THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Using institutional data to predict student course selections in higher education

Citation for published version:
Ognjanovic, I, Gasevic, D & Dawson, S 2016, 'Using institutional data to predict student course selections in higher education' Internet and Higher Education, vol. 29, pp. 49-62. DOI: 10.1016/j.iheduc.2015.12.002

Digital Object Identifier (DOI):
10.1016/j.iheduc.2015.12.002

Published In:
Internet and Higher Education

Using institutional data to predict student course selections in higher education

Ivana Ognjanovic[1], Dragan Gasevic[21], Shane Dawson[3]

[1]Mediterranean University, Montenegro
[2]Moray House School of Education and School of Informatics,
University of Edinburgh, United Kingdom
[3]University of South Australia, Australia
ivana.ognjanovic.edu@gmail.com, dgasevic@acm.org, shane.dawson@unisa.edu.au

Abstract.

The ability to predict what university course a student may select has important quality assurance and economic imperatives. The capacity to determine future course load and student interests provides for increased accuracy in the allocation of resources including curriculum and learning support and career counselling services. Prior research in data mining has identified several models that can be applied to predict course selection based on the data residing in institutional information systems. However, these models only aim to predict the total number of students that may potentially enrol in a course. This prior work has not examined the prediction of the course enrolments with respect to the specific academic term and year in which the students will take those courses in the future. Moreover, these prior models operate under the assumption that all data stored within institutional information systems can be directly associated with an individual student's identity. This association with student identity is not always feasible due to government regulations (e.g., student evaluations of teaching and courses). In this paper, we propose an approach for extracting student preferences from sources available in institutional student information systems. The extracted preferences are analyzed using the Analytical Hierarchy Process (AHP), to predict student course selection. The AHP-based approach was validated on a dataset collected in an undergraduate degree program at a Canadian research-intensive university (N=1061). The results demonstrate that the accuracy of the student course predictions was high and equivalent to that of previous data mining approaches using fully identifiable data. The findings suggest that a students' grade point average relative to the grades of the courses they are considering for enrolment was the most important factor in determining future course selections. This finding is consistent with theories of modern counselling psychology that acknowledges self-efficacy as a critical factor in career planning.

Keywords: course enrollment prediction, decision science, higher education, learning analytics, student information systems

1 Introduction

Contemporary higher education is confronted with numerous challenges stemming from increases in student numbers and diversity, alongside a global competitive education market and significant reductions in government funding (Srivastava, Gendy, Narayanan, Arun, & Singh, 2012). These pressures have resulted in universities re-thinking how education is best delivered and supported. This is well evidenced in the rise of online learning programs and Massive Open Online Courses (MOOCs) (Allen & Seaman, 2013). Even in this climate of

[1] Corresponding author. Email: dgasevic@acm.org; Phone: +44 131 651 6243; Fax: +44 131 650 4587; Address: School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh, Midlothian EH8 9LE, United Kingdom

economic austerity, changing education models and competition, education providers still need to continue to maintain efficient yet, effective and personalised student support services (e.g. curriculum support, learning support and career counselling). The ability to predict student enrolment trends for courses and programs provides an opportunity for an institution to not only effectively and efficiently allocate student and learning support resources (Marginson, 2013), but also to establish a higher quality learning experience and to better understand trends in program demand. However, the rhetoric of accurately predicting student course selections is far easier than the practice.

Although complex, accurate predictions regarding a student's future course selection can be developed through the analysis of student data extracted from institutional information systems (e.g. students' grades, course evaluations, and basic demographics). This has been well demonstrated in a number of research studies in educational data mining (Baker, 2010; Romero & Ventura, 2010). However, the predictive models derived from these studies tend to assume that all data stored within institutional information systems can be directly associated with an individual student's identity. This association with student identity is not always feasible as government and institutional policies can impose strict regulations to ensure privacy and anonymity. An example of such regulations can be found in the student evaluations of teaching and courses. In such cases indiscriminate use of data mining techniques cannot always be employed when student privacy must be protected (Vialardi, Bravo, Shafti, & Ortigosa, 2009). Yet, course and teacher evaluations do provide an important source of information that may influence a student's future course selections. Thus, there remains an imperative to attempt to incorporate these data into the development of a predictive model (MacFadyen, Dawson, Prest, & Gašević, in press; Kardan, Sadeghi, Ghidary, & Sani, 2013). To address this issue, we propose and test a novel approach for identifying a predictive model of student course selections drawing on identifiable and non-identifiable data. The approach is based on the Analytic Hierarchy Process (AHP) (Saaty, 1980), a well-established decision making technique for dealing with multi-dimensional and often contradictory preferences of individuals. As such, the study addresses four related goals. The study firstly aims to develop a predictive model of student course selections across a program of study. A second aim was to analyze the accuracy of the proposed AHP-based approach when applying different weights to the extracted preferences. The research also assesses the accuracy of the approach to predict the total number of students that may enrol in each course within a selected program of study. Finally, the study aims to compare the effectiveness of the findings with prior research. The study delimits these aims into the following four questions.

Research Question 1 assesses the overall accuracy of the predictive model:

- RQ1a. What is the accuracy of the predicted number of student enrollments for each course offering?
- RQ1b. What is the impact of each individual concern on the accuracy of the predicted number of student enrollments for each course offering?

Question 1 (RQ1a and RQ1b) addresses the need to validate the accuracy of the proposed approach against related published work – e.g. (Kardan, Sadeghi, Ghidary, & Sani, 2013). Specifically, RQ1b extends the current existing body of research to investigate the impact of each individual concern (i.e., the individual types of data stored in institutional information systems) on the accuracy of the predictions. For both questions, we have focused on the predictions of the number of enrollments for individual course offerings per semester and term in different academic years. The anticipated results from the research question may inform the institution of the number of enrollments for each course offering, and therefore informing the level of support and resource allocations.

Research Question 2 examines the accuracy of the prediction for exact course selections undertaken by each individual student per term/semester.

- RQ2a. What is the accuracy of the predictions of course selections for each individual student per semester/term?
- RQ2b. What is the impact of each individual concern on the accuracy of the predictions of course selections for each individual student per semester/term?

Research Question 3 assesses the accuracy of the prediction for exact course selections of each individual student per academic year.

- RQ3a. What is the accuracy of the predictions of course selections for each individual student per academic year?
- RQ3b. What is the impact of each individual concern on the accuracy of the predictions of course selections for each individual student academic year?

Questions 2 and 3 are focused at the individual student level. For each student, the accuracy of the approach to predict course selection in each semester/term and academic year is evaluated. These questions provide useful insights into how the model can potentially predict the future study plans for an individual student. To our knowledge, no study has been undertaken that has developed and reported an approach to predict course selections at the granularity of an individual student and the level of specific academic terms or years in which they may enrol.

Research Question 4 aims to compare the effectiveness of its contribution with respect to the published literature and cited methodologies.

- How does the proposed AHP method perform in respect to other adopted methods within the extant literature?

As the prior literature did not reveal any specific results related to research questions 2-3 (only research question 1), we compare the effectiveness of the AHP approach with a noted alternate method used to address the problem raised in research question 1.

In a broader sense, the study contributes to the body of research knowledge in the field of academic analytics (Campbell, DeBlois, & Oblinger, 2007). While the proposed research can be connected with the field of educational data mining and learning analytics, the questions and method adopted generates results and findings that are more applicable for institutional decision makers to aid their forecasting and resource planning.

2 Theoretical Background

2.1 Approaches to Modeling Preferences and Predicting Course Selection

There are numerous examples in the research literature illustrating the use of data mining techniques in order to develop course recommendation systems (Hsia, Shie, & Chen, 2008; Castellano & Martinez, 2008). For instance, Vialardi and colleagues (2009; 2010; 2011) have over a number of years, analysed historical student data such as course enrolments, course load, and academic performance to determine an individual's probability of success in a selected course. The intention of the research is to optimise the likelihood of an individual student's success through the provision of course recommendations. In terms of accuracy of the applied machine learning algorithms, Vialardi et al (2011) noted that the decision tree C4.5 outperformed both Naïve Bayes and KNN in predicting student success. Furthermore, the authors reported that the performance of their predictive model significantly improved (accuracy ~82%) when so-called syntactic attributes (i.e., potential for success and difficulty to complete the course successfully) are included with variables from the student information systems (e.g., number of course attempts, grade point average, course credits, number of credits, and final grade). While the work in this area has clearly demonstrated the potential for machine learning to provide course recommendations, there are data that could be used to further supplement this approach. Variables related to course and teacher evaluations, for example, are seldom incorporated despite their importance in influencing student performance and course selection (Paechter, Maier, & Macher, 2010; Sun, Tsai, Finger, Chen, & Yeh, 2008; Arbaugh & Duray, 2002).

While there has been much research in education focused on developing course recommender systems there has been comparatively less work examining the extraction and modeling of students' preferences to build decision support systems (Romero & Ventura, 2010). Decision support techniques have been long adopted and utilised in fields such as medical decision making (Berg, 1997; Sim, et al., 2001). However, the application of decision support techniques in education is only recently receiving attention (Romero, Ventura, Pechenizkiy, & Baker, 2010; Baker, 2010). Bala and Ojha (2012) argued that a core challenge facing the education sector is how to better utilise the vast amounts of captured student data to inform and improve managerial decision making. The application of decision support techniques provides a sound approach for addressing this challenge.

The decision science literature posits that decisions are inherently complex and influenced by an individual's expectations, attitudes and preferences. There is no one technique that can effectively model such a process. Hence, the selection of the most appropriate method for modeling and processing user preferences is based on the specific needs of the problem under investigation. For this present study, we adopted a framework based on the Analytical Hierarchical Process (AHP) as reported by Ognjanović et al. (2013). This framework was adopted as it can be used to model different preference structures of importance, and rank those preferences in terms of their importance, for student decision making in their course selection (Brafman & Domshlak, 2002; Wilson, 2011). The adopted framework is based on Saaty's (1980) well established Analytical Hierarchical Process.

