2015-06-08

Investigating the Efficacy of Algorithmic Student Modelling in Predicting Students at Risk of Failing in the Early Stages of Tertiary Education: Case study of experience based on first year students at an Institute of Technology in Ireland.

Geraldine Gray
Technological University Dublin, geraldine.gray@tudublin.ie

Investigating the Efficacy of Algorithmic Student Modelling in Predicting Students at Risk of Failing in the Early Stages of Tertiary Education:

Case study of experience based on first year students at an Institute of Technology in Ireland.

Geraldine Gray

Department of Informatics,
School of Informatics and Engineering,
Institute of Technology Blanchardstown

A thesis submitted to Institute of Technology Blanchardstown in fullfillment of the requirements for the degree of

Doctor of Philosophy

Principal supervisors:
Dr. Colm McGuinness & Dr. Philip Owende
Institute of Technology Blanchardstown, Dublin, Ireland

June 8th, 2015

To my husband Alan and children Sam and Sarah.

Acknowledgements

Thanks to Dr. Colm McGuinness and Dr. Philip Owende for their excellent supervision and guidance. Thanks to Dr. Brian Nolan, Dr. Aiden Carthy and Dr. Markus Hofmann for their support throughout this journey. I would like to acknowledge and thank Dr. Diarmuid O'Callaghan, Sandra McCullagh and my employer, the Institute of Technology Blanchardstown, for their support in facilitating this research. I also wish to thank the National Learning Network Assessment Services for their assistance in facilitating learner profiling during student induction.

Abstract

The application of data analytics to educational settings is an emerging and growing research area. Much of the published works to-date are based on ever-increasing volumes of log data that are systematically gathered in virtual learning environments as part of module delivery. This thesis took a unqiue approach to modelling academic performance; it is a first study to model indicators of students at risk of failing in first year of tertiary education, based on data gathered prior to commencement of first year, facilitating early engagement with at-risk students.

The study was conducted over three years, in 2010 through 2012, and was based on a sample student population (n=1,207) aged between 18 and 60 from a range of academic disciplines. Data was extracted from both student enrolment data maintained by college administration, and an online, self-reporting, learner profiling tool developed specifically for this study. The profiling tool was administered during induction sessions for students enrolling into the first year of study. Twenty-four factors relating to prior academic performance, personality, motivation, self-regulation, learning approaches, learner modality, age and gender were considered.

Eight classification algorithms were evaluated. Cross validation model accuracies based on all participants were compared with models trained on the 2010 and 2011 student cohorts, and tested on the 2012 student cohort. Best cross validation model accuracies were a Support Vector Machine (82%) and Neural Network (75%). The k-Nearest Neighbour model, which has received little attention in educational data mining studies, achieved highest model accuracy when applied to the 2012 student cohort (72%). The performance was similar to its cross validation model accuracy (72%). Model accuracies for other algorithms applied to the 2012 student cohort also compared favourably; for example Ensembles (71%), Support Vector Machine (70%) and a Decision Tree (70%).

Models of subgroups by age and by academic discipline achieved higher accuracy than models of all participants, however, a larger sample size is needed to confirm results. Progressive sampling showed a sample size > 900 was required to achieve convergence of model accuracy.

Results showed that factors most predictive of academic performance in first year of study at tertiary education included age, prior academic performance and self-efficacy. Kinaesthetic modality was also indicative of students at risk of failing, a factor that has not been cited previously as a significant predictor of academic performance.

Models reported in this study show that learner profiling completed prior to commencement of first year of study yielded informative and generalisable results that identified students at risk of failing. Additionally, model accuracies were comparable to models reported elsewhere that included data collected from student activity in semester one, confirming the validity of early student profiling.

