indonesia Open University. *Turkish Online Journal of Distance Education*, *19*(3), 156–168.

Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR, *Practical Assessment, Research, and Evaluation*, *19*(1), 1–15. https://doi.org/gk58v4

Bakker, E. J., Verhaegh K. J., Kox, J. H., van der Beek, A. J., Boot, C. R., Roelofs, P. D., & Francke, A. L. (2019). Late dropout from nursing education: An interview study of nursing students' experiences and reasons. *Nurse Education in Practice*, *39*, 17–25. https://doi.org/g4f5

Bennett, R. (2003). Determinants of undergraduate student drop out rates in a university business studies department. J*ournal of Further and Higher Education, 27*(2), 123–141. https://doi.org/br3mjj

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., Wallet, P. A., Fiset, M., & Huang, B. (2004). How does distance education compared with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research*, *74*(3), 379–439. https://doi.org/d8pd3p

Burns, M. (2013). Staying or leaving? Designing for persistence in an online educator training programme in Indonesia, *Open Learning: The Journal of Open, Distance and e-Learning*, *28*(2), 141–152. https://doi.org/ggx3rg

Cooper, D., & Schindler, P. (2011). *Business research methods* (11th ed.). McGraw-Hill International Edition.

Dolnicar, S. (2003). *Using cluster analysis for market segmentation-typical misconceptions, established methodological weaknesses and some recommendations for improvement*. Australia, Faculty of Commerce, University of Wollongong.

Flora, D. B., & Curran, P. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9(4), 466–491. https://doi.org/b6vmvn

Flynn, A. B. (2015). Structure and evaluation of flipped chemistry courses: Organic & spectroscopy, large and small, first to third year, English and French. *Chemistry Education Research and Practice*, *16*(2), 198–211. https://doi.org/gjk3h2

Franklin, S., Walker, C., & Statistics Canada. (2010). *Survey methods and practices*. Ministry of Industry. https://bit.ly/3qZpvWv

Garratt-Reed, D., Roberts, L. D., & Heritage, B. (2016). Grades, student satisfaction and retention in online and face-to-face introductory psychology units: A test of equivalency theory. *Frontiers in Psychology*, 7(673), 1–10. https://doi.org/f8kxbv

Garrison, D. R. (1987). Researching dropout in distance education. *Distance Education*, *8*(1), 95–101, https://doi.org/fhmc45

Hammond, D. E., & Shoemaker, C. (2014). Are there differences in academic and social integration of College of Agriculture Master's students in campus based, online and mixed programs? *NACTA Journal*, *58*(3), 180–188. https://bit.ly/32tbDcV

Jun, J. (2005). *Understanding dropout of adult learners in e-learning* [Doctoral dissertation, University of Georgia]. https://bit.ly/3qVwEqJ

Kember, D., Lai, T., Murphy, D., Siawet, I., & Yuen, K. S. (1992). Student progress in distance education: Identification of explanatory constructs. *British Journal of Psychology*, *62*, 285–298. https://doi.org/bhcn8r

Kember, D., Lai, T., Murphy, D., Siawet, I, & Yuen, K. S. (1994). Student progress in distance education courses: A replication study. *Adult Education Quarterly*, *45*(1), 286–301. https://doi.org/c7srnw

Jézégou, A. (2010). Se former à distance: Regard sur les stratégies d'autorégulation environnementale d'étudiants adultes. *Savoirs*, *24*(1), 79–99. https://doi.org/g63r

Kalet, A., Ellaway, R., Song, H., Nick, M., Sarpel, U., Hopkins, M., Hill, J., Plass, J., & Pusic, M. (2013). Factors influencing medical student attrition and their implications in a large multi-centre randomised education trial. *Advances in Health Sciences Education*, *18*(3), 439–450. https://doi.org/f46zm2

Lee, Y., & Choi, J. (2011). A review of online course dropout research: Implications for practice and future research. *Educational Technology Research and Development*, *59*(5), 593–618, https://doi.org/fws8x7