AHP as proposed by Saaty (1980) is a widely adopted multi-criteria decision making method that can assist in organizing and analyzing complex decisions (Chen & Wang, 2010). AHP enables decision making parties to deal with both tangible and intangible options and to monitor the degree of consistency in the judgments of the involved parties (Roper-Low, 1990). To date, AHP has predominantly been adopted in the important decision making domains such as forecasting, quality management, business process management, quality function deployment, and performance management (Chen & Wang, 2010; Forman & Gass, 2001). However, the structured technique has also been applied in education (Yüksel, 2012). For example, AHP has been used for the analysis of teaching quality (Liu'an, Xiaomei, & Lin, 2012) and the evaluation of educational effectiveness (Yüksel, 2012), based on hierarchical models of influencing criteria. The adopted AHP framework for predicting student course selection is summarized in Figure 1 and includes:

i) extracting the factors influencing course selection from institutional information systems, representing the extracted factors in a form suitable for processing by AHP, and assigning the values of the extracted factors to courses (Section 3); and
ii) gauging course selection preferences for each individual student based on the extracted factors (Section 2.2) in order to make predictions about their course selections as an AHP-based ranking of available courses (Section 3).

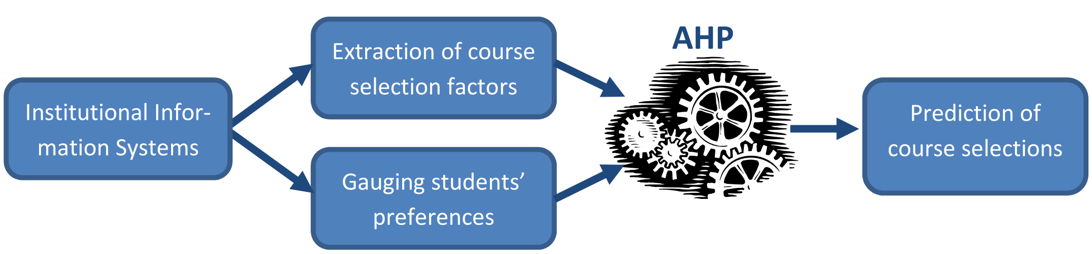

Figure 1. Overview of the adopted AHP framework. Illustrating the extraction of course selection factors and gauging student preferences for predicting course selections as an AHP-based ranking of available courses.

2.2 Extracting Factors influencing Students' Course Selection

This section outlines the factors that influence student course selection to inform a model for extraction from student information systems in order to transform the data into a suitable format for processing by the AHP based algorithm. Incoporation of the suite of factors was based on the available data and prior education literature. The values of the extracted factors are then assigned to courses offered by an institution, so that the predictions for course enrollments are made based on students' preferences.

2.2.1 Sources of Data

The student information system was the primary source of data for developing the predictive model. These systems store a wealth of data relating to a student's prior course enrolments, cumulative grade point average (GPA), as well as demographic data (e.g., gender and country of origin), potential career objectives (e.g., subject area of specialisation, major and minor), course scheduling, instructor demographics and course and teacher evaluations. In accordance with AHP, the students' preferences are defined as the relative importance between the quality characteristics that represent the important matters of interest (hereinafter *concerns*), and between the possible values of the concerns (hereinafter *qualifier tags*) (Ognjanović, Gašević, & Bagheri, 2013). For the present study, this would result in *concerns* such as disciplinary specialization and class scheduling. As such, the *qualifier tags* for the concern class scheduling could be early morning classes, mid-day classes, and evening classes. The hierarchical model adopted in the study (Figure 2) contains two main groups of concerns: (i) *course/programme factors* and (ii) *individual factors* (including both students' personal preferences and environmental factors). Based on the reviewed literature (referenced in the reminder of the following subsections), the groupings comprised a set of sub-concerns that were also included in the model. Either of the two groups of concerns is decomposed into three *sub-concerns*, and these concerns have their own specific *qualifier tags.* Details about each of these concerns and their qualifier tags are discussed in the following subsections and summarized in Table 1.

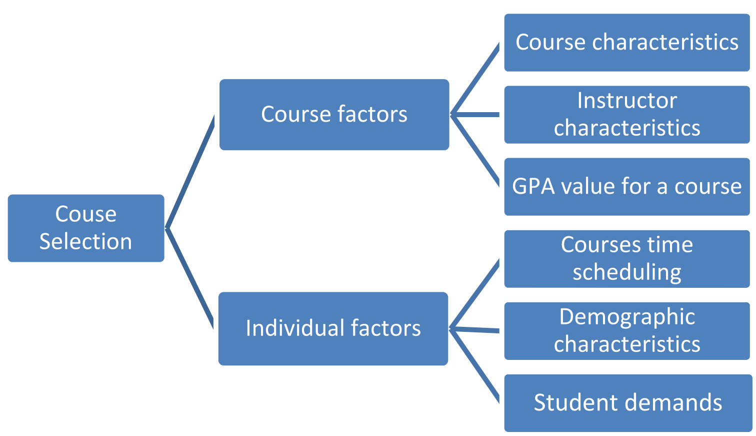

Figure 2. Hierarchical structure of concerns important for courses selection

2.2.2 Course Characteristics

Student evaluations of teaching (teachers and courses) are common occurrences in the higher education sector (Marsh, 2007; Spooren, Brockx, & Mortelmans, 2013). The vast majority of universities have in place online systems to administer evaluation surveys, and report results for all courses offered in an academic year. The evaluation surveys contain a series of items composed as statements that are ranked on a Likert-scale, which typically have five-points (strongly agree to strongly disagree). In addition, each item can be associated with a weight indicating their degree of importance in the entire survey (or each item may be considered of equal importance, and thus equally weighted). Since the arithmetic mean value is falling into exactly one of subintervals – e.g., for a five-point scale those would be [1,2), [2,3), [3,4), [4,5] – the set of qualifier tags is defined accordingly in our approach for the *course characteristics* concern. However, the arithmetic mean value is not a sufficiently good descriptive parameter. In fact, confidence intervals provide more meaningful information about the dissipation of values around the mean value (Blaikie, 2003). Thus, the score of a *k*-th course is an interval

$$\textit{Course Score Interval}_{k}= \left(\frac{1}{N}\sum_{j=1}^{N}\frac{\sum_{i=1}^{M_k} W_i C_{ij}}{\sum_{i=1}^{M_k} W_i} - \mathrm{z}\cdot\frac{\sigma_{\mathrm{k}}}{\sqrt{\mathrm{N}}}, \frac{1}{N}\sum_{j=1}^{N}\frac{\sum_{i=1}^{M_k} W_i C_{ij}}{\sum_{i=1}^{M_k} W_i} + \mathrm{z}\cdot\frac{\sigma_{\mathrm{k}}}{\sqrt{\mathrm{N}}}\right)$$

where W_i is the weight of *i*-th question in the evaluation survey, C_{ij} denotes the response of *j*-th student to *i*-th question, M_k is the number of questions about course characteristics in the survey, *N* is the total number of students participated in the survey, σ_k is the standard deviation value for *the k*-th course and *z*-score is defined by the required confidence interval.

For example, let us consider that the mean value of a course evaluation is 3.9 with a standard deviation of 0.49 on a five-point Likert scale question. A 95% confidence interval (with critical value z=1.96) for the mean value of the course characteristics course is *((3.9-1.96*0.49), (3.9-1.96*0.49))=(2.93, 4.86).* Hence, the evaluation results for each particular student in this example would fall into the interval (2.93, 4.86). As such, the course is annotated with qualifier tags *qt2* and *qt4* from Table 1 (as subintervals containing the lower and upper bounds of the course score interval).

The proposed model can support more than one course score. In fact, the model can use as many survey items as deemed necessary based on the data availability in a given institution. Typically, one of the main influencing factors for the first course selection is related to its perceived usefulness or value (Kardan, Sadeghi, Ghidary, & Sani, 2013). The item measuring the *'overall course'* value in an evaluation survey can be interpreted as the major summative evaluation indicator of the course (Babad & Tayeb, 2003). According to Marsh's (2007) dimensions of students' evaluations of university teaching, the course evaluations are an integral part of the learning value cluster and therefore, have been incorporated into the present study (as discussed in Sections 3-4).

Within the broader learning and teaching context, some student and course data are non-identifiable. Government privacy protection policies and legislation in various countries (e.g., Canada) often require an institution to ensure student anonymity in relation to submitted course evaluations. This can prohibit capacity to connect course evaluations with an individual student's identity. In this context, it is only possible to record if the student has or has not submitted an evaluation – there are no records of the quality or score of the evaluation submission. The absence of such data obviously limits how students' preferences concerning the specific characteristics of courses are gauged and requires alternative approaches, to be developed (detailed in Section 3.2).

2.2.3 Instructor Characteristics

Over the past several decades there has been much research undertaken to determine the association between students' evaluation scores and effective teaching (Centra, 2003; Babad & Tayeb, 2003; Marsh & Roche, 2000; Spooren, Brockx, & Mortelmans, 2013; MacFadyen, Dawson, Prest, & Gašević, in press). While there continues to be debate regarding the legitimacy of course and teacher evaluations, Marsh's (2007) well noted work in this area demonstrates that student evaluations are reliable measures of teaching effectiveness. As such, we would surmise that a student's negative or positive course experience with an instructor (teacher) will ultimately influence their decision to enrol in future courses associated with an individual instructor or even disciplinary field. In this context, the data regarding students' perceived "overall satisfaction" with the instructor are included in the study in order to calculate the score of the *l*-th instructor over the *k*-th course as a confidence interval. The interval was defined in a similar way to the confidence intervals associated with the course characteristics (Sect. 2.2.2). Thus, the intervals for a *k*-th course are calculated separately for all instructors who were involved in teaching the course. The set of qualifier tags (see Table 1) is defined in a similar way as described for course characteristics in the prior subsection.

2.2.4 Grade Point Average Value for a Course

Several studies have shown that a positive relationship exists between a student's academic performance in a course and their overall evaluations (Marshall, Greenberg, & Machun, 2012; Svanum & Aigner, 2011; Greenwald & Gillmore, 1997). Based on the work of Marsh and Roche (2000), the GPA value for each student defines their overall learning effectiveness. As Centra (2003) observed, the learning effectiveness, represented as a course grade, is influenced by the student's level of interest and motivation in the particular course. Hence, the data concerning overall Grade Point Average (GPA) for all enrolled students in a course is a necessary inclusion for the present study. All students enrolled in a specific course are divided into groups based on their GPA. The numbers of students in each group are then calculated based on the score of *i*-th course defined as a quadruplet

$$GPA\text{-}course_i=(q_1, q_2, q_3, q_4)$$

where q_j, $j=1,..,4$ is the number of students enrolled to *i*-th course with GPA falling into GPA ranges [60-70), [70-80), [80-90), and [90-100], respectively. The set of qualifier tags is defined based on this division of GPA values on subintervals [60-70), [70-80), [80-90), [90-100]. Those subintervals correspond to distribution of letter grades A-, A, A+ (over [90-100]), B-, B, B+ (over [80-90)), C-, C, C+ (over [70-80)), and D+, D (over [60-70)), commonly used in many universities. Qualifier tags provide good insight into the distribution of students' interests in taking the course with respect to their previous academic success, i.e., it is a detailed snapshot of the students who take the course and how they have performed on previous course assessments.

2.2.5 Scheduling Time of Courses

Individual preferences concerning course scheduling are of importance when making decisions about course selection (Baker, 2010). For example, students may attempt to avoid particular times as a result of personal conflicts such as part-time work (Schuhmann & McGoldrick, 1999). Similarly, scheduling issues can arise as a result of the time necessary to travel to and from class. Course scheduling preferences could be elicited from surveys. However, an alternate automated approach can be determined from analysis of the scheduling of elective courses a student has previously enrolled in. To this end, we divided a period commonly used by many universities for scheduling classes, from 8:00am to 8:00pm, into the following time intervals: [8:00-10:00), [10:00-12:00), [12:00-14:00), [14:00-16:00), [16:00-18:00], [18:00-20:00]. These time intervals therefore, define a set of qualifier tags. Each course is annotated with two of these six intervals denoting both start and end of class.

2.2.6 Demographic characteristics

Several studies have reported a relationship between student characteristics and decisions concerning their learning and study interests (Dutton, Dutton, & Perry, 2002; Qureshi, Morton, & Antosz, 2002; Stewart, Bachman, & Johnson, 2010). To include demographic characteristics in the AHP-based model, we took into consideration the following:

- The analysis of differences between possible values for each particular demographic characteristic, stored in a student information system, is performed and super-groups are created by merging groups with no difference;
- A Cartesian product of the created groups is then used as a basis for creating qualifier tags of the demographics considered (e.g., if gender and domestic/international students are characteristics with significant differences, the following groups are created: domestic–male, domestic–female, international–male, and international–female). The numbers of students falling into each group are subsequently calculated.

Thus, the score of demographic characteristics with *k* significant groups is a *k*-tuple

$$\textit{Demographic characteristics score}_i = (c_1, c_2, \ldots, c_k)$$

where c_k is the number of students previously enrolled in an i-th course with the demographic characteristics specified within the k-th group. These values outline the distribution of the students' characteristics depending on their prior course selections. For instance, using this approach, the number of international female students interested in a particular course can be demonstrated.

2.2.7 Student Demands

A students' interest in a course can be influenced by numerous factors such as the subject area, academic ability and skill, and identified career goals (Babad, Darley, & Kaplowitz, 1999). These criteria can determine the level of personal interest for each student and can be extracted from existing (course evaluation) surveys available in the institutional information systems. It is reasonable to assume that a student's implicit expression of interest in a subject area can be obtained through the analysis of elective courses he/she has previously engaged in. The student's interests can be extracted in a similar way as previously shown for the extraction of preferences over scheduling time for specific courses. Alternatively, the student may want to change their subject area. In such cases, the student should be asked to explicitly define his/her interest or declare indecision and ambiguity (in terms of the relative importance between concerns/qualified tags as requested by AHP).

2.2.8 Hard constraints in Courses Selection Process: Program and Institutional Rules and Requirements

While course selection is a decision making process influenced by an individual's personal and academic interests and characteristics, there are also specific institutional rules and requirements for a given program of study. The present study sought to incorporate:

(i) institutional rules concerning majors in programs (e.g., foundational, core, and elective courses) and defined prerequisites;
(ii) the number of required credits for each term/semester and academic year to maintain a specific status (i.e., full or part-time);
(iii) the maximum number of students that can be enrolled in a given course. In such cases, institutions may give a priority for course selection to students with higher academic achievement (i.e. higher GPA); and
(iv) where the overlap between course offerings is not permitted.

Most existing approaches examining the problem of predicting course selection, consider the program and institutional rules to be of the same nature as all other factors (Baker, Corbett, & Aleven, 2008; Kardan, Sadeghi, Ghidary, & Sani, 2013). However, if course selection is modeled as a decision making process (Babad, 2001) that is required to adhere to the program and institutional rules, there is a need to consider these rules as hard constraints in the AHP-based course selection method proposed in this paper. Therefore, in order to obtain a feasible set of courses, we should remove any combination of courses that violate an established hard constraint.

2.2.9 Qualifier tags for Two-layered Structure

A two-layered structure of concerns and qualifier tags (outlined in Figure 2), is summarized in Table 1. The concerns and qualifier tags should be used as a basis for extracting students' preferences, defining the measurement of those preferences, and computing an optimal set of courses gauged as the most appropriate for each student.

Table 1. Concerns and qualifier tags

Concern	***Qualifier tags / values***				
Course characteristics	***qt1***	***qt2***	***qt3***	***qt4***	
	[1,2)	[2,3)	[3,4)	[4,5]	

Instructor characteristics	***qt1***	***qt2***	***qt3***	***qt4***		
	[1,2)	[2,3)	[3,4)	[4,5]		
GPA value for Course	***qt1***	***qt2***	***qt3***	***qt4***		
	[60-70)	[70-80)	[80-90)	[90-100]		
Course time scheduling	***qt1,1- beginning hours***	***qt1,2- beginning hours***	***qt1,3- beginning hours***	***qt1,4- beginning hours***	***qt1,5- beginning hours***	***qt1,6- beginning hours***
	qt2,1- finishing hours	***qt2,2- finishing hours***	***qt2,3- finishing hours***	***qt2,4- finishing hours***	***qt2,5- finishing hours***	***qt2,6- finishing hours***
	[8:00-10:00)	[10:00-12:00)	[12:00-14:00)	[14:00-16:00)	[16:00-18:00]	[18:00-20:00]
Demographic characteristics	***qt1***	***qt2***	***…***	***qtn***		
	Depending on questionnaire contents					
Student demands	***qt1***	***qt2***	***…***	***qtm***		
	Depending on questionnaire contents					

3 Analytical Hierarchy Process for Course Selection

This section describes our adaptation of the AHP (Saaty, 1980) over the two-layered structure of factors influencing student course selection (c.f. Table 1 and Section 3.1). The outcomes of AHP (Section 3.3) are the ranks over the set of available courses that best fit the measured level to students' preferences extracted from the institutional information systems (Section 3.2).

3.1 Overview of Analytical Hierarchy Process

In order to use AHP, the relative importance of each of the available criteria (i.e., concerns and qualifier tags) compared to others was determined. Relative importance is typically defined with odd numbers ranging from 1 (equal importance) to 9 (extreme importance of one over the other). That is, in the *concern prioritization step*, the relative importance of each concern $\{c_1,\ldots,c_n\}$ with respect to the others is defined by the stakeholders. The concerns are compared in a pair-wise way, and the relative priorities $\{r_{c_1},\ldots,r_{c_n}\}$ are calculated for each of them, defining their ranks. The ranking of available options (i.e., course in our case) are then formed. The options (i.e., courses) $\{o_1,\ldots,o_n\}$ available to the students are also associated with qualifier tags, $o_j = \left\langle qt^1_{j_1},\ldots,qt^m_{j_m}\right\rangle, 1 \le j \le n$, as shown in Section 2.2. During the course ranking process, to establish the actual priority and importance of the available courses, the relative importance of the qualifier tags for each concern is computed by performing AHP, assigning them $\{r^1_{qt_1},\ldots,r^1_{qt_{|QT_1|}}\},\ldots,\{r^n_{qt_1},\ldots,r^n_{qt_{|QT_n|}}\}$, which are the ranks of qualifier tags of the *1*st*,…, n*th concern, respectively. Afterwards, the rank of each course is determined based on the ranks of the qualifier tags that are associated with the courses. That is $r\left(\left\langle qt^1_{j_1},\ldots,qt^m_{j_m}\right\rangle\right) = f\left(r_{c_1}\cdot r^1_{qt_{j_1}},\ldots,r_{c_m}\cdot r^m_{qt_{jm}}\right), 1 \le j \le n$, where *f* is a predefined function (i.e., minimum, maximum, or mean). The goal of this stage is to assign higher ranks to the courses which are related to the student's most important concerns.

The AHP approach in this paper is applied to develop a predictive model – which makes use of the existing data from student information systems – to address the tasks introduced in research questions in Section 1. Therefore, expressions of relative importance – as each student's preferences – are extracted from data as described in Section 3.2.

3.2 Extraction of Students' Preferences as Relative Importance over the Two-layered Concern Structure

To ascertain the degree to which a course is appropriate for a particular student, a pairwise comparison between each pair of concerns and each pair of qualifier tags per each concern is required. This provides the essential input values for the AHP algorithm. In the present study, we propose extracting student preferences, (as described in Figure 2 and Table 1), from the institutional information systems in lieu of asking individual students to express their personal judgement.

3.2.1 Preferences over Characteristics of Courses and Instructors

Course and instructor evaluation score intervals are based on the standardized mean of the weighted average of course and instructors characteristics. Since those intervals usually do not correspond to the set of qualifier tags, the following steps are considered in order to obtain their quantitative measurement:

(i) Qualifier tags are compared based on the assumptions that a course with the highest evaluation score (i.e. qualifier tag [4, 5]) is the most preferable. Higher course evaluation intervals are more preferable (e.g., course evaluations in interval [3-4) are more preferred than scores in interval [2-3)). Thus, the ranks of qualifier tags about characteristics of courses and instructors can be determined by the standard AHP calculations (see Table A1 in Appendix A) by giving the following rank values: $r_1 = 0.06$, $r_2 = 0.12$, $r_3 = 0.26$, $r_4 = 0.56$ for grade ranges (i.e., qualifier tags) [1-2), [2-3), [3-4), and [4-5], respectively.