Leeds, E., Campbell, S., Baker, H., Ali, R., Brawley, D., & Crisp, J. (2013). The impact of student retention strategies: An empirical study. *International Journal of Management in Education*, *7*(1/2), 22–43. https://doi.org/gqvc

Li, Y., Chen, H., & Han, Y. (2010). Why do they drop out of school? Interview survey of learners in open and distance education. *Open Education Research*, *2*, 1–11. https://bit.ly/3cEdvRD

Li, Y., Niu, J., & Ding, X. (2012). A Follow-up study of the dropouts from the English program of open and distance learning (part 2). *Open Education Research*, *18*(6), 80–86. https://bit.ly/3DFT8zq

Marcus, M., & Minc, H. (1988). *Introduction to linear algebra*. Dover Publications.

Monteiro, S., Lencastre, J. A., Silva, B. D., Osório, A. J., de Waal, P., İlin, S. Ç., & İlin, G. (2017). A systematic review of design factors to prevent attrition and dropout in e-learning courses. In G. İlin, S. C. İlin, B. D. da Silva, A. J. Osório, & J. A. Lencastre (Eds.), *Better e-learning for innovation in education* (pp. 135-153). Istanbul: ÖzKaracan. http://hdl.handle.net/1822/47770

Nichols, M. (2010). Student perceptions of support services and the influence of targeted interventions on retention in distance education. *Distance Education*, *31*(1), 93–113. https://doi.org/cxbxbr

No, F., & Hirakawa, Y. (2012). Identifying causes of dropout through longitudinal quantitative analysis in rural Cambodian basic schools. *Journal of International Development and Cooperation*, *19*(1), 25–39, http://doi.org/10.15027/33828

Packham, G., Jones, G., Miller, G., & Thomas, B. (2004). E-learning and retention: Key factors influencing student withdrawal. *Education and Training*, *46*(6/7), 335–342. https://doi.org/brq8xr

Pittenger, A., & Doering, A. (2010). Influence of motivational design on completion rates in online self-study pharmacy-content courses. *Distance Education*, *31*(3), 75–293. https://doi.org/dh3x33

Ramsdal, G. H., Bergvik. S., & Wynn, R. (2018). Long-term dropout from school and work and mental health in young adults in Norway: A qualitative interview-based study. *Cogent Psychology*, *5*(1), 1–16, https://doi.org/g4f3

Rekkedal, T. (1985). *Introducing the personal tutor/counsellor in the system of distance education. Project Report 1: Experiment Description*. ED235774. https://eric.ed.gov/?id=ED235774

Robinia, K. J., Maas, N. A., Johnson, M. M., & Nye, R. M. (2012). Program outcomes following implementation of a hybrid curriculum at the certificate level. *Nursing Education Perspectives*, *33*(6), 374–377. https://bit.ly/3CQQPIP

Sauvé, L., Fortin, A., Landry, F., Viger, C., Côté, S., Bégin, S., & Mupalanga. I. (2014). *La persévérance et la réussite universitaires d'étudiants inscrits à des programmes de premier cycle en sciences comptables au Québec. Rapport de recension*. Société pour l'Apprentissage à Vie (SAVIE). https://bit.ly/3CE8fs0

Tan, P. N., Steinbachet, M., & Kumar, V. (2013). Data mining cluster analysis: basic concepts and algorithms. In P.-N. Tan, M. Steinbach, A. Karpatne, & V. Kumar (Eds.), *Introduction to data mining* (pp. 487-533). Pearson Publishing.

Tan, M., & Shao, P. (2015). Prediction of student dropout in e-learning program through the use of machine learning method. *International Journal of Emerging Technologies in Learning*, *10*(1), 11–17. https://doi.org/gmd8x3

Willging, P. A., & Johnson, S. D. (2009). Factors that influence students' decision to dropout of online courses. *Journal of Asynchronous Learning Networks*, *13*(3), 115–127. https://bit.ly/3oI6HYY

Yasmin, D. (2013). Application of the classification tree model in predicting learner dropout behaviour in open and distance learning, *Distance Education*, *34*(2), 218–231. https://doi.org/gf278z