(ii) Since each course (and instructor) is annotated (see Section 2.2.2) with a score interval [*lower_bound, upper_bound*] for the course characteristics concerned and where the lower and upper bound values fall into different grade ranges (i.e., qualifier tags), the continuous measurement is used to generate ranks based on the specified rank values for grade ranges [1-2), [2-3), [3-4), and [4-5]. The continuous measurements are defined by constructing a polynomial measure over the obtained ranks of qualifier tags (Ognjanović, Mohabbati, Gašević, Bagheri, & Bošković, 2012), and is used for the calculation of the ranks for bound values with:

$$r(x) = r_{j-1} + \frac{x - l_{j-1}}{u_{j-1} - l_{j-1}} * \left(r_j - r_{j-1}\right) = r_{j-1} + \left(x - l_{j-1}\right) * \left(r_j - r_{j-1}\right)$$

where $x \in \lfloor l_j, u_j \rfloor$, and r_j is the measure of interval $\lfloor l_j, u_j \rfloor$ to which grade x belongs,

on the basis of which, the rank value for the interval is calculated:

$$r(lower\_bound, upper\_bound) = (r(upper\_bound) + r(lower\_bound))/2.$$

For example, consider a course with the score interval *(2.93, 4.86).* The ranks for the lower and upper bounds are calculated as: *r(2.93)=0.06+(2.93-2)*(0.12-0.06)=0.11* and *r(4.86)=0.26+(4.86-4)*(0.56-0.26)=0.52*. Finally, the rank for the score interval is *r(2.93, 4.86)=(0.11+0.52)/2=0.315.* It is also interesting to compare the obtained value with the rank value for the course average score *r(3.9)=0.12+(3.9-3)*(0.26-0.12)=0.246 < r(2.93, 4.86)*, quantifying the necessity for considering the confidence interval for course annotation and measurement.

When a course is offered for the first time, there is no evaluation available to compute the score of the course. In such cases, AHP allows for defining the so-called 'unknown' preference, which has the rank of value *r=0.25* (Ognjanović, Gašević, & Bagheri, 2013) (see Table A2 in Appendix A).

3.2.2 Preferences over GPA Value for a Course

The *GPA value for a course* describes the academic range of students previously enrolled in a course. The following steps attempt to measure the relative importance of academic ability (i.e. GPA) for a given student in comparison to other enrolled students when selecting a course:

(i) Let a_1, a_2, a_3, and a_4 be a frequency of the students with GPA groups falling into subintervals (i.e. qualifier tags) [60-70), [70-80), [80-90), and [90-100], respectively.
(ii) Ranking of qualifier tag a_s which is the subinterval for the GPA value of the given student is defined with:

$$r(a_s) = \frac{\sum_{1 \le i \le s} a_i}{\sum_{i=1}^{4} a_i} = \frac{\sum_{1 \le i \le s} a_i}{100}.$$

For instance, let us consider "Student A" with a GPA of 73% (i.e. the student's GPA interval is [70-80]). If the distribution of the GPA intervals for an Economics course is a_1=15%, a_2=30%, a_3 =35% and a_4=20%, the rank for Student A to enrol in the course is (15+30)/100=0.45. The value of 0.45 can be interpreted that the course is more preferable for students with better grades, but the student under consideration is academically sufficiently close and thus, can also select the course.

3.2.3 Preferences over the Scheduling Time for a Course and other Personal Preferences

As discussed in Section 2.2.5, each course is annotated with qualifier tags representing the earliest beginning and latest ending of the classes. Preferences for each individual student over the scheduling time characteristics can be extracted, (as per Section 3.2.2) by defining the preference of a *s*-th student as a pair of six-tuples.

$$\textit{Course-scheduling score}_s=((b_1, b_2, b_3, b_4, b_5, b_6),(f_1, f_2, f_3, f_4, f_5, f_6))$$

where b_i, $i=1,\ldots,6$ and f_i, $i=1,\ldots,6$ are the numbers of elective courses that the *s*-th student have already selected with the earliest beginning and the latest ending in an *i*-th interval. Considering that each beginning time after the lower bound and each ending time prior to the upper bound are acceptable for the student, the measurement for the students' preferences over the scheduling time is created by using the historical data for all elective courses that the student has previously enrolled in. Calculations are analogous to those for GPA values (Section 3.2.2.) and the mean values for both criteria (beginning and ending hours) are used as rank values for the scheduling time concern. Similar considerations to those for course scheduling times are used for measuring students' preferences over other concerns (i.e., demographic characteristics and student demands).

3.3 Predicting Course Enrollment Sequences

Once the student preferences have been extracted (see section 3.2), predicting the selection of a set of courses for each student (per semester and/or per academic year), is determined through the following steps:

(i) AHP generates the level of suitability of each course, based on the extracted preferences for a given student (Section 3.2);
(ii) From the AHP calculations, different combinations of available courses are generated by simultaneously checking the level of satisfaction of the hard constraints and by maximizing the overall suitability of the selected course with respect to the student's preferences.

4 Method

4.1 Context

This study was conducted on a dataset derived from student course enrolments in a Bachelor of Arts (BA) degree in Psychology at a research-intensive university in Canada. The program investigated comprised some 47 courses offered across two semesters – winter and summer. The main academic year is completed in the winter session over two academic terms. Although the summer session is based on a two term model, the overall term duration is shorter in comparison to the winter semester. The courses offered during the summer period are generally less well attended and the quantity and diversity of available courses are reduced. Although the data centers on a psychology program, the students are encouraged to select additional courses offered in alternate departments and faculties in the university. Furthermore, through inter-institutional agreements, students were permitted to undertake courses offered by other Canadian universities. Each course was worth a specific number of credits with a typical one-semester long course being worth three (and in rare cases four) credits. Adding further complexity to the data, a subset of courses offered the option for students to enrol in a reduced number of sections (i.e. partial course completion). Upon completion of these sections students are awarded a pro-rated number of credits. In order to fulfill the requirements for the BA degree psychology, students had to complete a set number of credits from the following six groups of courses:1) Writing and Research intensive courses, 2) Language sources, 3) Science courses, 4) Literature courses, 5) at least 30 credits in senior psychology courses, and 6) credits in an area of psychology specialization.

4.2 Data Collection

The data was collected over five academic years from 2007 to 2011. The data set included course evaluations (see Table A3 in Appendix A), student grades and demographics. For the analyses, only data about students who had started and completed their degrees during the five academic years were considered. This decision was based on the available data. While students may have taken a longer period of time to graduate than the 5 year time span indicated in this study, the available data did not afford opportunity to identify if such students actually completed the program of study or had left the university. Thus, the total number of students in the study was 1061. The mean value of GPA for the included participants was 70.89 (SD = 8.46) and the distribution of GPA over the four GPA intervals defined in Section 3 (c.f. Table 1) was: 6 students in [90-100], 137 in [80-90), 449 in [70-80), and 469 in [60-70). There were 788 female (97 international) and 273 male (38 international) students. On average students undertook 12.31 courses (SD = 3.25) in Year 1 of their studies, 11.88 (SD = 3.19) in Year 2, 12.25 (SD = 3.35) in Year 3, and 12.35 (SD = 3.29) in Year 4[2].

A total of 47 psychology courses were included in the dataset. A total of 192 instructors taught in these courses. Ten psychology courses did not have any pre-requisites. Twelve courses had only one pre-requisite course and the other courses had different combinations of pre-requisites consisting of one or two courses. Of those, the students in our dataset enrolled in a total of 37 courses. The total number of courses available at all other departments in the university was 921, and of those, the students included in the study enrolled in 526 courses. The courses included in the study received on average 120.5 (SD = 168.1) evaluations per academic year. Due to the confidentiality policy followed by the institution, the number of submitted evaluations per student was not available nor was the link between course evaluations and individual students who completed them. Instructors who taught the courses included in our dataset received on average an evaluation score of 3.89 (SD = 0.49) on a five-point Likert scale. The minimal average evaluation score of an instructor was 2.68 and the maximal average score was 4.87.

4.3 Ethics and Privacy

The extraction and analysis of student and faculty online behaviour can provide useful insights into the learning process and the possible impact of implemented pedagogical practices. However, these analyses also raise concerns about the ethics and privacy of these forms of analysis and research. In this context our approach was informed by and adhered to the institution's policies on research involving human subjects and the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2—2nd edition of Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, 2010). This study complied with all stated ethics and privacy policies.

4.4 Measurement

The training data set was used as an input for the proposed AHP-based approach. The output data, that is the sets of predicted course selections, were then compared to the initial test dataset. The comparison between

[2] It is important to note that the undergraduate four-year bachelor's degrees (for both science and arts), in Canada, have a requirement for students to undertake at least 120 credits (e.g., 40 x 3-credit courses) for degree completion. The present study is consistent with the literature in identifying that students typically take more courses and therefore accrue more credits than is required for degree completion.

the predicted course selections and the test dataset was defined as distances between two (predicted and test) set of courses as follows.

Let us consider that the predicted number of students enrolling into course *c* is n_1. The estimation error of the actual number of the students enrolled to a course n_c was measured as a relative distance of n_1 to n_c , i.e.: $d^c(n_c, n_1) = \frac{n_1 - n_c}{n_c}$. This measure of relative distance is used to address research question 1.

Let us now consider two sets of courses $o_1 = (c_{1,1}, c_{1,2}, ..., c_{1,k}, c_{1,k+1}, ..., c_{1,2k}, c_{1,2k+1}, ...c_{1,3k})$ and $o_2 = (c_{2,1}, c_{2,2}, ..., c_{2,k}, c_{2,k+1}, ..., c_{2,2k}, c_{2,2k+1}, ...c_{2,3k})$, consisting of courses in (i) term 1 (courses indexed by *1* to *k*) and term 2 (courses indexed by *k+1* to *2k*) of a winter semester and (ii) a summer semester (courses indexed by *2k+1* to *3k*). The distance between two options o_1 and o_2 is defined as

$$d(o_1, o_2) = diff(\{c_{1,1}, c_{1,2}, ..., c_{1,k}\}, \{c_{2,1}, c_{2,2}, ..., c_{2,k}\}) + \\ diff(\{c_{1,k+1}, ..., c_{1,2k}\}, \{c_{2,k+1}, ..., c_{2,2k}\}) + diff(\{c_{1,2k+1}, ..., c_{1,3k}\}, \{c_{2,2k+1}, ..., c_{2,3k}\}),$$

where $diff(\{.\},\{.\})$ denotes the number of different elements between those two sets. The measure allows only combinations of courses that are valid under hard constraints as defined in Section 2.2.8 and maximizes the satisfaction of the students' preferences as extracted through the process introduced in Section 3. The above measure corresponds to the aim of the proposed AHP-based approach that is to predict course selections over semesters/terms and academic years, and was therefore used to address research questions 2 and 3**.** The values of the proposed measures could take were in the interval 0-1 where 0 meant no correct prediction was made and 1 all correct predictions were made.

Finally, for each research question, we also calculated: i) Pearson's R correlation coefficient between the predicted outcome by the proposed model and the actual values from the test data set; and ii) root mean squared error (RMSE) to measure how far each observation from the test dataset was from its mean estimated by the predictive model. These measures were implemented in order to compare the approach undertaken in the present study with the findings from related studies adopting alternate methods.

4.5 Procedures

To address RQ1a; RQ2a and RQ3a, we first annotated the courses available in our dataset with the concerns and qualifier tags as described in Section 2.2. We then used the brute force algorithm to determine all the valid combinations of courses in order to enforce the hard constraints. Finally, we ran the AHP-based approach on the set of the allowed combinations of courses. In this process, the preferences of individual students were qualifier tags extracted as described in Section 3. Each concern had an equal importance.

To undertake RQ1b; RQ2b and RQ3b, we controlled the importance of individual concerns. In the first group of experiments, we assigned a higher importance to a concern, one at a time, while keeping all other concerns unchanged (equal) importance. Based on the results of the first group of the experiments, the importance distributions were created by grouping concerns with similar individual values. Next, the AHP-based approach with the revised importance distributions was performed for each of the three research questions. The results were then compared to those obtained under the conditions of the parts a) of the research questions.

To test the accuracy of the proposed approach, we undertook a five-fold cross validation. This approach aimed to minimize problems such as over fitting and analyse how the model can be used for making predictions of newly enrolled students, i.e. students with having no prior experience with the University under consideration. To this end, five-fold cross validation made use of the training and test sets. The training set consisted of data from the students' course enrollment and evaluation data during the first two years of their degree programs. The test set contained the students' course enrollment and evaluation data during their third and fourth academic years of their degree programs. The students from the training and testing sets are then randomly split into five smaller subsets (i.e. five smaller training sets and five corresponding test sets). Finally, the following procedure is applied to each of the five "folds": *(i)* a model is trained using four folds as training data (i.e. 80% students from years 1 and 2); and *(ii)* the resulting model is validated on the remaining part of the test sets (i.e., remaining 20% of students and enrolment data from years 3-4). The average values for the accuracy of the predication are then reported in the paper.

Finally, to address research question 4, we incorporated neural networks as this is a well-regarded methodology in the extant literature (Kardan, Sadeghi, Ghidary, & Sani, 2013). The neural networks are designed separately for each of the three research questions, and each developed system consists of two stages (Zhang, Patuwo, & Hu, 1998): the input system stage, and the neural networks application stage. The in-sample data-set is used for the design of the input selection stage and design of neural networks. The out-of-sample set is reserved for the final test of the system. Five-fold cross-validation was also used for neural network models. We used fixed ratio of in-sample and out-of-sample sets on approximately 75% and applied repeated random partitioning procedure for the in-sample set into the training set and validation set, and repeated the training and validation for the different partitions (Refenes, 1995). Since training a network till 'death' for highly noisy applications can introduce some over-fitting, we incorporated an early stopping procedure. The best number of iterations is determined with the help of the validation set at 100.

4.6 Data Analysis

Due to the nature and type of the data collected, the analyses employed were standard descriptive statistics (Blaikie, 2003) including reported mean (M) and standard deviation (SD) values. The ANOVA test was used to check for significant differences in the accuracy in the prediction as a result of changing the importance of individual concerns. The sphericity assumption was checked by Mauchly's sphericity test for all the uses of the ANOVA test in the paper and this test confirmed the sphericity assumption for all the uses of ANOVA in the study. A t-test was adopted to assess if the changed distribution of the concern importance significantly improved the accuracy of the predictions as compared to the condition when the concerns had equal importance. Shapiro-Wilk test of normality was used and confirmed that the t-test assumptions were satisfied for all the uses o the t-test in the study. Cohen's *d* for the t-tests and partial η for the ANOVA were reported as measures of effect sizes (Cohen, 1992).

5 Results

The results are organized according to the four research questions. Table 2 provides an overview of the results for research questions 1-3, according to the evaluation approach outlined in Section 4.5. Table 4 reports the results of relevance for research question 4.

5.1 Research Question 1

***RQ1a*.** This research question compares the accuracy of the proposed approach to predict the number of enrollments per course offering against the reported related literature such as Kardan et al. (2013). The mean value of the relative distances $d^c()$ between the number of enrollments predicted by the AHP-based approach and the actual number of enrollments for each course was 22.48% (SD = 15.52%). Thus, the AHP-based approach accurately predicted 77.52% of course selections in the test dataset. This result shows that presented approach may be considered as appropriate for making estimations of the total number of students enrolling to the course. However, the accuracy of the model could be increased in order to enhance its potential practical value. This aspect is addressed in RQ1b.

***RQ1b*.** The descriptive statistics for each of the six experiments (groups C1-6) are reported in Table 2. As the collected data were not normally distributed, a one way ANOVA test was used over the log-transformed data to compare the means of the dependent variable. The results demonstrate a significant difference in the distances to the test dataset in the case of higher importance coefficients to individual concerns (*F(5; 546) = 32.08; p = 0.003, partial* η^2*=0.44*). The Tukey post-hoc test revealed that there was no significant difference between groups C2, C4, C5 and C6. Thus, the accuracy of the proposed approach may be increased by adjusting the importance of the identified concerns.

Table 2. The descriptive statistics of the three research questions from the conducted study

Part of research question	Measure	RQ1	RQ2	RQ3
A	d – M, SD	0.2248, 0.1652	0.5707, 0.1164	0.4634, 0.1842
	A	0.7752	0.4293	0.5366
	R	0.739	0.393	0.542
	MSE	0.0085	0.0009	0.0075
	RMSE	0.092	0.003	0.087
B	d – C1	0.2055, 0.1732	0.7412, 0.1147	0.6619, 0.2315
	d – C2	0.2368, 0.1530	0.6589, 0.1488	0.6148, 0.2145
	d – C3	0.2192, 0.18.64	0.6523, 0.1256	0.5647, 0.3512
	d – C4	0.2589, 0.1128	0.6625, 0.1321	0.7514, 0.1535
	d – C5	0.2312, 0.1702	0.6489, 0.1544	0.7275, 0.2861
	d – C6	0.2673, 0.1689	0.7566, 0.17.89	0.6965, 0.2273
	d – M, SD	0.176, 0.1982	0.4623, 0.1235	0.3573, 0.2416
	A	0.8240	0.5377	0.6427
	R	0.791	0.502	0.596
	MSE	0.0036	0.0058	0.0040
	RMSE	0.06	0.076	0.063

Legend: d – mean and standard deviation values of the distance of the overall as defined in Section 4.4; A – accuracy; R – Pearson's R coefficient; MSE – mean squared error; RMSE – root mean squared error; d – C_i (i = 1..6) – mean values and standard deviation values for the distance as defined in Section 4.4 when individual concerns are prioritized: C1 – the course characteristics concern; C2 – the instructor characteristics concern; C3 – the GPA value for course concern; C4 – the course time scheduling concern; C5 – the student demands concern; and C6 – the demographic characteristics concern.

Based on the results, we can infer that the course characteristics and GPA distributions of past students enrolled to the course have the highest accuracy in predicting the expected number of students. The accuracy is higher than when all concerns are combined with equal importance as reported in research question RQ1a. Thus, the model may be improved by assigning a higher importance to the two concerns over others. For example, consider the predictive model was formalized with equation: $y = \frac{3*(x_1 + x_3) + x_2 + x_4 + x_5 + x_6}{6}$, where variables x_i corresponds to the variables previously emphasized by *i*-th group. We ran an experiment with the revised model that used *y* to evaluate its accuracy against the test dataset. The mean value of the obtained relative distances in the combinations of courses was 17.6% (SD = 19.82%). Thus, the accuracy of the enhanced approach was 82.4%. A t-test confirmed that the increase in the accuracy was significant (*t(1060)=15.38%, p<0.05, d=0.56*) as compared to the accuracy of the model where all concerns were assigned equal importance.

5.2 Research Question 2

RQ2a. This research question evaluates how accurately the proposed approach can predict the course enrolments for each student per semester/term. That is, *d(.)* is a measure that compares sets of courses at the level of course selections per semester/term for each individual student. This is the most difficult prediction task in the evaluation due to the requirement to predict a particular set of courses selected in each semester by a student – especially given the quantity and diversity of courses available across the University. The mean value of the obtained relative distances between the combinations of the courses predicted by the AHP-based approach and those available in the test dataset was 57.07% (SD = 11.64%). Thus, the prediction accuracy of the AHP-based approach was 42.93% of the students' course selections for each semester in the test set – based on the extracted preferences of the students and course annotations from the training dataset.

RQ2b. The results showed a significant difference in the distances to the test dataset in case of higher importance coefficients to individual concerns (*F(5; 546) = 25.47; p = 0.000, partial* η^2*=0.33*). The Tukey post-hoc test revealed that there was no significant difference between groups C1 and C6, C4 and C2-3. Thus, the accuracy of the proposed approach may be increased by adjusting the importance of the identified concerns.

Based on the results we can conclude that the data about a student's interests and preferences are more important contributors to the accuracy of the predictive model than other data types. The model may be improved by assigning a higher importance to this concern, lower importance to the concerns from groups C3, C2 and C4, and a further reduction in the importance of the remaining concerns. For example, consider the predictive model formalized with equation: $y = \frac{3*x_5 + 2*(x_4 + x_2 + x_3) + x_1 + x_6}{6}$, where variables x_i corresponds to the variables previously emphasized by *i*-th group. We ran an experiment with the revised model with *y* that used to evaluate its accuracy against the test dataset. The mean value of the obtained relative distances in the combinations of courses was 46.23% (SD = 12.35%). Thus, the accuracy of the enhanced approach was 53.77%. A t-test confirmed that the increase in the accuracy was significant (*t(1060)=4.71%, p<0.05, d=0.34*) as compared to the accuracy of the model was had all concerns with equal importance (reported under research question RQ2a).

5.3 Research Question 3

RQ3a. The final research question tests the accuracy of the proposed AHP-based approach when the prediction task was relaxed to predict the course enrolments for each student per academic year (i.e. semesters and terms belonging to the same academic year were joined). That is, *d(.)* is a measure that compares sets of courses at the level of course selections per academic year for each student. The mean value of the obtained relative distances between the combinations of courses predicted by the AHP-based approach and those available the test dataset was 46.34% (SD = 18.42%). Thus, the prediction accuracy of the AHP-based approach was 53.66% of the students' course selections for each semester in the test set, based on the extracted preferences of the students and course annotations from the training dataset.

RQ3b. The results indicate a significant difference in the distances between the test dataset in the case for higher importance coefficients of individual concerns (*F(5; 546) = 48.25; p = 0.018, partial* η^2*=0.45*). The Tukey post-hoc test revealed that there was no significant difference only between groups C3 and C1-2, C4-6 and between groups C2-4.

The findings from the study suggest that the data about a student's interests and preferences in the subject areas has a corresponding positive impact on the accuracy of the course prediction model for each student per semester/term. Thus, the model may be improved by elevating the importance of this concern over the other concerns in the remaining groups. For example, consider the predictive model formalized in the equation: $y = \frac{3*x_3 + 2*(x_1 + x_2 + x_5 + x_6) + x_4}{6}$, where variables x_i corresponds to variables previously emphasized by the *i*-th group. We ran an experiment with the revised model that used *y* to evaluate its accuracy against the test dataset. The mean value of the obtained relative distances in the combinations of courses was 35.73% (SD = 24.16%). Thus, the accuracy of the enhanced approach was 64.27%. A t-test confirmed that the increase in the accuracy was significant (*t(1060)=4.13, p<0.05, d=0.55*) as compared to the accuracy of the model where all concerns were assigned equal importance (reported under research question RQ3a).

5.4 Research Question 4

Table 4 shows the results obtained by the use of neural networks for the tasks addressed in the previous three research questions and applying the procedure described in Section 4.5. The use of neural networks (Table 4) revealed a considerable decrease in the performance compared to the results of the proposed method (Tables 2-3). Specifically, the accuracy values for the neural network model compared to the accuracy values of the cross-validated model (Table 3) are lower by 17.26% for the task covered by research question 1, by 13.02% for research question 2, and by 20.56% for research question 3.

Table 4. The descriptive statistics of the three research questions from the use of neural networks

Measure	RQ1	RQ2	RQ3
A	0.6514	0.4075	0.4371
R	0.691	0.318	0.373
MSE	0.0052	0.0046	0.0039
RMSE	0.072	0.068	0.062

Legend: A – accuracy; R – Pearson's R coefficient; MSE – mean squared error; RMSE- root mean squared error

6 Discussion

The results of the evaluation of the prediction task studied in research question RQ1 produced the highest accuracy. This was most pronounced when the importance of the concerns about course characteristics and the GPA value for courses were increased in comparison to the other concerns. To assist in the interpretation and impact of these findings, we compared the accuracy level of the AHP-based approach with those of other related published studies. Although, to date there have been limited investigations of this kind to undertake a comprehensive comparison and evaluation. However, in a similar study, Kardan et al. (2013)applied neural networks to predict course selection in two fully online master's program. Table 5 provides a comprehensive comparison of the proposed AHP-based approach adopted in the present study in comparison to the neural networks-based model proposed by Kardan et al. (2013).

Table 5. A comparative comparison of the AHP-based course selection prediction approach and the neural network-based model proposed by Kardan et al. (2013)

Approach	Experimental condition	Measures		Variables									
		R	*MSE*	*C1*	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*	*C8*	*C9*	*C10*
Neural network	Kardan et al.'s experiment 1[a]	0.896	0.0036	X	x	x	x	x	x	x	x		
	Kardan et al.'s experiment 1[b]	0.923	0.0029	X	x	x	x	x	x	x	x	x	
	Experiment in this study	0.691	0.0052	X	x		x	x	x			x	x
AHP	Different importance	0.802	0.0037	X	x		x	x*	x			x	x

Legend: [a] - Kardan et al's (2013) study regression calculated without considering student demands and [b] – considering student demands; R – Pearson's correlation coefficient; MSE – mean squared error; C1 – course characteristics; C2 – instructors' characteristics; C3 – students' workload; C4 – course grade; C5 – course type; C6 – course time; C7 – number of time conflicts; C8 – final examination time; C9 – student demands; and C10 – demographic characteristics; * hard constraints, not as a variable included in the model.

A comparison between alternate approaches (AHP vs. Neural networks) indicates there is a considerable overlap between the sets of variables employed in developing the predictive models. However, there were also some observed differences. For example, the neural network-based model employed by Kardan et al. (2013)

incorporated a larger set of variables than those employed in the present study. As detailed in Table 5, the added variables included (C3) student workload; (C7) number of time conflicts; and (C8) the final examination time. In contrast to the Kardan et al. (2013) study, the AHP-based model incorporated demographic variables.

The values of Pearson's R correlation coefficient and MSE (see Table 5) can be used as a comparison for the two modeling approaches[3]. A higher R value associated with the neural network model with the overall number of registrations trained with dataset from the the Kardan et al. study was observed in comparison to the AHP-based approach. However, when neural networks were applied to the dataset used in our study, the values for R coefficients were considerably lower compared to those obtained with the use of the proposed AHP method. The R values for the neural network models compared to the AHP models were lower by 0.111 for the task covered by research question 1, by 0.184 for research question 2, and by 0.223 for research question 3.

The considerably weaker performance of the neural network model on the accuracy measure and Pearson's correlation coefficient observed in this study compared to the neural model used by Kardan et al. (2013) is likely to be due to the type of datasets adopted in the different studies. As shown in Section 4.2, the results of the present study were derived from an undergraduate program offering over 900 courses and requiring completion of a minimum of 120 credits (typically equivalent to 40 courses). In comparison, the neural network model of the Kardan et al. (2013) study was based on two online master's programs, each of which requested students to take four out of six core (i.e., mandatory) courses and four out of nine elective courses (i.e., students were requested to take eight courses to get their master's degrees). This is a clear indication that the context and setting in the present study was of a far greater level of complexity. This complexity was a result of the larger flexibility of course selection options and the greater number of course offerings available to students considered in the present study. For example, the dataset of the present study included information from some 50,000 course selections. In comparison, the Kardan et al. study contained only 5,937 course selections. Furthermore, the Kardan et al. study derived data from repeated enrolments across the same 17 courses over a period of eight years (i.e., more than one cohort of students who completed their degrees). The present study analysed data from only a single cohort of students with a degree completion within a five year period.

The results of the comparisons between the AHP-based approach and the neural network models on their RSME scores were mixed. While the AHP-based model outperformed the neural network models (0.06 vs. 0.072) in the prediction of the overall number of course registrations (RQ1), the neural network model outperformed the AHP-based approach (0.076 vs 0.068) for the predictions of course selection for each individual student per semester/term (RQ2). The models were almost tied in their RMSE performance (0.63 vs 0.62) in the predictions of course selection for each individual student per academic year (RQ3). In our research questions, we were interested in the prediction tasks for which the accuracy measure is most commonly used in existing research. However, these mixed results of RMSE with both accuracy measure and Pearson's correlation coefficient warrant future research.

Intrinsic curricular differences – namely, the number of course options offered to students to take – could be an important reason for better performance of one approach over another one in the prediction task studied in this paper. As already mentioned, the dataset used in the study was based on an academic program offered a wide rage of courses to students to take. This probably resulted in the insufficient dataset for neural networks to train a highly accurate model. This reason is traditionally associated with neural networks (Kramer & Leonard, 1990). On the other hand, the AHP-based approach is not so much dependent on the size of the training set, and thus, produced higher accuracy with the dataset used in this study. Neural networks are more likely to perform well on the training datasets generated in academic curricula with much lower variability, as it was the case of the datasets used in the Karman et al (2013) study. Future research needs to validate these discussion points and conditions under which different modelling approaches are more suitable to be used.

[3] Kardan et al. (2013) provide only the R and MSE values in the results of their paper.

The accuracy of the AHP-based predictions of course selection for each individual student per semester/term (RQ2) and academic year (RQ3) were clearly lower than the accuracy of the predictions of the overall number of course registrations (RQ1). Given the obvious importance for establishing high accuracy in predictive modeling, the output from the AHP-based model are only moderately associated with the actual course selections of each student per semester/term (RQ2) or academic year (RQ3). Clearly, the prediction tasks for RQ2 and RQ3 were more difficult. This is in part due to the additional temporal dimension associated with the course selection of each individual student for RQ2 and RQ3. For example, a possible interpretation of this finding is that the overall interest of all students in a particular course (RQ1) is easier to predict than a set of individual students who will finally register for a course in a particular semester/term or an academic year (RQ2 and RQ3). Unfortunately, a comparison of the results of the AHP-based prediction was not possible due to the lack of available studies that address the course selection tasks as defined in research questions RQ2 and RQ3. It is envisaged that the findings presented here will serve as a benchmark for future comparative studies.

A common trend observed for all prediction tasks in the research questions was that the concerns used for modeling students' preferences had a different importance for the overall accuracy. These varied importance values offer some evidence that students' have different preferences and priorities in their decision making process for selecting their future courses. These preferences differ according to the particular prediction tasks (see Table 6). For example, when attempting to predict the number of students that will register for a course, the course characteristics and the GPA value for a course were observed to be the most important. When predicting the course selections of each student per semester/term then the concerns relating to student demands were the most important. In predicting the course selections of each student per academic year, the concern relating to course scheduling time was considered the most important. While the changed importance of individual concerns had a medium effect size for the accuracy of prediction tasks studied in RQ1 and RQ3 (d_{RQ1} = 0.56 and d_{RQ3} = 0.55), the effect size for the accuracy of the prediction task studied in RQ2 was small (d_{RQ2} = 0.34) (Cohen, 1992). The effect size differences were expected, given the complexity of the prediction task for RQ2 (the course selection for each student per semester/term). Additional types of information would be required related to student interests and preferences in order to enhance the accuracy of the model.

The results of the study indicate that robust predictions concerning student course selections are achievable despite access to limited and complex data. A relatively accurate predictive model can be developed even in instances where student anonymity (from course evaluations) is required. Although the protection of student anonymity is a necessary practice, it does in this context create a high level of complexity as the student's individual evaluative rankings of courses and instructors cannot be tied to other information sets such as grades or course scheduling. However, as this study demonstrates, an accurate model capable of predicting student course selections is a feasible goal while still supporting and adhering to, privacy and ethics legislation. The ability to develop such a predictive model including non-identifiable data would not be possible using more conventional data mining techniques. Therefore, an outcome of this research is to stress the importance of drawing on alternative predictive modeling techniques that are frequently encountered in disciplines such as decision science in lieu of the more traditional data-mining approaches.

Table 6. Comparative analysis of the importance observed for individual concerns in the empirical evaluation for each of the three research questions, and the effect size of the changed importance on the prediction accuracy

Variable/ Effect size	**RQ1**	**RQ2**	**RQ3**
C1	H	L	M
C2	L	M	M
C3	H	M	H
C4	L	M	L
C5	L	H	M
C6	L	L	M

Cohen's *d*	0.56	0.34	0.55

Legend: C1 – the course characteristics concern; C2 – the instructor characteristics concern; C3 – the GPA value for course concern; C4 – the course time scheduling concern; C5 – the student demands concern; and C6 – the demographic characteristics concern

Overall, the established predictive model produced a high level of accuracy for course selections. However, this accuracy further diminished when integrating more temporally based variables such as semester/ term or academic year. We see opportunities for improving this outcome through investigations into the following two areas. First, the approach proposed in this paper and other solutions studied in the literature do not consider the impact of individual needs and preferences that could be characteristic of different student subpopulations. Modern counselling psychology deems learners as active agents in their course selection process (Bandura, 2006; Haggard & Tsakiris, 2009). Therefore, to improve the accuracy of the predictive model further consideration is required regarding the impact of individual differences (e.g., self-efficacy) in course selection. The results for the research questions suggest that the GPA value for a course concern was consistently of a high importance. There is a high association between an individual student's academic performance (i.e., GPA) and the distribution of grades in a particular course (Marsh & Roche, 2000; Svanum & Aigner, 2011). If the academic performance of a student is lower than the average grade of a course, the student is less motivated to enrol in that course. One interpretation for this finding is that there is a level of incongruence between the student's self-efficacy and efficacy expectation, and that of the course under consideration. Simply put, the perceived capacity for a student to succeed in a course is evidenced through a social comparison of peer grades. A high grade differential will result in the student selecting an alternate course. Self-efficacy and efficacy expectations are well-established as strong predictors of academic achievement (Robbins, et al., 2004) and career planning (Lent, Brown, Brenner, Lyons, & Treistman, 2003). Therefore, it seems a promising avenue for future research to investigate different approaches to the identification of student sub-populations, who for example share similar characteristics, make similar courses selections, have similar learning achievements and share similar subject domain interests. As demonstrated in detecting learner profiles from data about interactions with learning environments, the use of clustering techniques such as K-means or hierarchical clustering is a fruitful research direction for identifying student subpopulations (Lust, Vandewaetere, Ceulemans, Elen, & Clarebout, 2011).

Consideration of the temporal dependencies in course selection is essential for improving the accuracy of the predictive model for course selection. As shown in the results, the predictive accuracy of course selection for each student per semester/term and academic year was relatively poor and too immature for any practical adoption to support any institutional decision-making process. Building on these preliminary findings, we would suggest that the treatment of students' course selection throughout their academic programs should be considered a process, rather than aggregate values of certain variables of interest. The algorithms and tools associated with process mining would appear to be a promising avenue to pursue in order to address this temporal challenge (Pechenizkiy, Trčka, Vasilyeva, van der Aalst, & De Bra, 2009; van der Aalst, 2012). For instance, process mining can be used for discovering "emergent curricula", which are commonly followed by students throughout their degree programs. If combined with clustering, such emergent curricula can be discovered from different student subpopulations. Moreover, the predictions about student course selection sequences can be considered a process configuration problem, which already has approaches based on the use of AHP (Ognjanović, Mohabbati, Gašević, Bagheri, & Bošković, 2012).

The modelling of qualifier tags used in the AHP-approach was based on discretization decisions made in the scope of the study. For example, sections 2.2.3 and 2.2.4 explain the rationale why four qualifier tags were created for course characteristics and instructor characteristics. Section 2.2.5 likewise explains the rational for the four qualifier tags for grade average point. We acknowledge that some other methods could be used for their grouping and possible influence some of the results reported in the paper. Our choices are transparently explained with the goal to enable other researchers replicate our study. Future research should investigate the effectiveness of these decisions and examine some alternative ways for the creation of qualifier tags and their influence on the results.

The further consideration of alternate and additional variables that can aid in the prediction of course selection is essential next step. Well-established theories of student retention posit that academic and social integration of students is critical for their success in higher education (Tinto, 2006). For this reason and many other established educational benefits, many institutions are attempting to implement strategies and practices aimed towards fostering learning communities to build a network of support for their students (Dawson, 2006; Smith, MacGregor, Matthews, & Gabelnick, 2004). A student's decisions are often influenced by different contextual/environmental factors (Babad & Tayeb, 2003). For example, enrolling with a group of friends is a common influencing factor for course selection. Therefore, consideration of the social structures and position of students in social networks is an important source to consider in order to enhance the prediction accuracy of course selections (Dawson, Macfadyen, Lockyer, & Mazzochi-Jones, 2011; Gašević, Zouaq, & Jenzen, 2013).

The higher education landscape is increasingly complex requiring students to identify future career plans and pathways early in their academic candidature. Understandably, not all students are job-focused with the necessary insight and motivation to make productive career decisions. This is well recognised across the higher education sector with the vast majority of universities now offering career counselling and support services. While this approach offers much personal value there is limited scalability in practice. Not all students would or could avail themselves of these services. There simply are insufficient resources to cope with a large increase in demand. However, as the education sector increases its application of data and analytics to the decision making process (strategic and personalised) (Siemens, Dawson, & Lynch, 2014), there is an opportunity to better align support resources to assist students in their career plans. As this study well illustrates the application of student data derived from information sources can provide valuable predictive insights into student course selections, demands and drivers in order to better align the university support resources to achieve a greater return on investment.

7 Bibliography

Allen, I. E., & Seaman, J. (2013). *Changing Course: Ten Years of Tracking Online Education in the United States.* Oakland, CA: Babson Survey Research Group and Quahog Research Group, LLC. Retrieved from http://www.onlinelearningsurvey.com/reports/changingcourse.pdf

Arbaugh, J. B., & Duray, R. (2002). Technological and structural characteristics, student learning and satisfaction with web-based courses- An exploratory study of two on-line MBA programs. *Management Learning, 33*(3), 331-347. doi:10.1177/1350507602333003

Babad, E. (2001). Students' course selection: Differential considerations for first and last course. *Research in Higher Education, 42*, 469-492. doi:10.1023/A:1011058926613

Babad, E., & Tayeb, A. (2003). Experimental analysis of student's course selection. *British Journal of Eductaional Psychology, 73*, 373-393. doi:10.1348/000709903322275894

Babad, E., Darley, J. M., & Kaplowitz, H. (1999). Developmental Aspects in Students' Course Selection. *Journal of Educational Psychology, 91*(1), 157-168. doi:10.1037/0022-0663.91.1.157

Baker, R. S. (2010). Data mining for education. In B. McGaw, P. Peterson, & E. Baker, *International Encyclopedia of Education* (pp. 112–118). Oxford, UK: Elsevier. doi:10.1016/B978-0-08-044894-7.01318-X

Baker, R. S., Corbett, A. T., & Aleven, V. (2008). More Accurate Student Modeling Through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, (pp. 406-415). doi:10.1007/978-3-540-69132-7_44

Bala, M., & Ojha, D. B. (2012). Study of applications of data mining techniques in education. *International Journal of Research in Science And Technology, 1*(4), 1-10.

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on psychological science, 1*(2), 164-180. doi:10.1111/j.1745-6916.2006.00011.x

Berg, M. (1997). *Rationalizing medical work: decision-support techniques and medical practices.* Cambridge, MA: MIT Press.

Blaikie, N. (2003). *Analyzing Quantitative Data.* London: Sage.

Borgida, E. (1978). Scientific deduction-evidence is not necessarily informative: A reply to Wells and Harvey. *Journal of Personality and Social Psychology, 36*(5), 477-482.

Brafman, R. I., & Domshlak, C. (2002). Introducing variable importance tradeoffs into CP-nets. *Proceedings of the Eighteenth Conference on Uncertainty in AI*, (pp. 69-76). Canada.

Campbell, J., DeBlois, P., & Oblinger, D. (2007). Academic analytics: a new tool for a new era. *Educause Review, 41*(4), 40–57. Retrieved from https://net.educause.edu/ir/library/pdf/PUB6101.pdf

Castellano, E., & Martinez, L. (2008). ORIEB, A CRS for Academic Orientation Using Qualitative Assessments. *Proc. of the IADIS International Conference E-Learning*, (pp. 38-42).

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education, 44*(5), 495-518. doi:10.1023/A:1025492407752

Chen, M. K., & Wang, S. (2010). The critical factors of success for information service industry in developing international market: Using analytic hierarchy process (AHP) approach. *Expert Systems with Applications, 37*(1), 694–704. doi:10.1016/j.eswa.2009.06.012

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. doi:10.1037/0033-2909.112.1.155

Dawson, S. (2006). Online forum discussion interactions as an indicator of student community. *Australasian Journal of Educational Technology, 22*(4), 495-510.

Dawson, S., Macfadyen, L., Lockyer, L., & Mazzochi-Jones, D. (2011). Using social network metrics to assess the effectiveness of broad based admission practices. *Australasian Journal of Educational Technology, 27*(1), 16-27.

Dutton, J., Dutton, M., & Perry, J. (2002). How do online students differ from lecture students? *Journal of Asynchronous Learning Networks, 6*(1), 1-20.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5/6), 304–317. doi:10.1504/IJTEL.2012.051816

Forman, E. H., & Gass, S. I. (2001). The analytical hierarchy process—an exposition. *Operations Research, 49*(4), 469-487. doi:opre.49.4.469.11231

Gašević, D., Zouaq, A., & Jenzen, R. (2013). “Choose your Classmates, your GPA is at Stake!” The Association of Cross-Class Social Ties and Academic Performance”. *American Behavioral Scientist, 57*(10), 1460-1479. doi:10.1177/0002764213479362

Greenwald, A. G., & Gillmore, G. M. (1997). No Pain, No Gain? The Importance of Measuring Course Worklaod in Student Ratings of Instruction. *Journal of Educational Psychology, 89*(4), 743-751. doi:10.1037/0022-0663.89.4.743

Haggard, P., & Tsakiris, M. (2009). The experience of agency: Feelings, judgments and responsibility. *Current Directions in Psychological Science, 18*(4), 242-246. doi:10.1111/j.1467-8721.2009.01644.x

Hsia, T.-C., Shie, A.-J., & Chen, L.-C. (2008). Course planning of extension education to meet market demand by using data mining techniques - an example of Chinkuo technology univeristy in Taiwan. *Expert Systems with Applications, 34*, 596-602. doi:10.1016/j.eswa.2006.09.025

Kardan, A. A., Sadeghi, H., Ghidary, S. S., & Sani, M. R. (2013). Predicition of student course selection in online higher education institutes using neural network. *Computers & Education, 65*, 1-11. doi:10.1016/j.compedu.2013.01.015

Kramer, M. A., & Leonard, J. A. (1990). Diagnosis using backpropagation neural networks—analysis and criticism. *Computers & chemical engineering, 14*(12), 1323-1338. doi:10.1016/0098-1354(90)80015-4

Lent, R. W., Brown, S. D., Brenner, B., Lyons, H., & Treistman, D. (2003). Relation of contextual supports and barriers to choice behavior in engineering majors: Test of alternative social cognitive models. *Journal of Counseling Psychology, 50* (4), 458–465. doi:10.1037/0022-0167.50.4.458

Liu'an, K., Xiaomei, W., & Lin, Y. (2012). The Research of Teaching Quality Appraisal Model Based on AHP. *International Journal of Education and Management Engineering, 9*, 29-34. doi:10.5815/ijeme.2012.09.05

Lust, G., Vandewaetere, M., Ceulemans, E., Elen, J., & Clarebout, G. (2011). Tool-use in a blended undergraduate course: In Search of user profiles. *Computers & Education, 57*(3), 2135-2144. doi:10.1016/j.compedu.2011.05.010

MacFadyen, L. P., Dawson, S., Prest, S., & Gašević, D. (in press). Whose feedback? A multilevel analysis of student completion of end-of-term teaching evaluations. *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2015.1044421

Marginson, S. (2013). The impossibility of capitalist markets in higher education. *Journal of Education Policy, 28*(3), 353-370. doi:10.1080/02680939.2012.747109

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry, & J. C. Smart, *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319-383). Dordrecht, The Netherlands: Spring. doi:10.1007/1-4020-5742-3_9

Marsh, H. W., & Roche, L. A. (2000). Effects of Grading Leniency and Low Workload on Students' Evaluations of Teaching: Popular Myth, Bias, Validity, or Innocent Bystanders? *Journal of Educational Psychology, 92*(1), 202-228. doi:10.1037/0022-0663.92.1.202

Marshall, J., Greenberg, H., & Machun, P. A. (2012). How would they choose? Online student preferences for advance course information. *Open Learning: The Journal of Open, Distance and e-Learning, 27*(3), 249-263. doi:10.1080/02680513.2012.716656

Ognjanović, I., Gašević, D., & Bagheri, E. (2013). A Stratified Framework for Handling Conditional Preferences: an Extension of the Analytic Hierarchy Process. *Expert Systems with Applications, 40*(4), 1094-1115. doi:10.1016/j.eswa.2012.08.026

Ognjanović, I., Mohabbati, B., Gašević, D., Bagheri, E., & Bošković, M. (2012). A Metaheuristic Approach for the Configuration of Business Process Families. *Proceedings of the 9th IEEE International Conference on Servicea Computing (SCC2012)* (pp. 25-32). Hawaii, USA: IEEE Press. doi:10.1109/SCC.2012.6

Paechter, M., Maier, B., & Macher, D. (2010). Students' expectations of, and experiences in e-learning: Their relation to learning achievements and course satisfaction. *Computers & Education, 54*, 222-229. doi:10.1016/j.compedu.2009.08.005

Pechenizkiy, M., Trčka, N., Vasilyeva, E., van der Aalst, W., & De Bra, P. (2009). Process Mining Online Assessment Data. *Proceedings of the 2nd International Conference on Educational Data Mining*, (pp. 279-288). Cordoba, Spain.

Qureshi, E., Morton, L. L., & Antosz, E. (2002). An interesting profile-University students who take distance education courses show weaker motivation than on-campus students. *Online Journal of Distance Learning Administration, 5*(4).

Refenes, A. N. (1995). *Neural network design considerations.* New York: John Wiley.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do Psychosocial and Study Skill Factors Predict College Outcomes? A Meta-Analysis. *Psychological Bulletin, 130*(2), 261-288. doi:10.1037/0033-2909.1

Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics- Part C: Applications and Reviews, 40*(6), 601-618. doi:10.1109/TSMCC.2010.2053532

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. (2010). *Handbook of Educational Data Mining.* Boca Raton, FL: CRC Press, Taylor & Francis Group.

Saaty, T. L. (1980). *The Analytic Hierarchy Process.* New York: McGraw-Hill.

Schuhmann, P., & McGoldrick, K. (1999). A conjoint analysis of student registration decision making: implications for enrollment. *Journal on Excellence in College Teaching, 10*(3), 93-121.

Siemens, G., Dawson, S., & Lynch, G. (2014). *Siemens, G., Dawson, S., & Lynch, G. (2014). Improving the Quality and Productivity of the Higher Education Sector - Policy and Strategy for Systems-Level Deployment of Learning Analytics.* Camberra, Australia: Office of Learning and Teaching, Australian Goverment. Retrieved from http://solaresearch.org/Policy_Strategy_Analytics.pdf

Siemens, G., Gašević, D., Haythornthwaite, C., Dawson, S., Buckingham Shum, S. F., Duval, E., . . . Baker, R. S. (2011). *Open Learning Analytics.* Edmonton, AB, Canada: Society for Learning Analytics Research. Retrieved from http://solaresearch.org/OpenLearningAnalytics.pdf

Sim, I., Gorman, P., Greenes, R. A., Haynes, R. B., Kaplan, B., Lehmann, H., & Tang, P. C. (2001). Clinical decision support systems for the practice of evidence-based medicine. *Journal of the American Medical Informatics Association, 8*(6), 527-534. doi:10.1136/jamia.2001.0080527

Smith, B., MacGregor, J., Matthews, R., & Gabelnick, F. (2004). *Learning communities: Reforming undergraduate education.* San Francisco, CA: Jossey-Bass.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the Validity of Student Evaluation of Teaching - The State of the Art. *Review of Educational Research, 83*(4), 598-642. doi:10.3102/0034654313496870

Srivastava, R., Gendy, M., Narayanan, M., Arun, Y., & Singh, J. (2012). *University of the future – A thousand year old industry on the cusp of profound change.* Melbourne, Australia: Ernst & Young. Retrieved from http://www.ey.com/Publication/vwLUAssets/University_of_the_future/$FILE/University_of_the_future_2012.pdf

Stewart, C., Bachman, C., & Johnson, R. (2010). Students' Characteristics and Motivation Orientations for Online and Traditional Degree Programs. *MERLOT Journal of Online Learning and Teaching, 6*(2), 367-379.

Sun, P.-C., Tsai, R. J., Finger, G., Chen, Y.-Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education, 50*, 1183-1202. doi: 10.1016/j.compedu.2006.11.007

Svanum, S., & Aigner, C. (2011). The influences of course effort, mastery and performance goals, grade expectancies, adn earned course grades on student ratings of course satisfaction. *British Journal of Educational Psychology, 81*, 667-679. doi:10.1111/j.2044-8279.2010.02011.x

(2010). *TCPS 2—2nd edition of Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.* Ottawa, ON, Canada: Goverment of Canada. Retrieved from http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/Default/

Tinto, V. (2006). Research and Practice of Student Retention: What Next? *Journal of College Student Retention: Research, Theory and Practice, 8*(1), 1-19. doi:10.2190/4YNU-4TMB-22DJ-AN4W

van der Aalst, W. (2012). Process Mining: Overview and Opportunities. *ACM Transactions on Management Information Systems, 3*(2), Article 7, 17 pages. doi:10.1145/2229156.2229157

Vialardi, C., Bravo, J., Shafti, L., & Ortigosa, A. (2009). Recommendation in higher education using data mining techniques. *Proceedings of the 2nd International Conference on Educational Data Mining* (pp. 190-198). Montreal, Canada: IEDMS.

Vialardi, C., Chue, J., Barrientos, A., Victoria, D., Estrella, J., Ortigosa, A., & Peche, J. (2010). A case study: data mining applied to student enrollment. *Proceedings of the 3rd International Conference on Educational Data Mining* (pp. 333–335). Pittsburgh, PA: IEDMS.

Vialardi, C., J. Chue, J. P., Alvarado, G., Vinatea, B., Estrella, J., & Ortigosa, A. (2011). A data mining approach to guide students through the enrollment process based on academic performance. *User Modeling and User-Adapted Interaction, 21*(1-2), 217-248. doi:10.1007/s11257-011-9098-4

Wilson, N. (2011). Computational techniques for a simple theory of conditional preferences. *Artificial Intelligence, 175*(7-8), 1053-1091. doi:10.1016/j.artint.2010.11.018

Yüksel, M. (2012). Evaluating the Effectiveness of the Chemistry Education by Using the Analytic Hierarchy Process. *International Education Studies, 5*(5), 79-91. doi:10.5539/ies.v5n5p79

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14*(1), 35–62.

Appendix A

Table A1. The AHP ranks for the interval grades

	1-2	**2-3**	**3-4**	**4-5**	**Ranks**	**Normalization**
1-2	1.00	0.33	0.20	0.14	0.23	**0.06**
2-3	3.00	1.00	0.33	0.20	0.49	**0.12**
3-4	5.00	3.00	1.00	0.33	1.05	**0.26**
4-5	7.00	5.00	3.00	1.00	2.23	**0.56**
	16.00	9.33	4.53	1.68	4.00	

Table A2. AHP ranks for the ‘unknown’ preferences over the interval grades

	1-2	**2-3**	**3-4**	**4-5**	**Ranks**	**Normalization**
1-2	1.00	1.00	1.00	1.00	1.00	**0.25**
2-3	1.00	1.00	1.00	1.00	1.00	**0.25**
3-4	1.00	1.00	1.00	1.00	1.00	**0.25**
4-5	1.00	1.00	1.00	1.00	1.00	**0.25**
	4.00	4.00	4.00	4.00	4.00	

Table A3. Evaluation questions

University Questions[1]
U1: The instructor made it clear what students were expected to learn. U2: The instructor communicated the subject matter effectively. U3: The instructor helped inspire interest in learning the subject matter. U4: Overall, evaluation of student learning (through exams, essays, presentations, etc.) was fair. U5: The instructor showed concern for student learning. U6: Overall, the instructor was an effective teacher.

Faculty Specific Questions[2]
F1: In classes where the size of the class and content of the course were appropriate, student participation in class was encouraged by the instructor. F2: High standards of achievement were set. F3: The instructor was generally well prepared for class. F4: The instructor was readily available to students outside of class (e.g., through email, office hours, or by appointment). F5: The instructor treated students with respect. F6: Considering everything how would you rate this course?

[1]Asked in all course evaluations of the university

[2]Asked in all course evaluations of the faculty the degree program, used in the study, was offered by