Contents

List of Tables

List of Figures

List of Symbols, Acronyms & Abbreviations

BPNN	Back-propagation Neural Network
CAO	Central Applications Office; it has responsibility for processing applications for undergraduate courses in Higher Education Institutes in Ireland
DT	Decision Tree
FET	Fisher's Exact test
GM	Geometric Mean
GPA	Grade Point Average, an aggregate score measuring academic performance in year 1
IoT	Institutes of Technology (`http://www.ioti.ie`)
ITB	Institute of Technology Blanchardstown (`http://www.itb.ie`)
ITS	Intelligent Tutoring System
k-NN	k-Nearest Neighbour
LR	Logistic regression
LOF	Local outlier factor
m	Mean
$Model_{2012}$	Models trained on 2010 and 2011 participants and tested on 2012 participants
$Model_{all}$	Models trained on all 24 study factors
$Model_{NCog}$	Models trained on non-cognitive factors of learning, gender and age
$Model_{Prior}$	Models trained on attributes relating to prior academic performance, gender and age
$Model_{ran}$	Mean model accuracy based on 50 random samples selected from all participants. Models were trained on 2010 and 2011 participants and tested on 2012 participants
$Model_{sub}$	Models trained and tested on a subgroup of the dataset, either by course of study, age group or gender
$Model_{XVal}$	Models trained using 10-fold cross validation using stratified sampling
n	Sample size
NB	Naïve Bayes
NLN	National Learning Network (`http://www.nln.ie`)
r	Pearson product-moment correlation coefficients
R	The R software environment for statistical computing (`www.r-project.org`). Version 3.0.2 was used in this study
s	Standard deviation
SVM	Support Vector Machine
VLE	Virtual Learning Environment

Chapter 1

Introduction

1.1 Study background

It is increasingly evident that significant numbers of college students do not complete the courses on which they enrol, particularly for courses with lower entry requirements [ACT, 2012; Mooney et al., 2010]. Enrolment numbers to tertiary education are increasing, as is the diversity in student populations [OECD, 2013; Patterson et al., 2014]. This adds to the challenge of both identifying students at risk of failing, and provisioning appropriate supports to enable all students to perform optimally [Mooney et al., 2010]. Tertiary education providers collect a lot of data on students, particularly activity data from virtual learning environments and other online resources [Drachsler and Greller, 2012]. As a result, the application of data analytics to educational settings is emerging as an evolving and growing research discipline [Mirriahi et al., 2014; Sachin and Vijay, 2012; Siemens and Baker, 2012]. Its primary aim is exploring the value of data gathered in providing learning professionals, and students, with actionable information that could be used to enhance the learning environment [Chatti et al., 2012; Siemens, 2012]. A key challenge for learning analytics is the need to develop capability to explore and identify data that will contribute to improvement of learning models, including data that are currently not systematically gathered by tertiary education providers [Buckingham Shum and Deakin Crick, 2012; Tempelaar et al., 2013].

Learning is a latent variable, typically measured as academic performance in assessment work and examinations [Mislevy et al., 2012]. Factors impacting on academic performance have been the focus of research for many years, for example Allick and Realo [1997]; Farsides and Woodfield [2003]; Hembree [1988]; Lent et al. [1994]; Moran and Crowley [1979]; Powell [1973]. It still remains an active research topic [Buckingham Shum and

Deakin Crick, 2012; Cassidy, 2011; Jayaprakash et al., 2014; Komarraju and Nadler, 2013; Nandagopal and Anders Ericsson, 2012] indicating the inherent difficulty in both measurement of learning [Knight et al., 2013; Tempelaar et al., 2013], and modelling learning process, particularly in tertiary education [Pardos et al., 2011]. Cognitive ability remains an important determinant of academic performance [Cassidy, 2011], often measured as prior academic ability. Age has also been cited as significant [Naderi et al., 2009], as are data gathered from learner activity on online learning systems [Bayer et al., 2012; López et al., 2012]. In addition to the data systematically gathered by providers, there are other non-cognitive factors that can be measured prior to commencing tertiary education, which could be useful in modelling learner academic performance. For example, models predicting academic performance that include factors of motivation (e.g. self-efficacy, goal setting) with cognitive ability yield a lower error variance than models of cognitive ability alone (reviewed in Boekaerts [2001]; Robbins et al. [2004]). Research into personality traits, specifically the Big Five factors of openness, conscientiousness, extroversion, agreeableness and neuroticism, suggests some personality factors are indicative of potential academic performance [Chamorro-Premuzic and Furnham, 2004, 2008; De Feyter et al., 2012]. For example, Chamorro-Premuzic and Furnham [2008] found conscientiousness was correlated with academic performance, but not with IQ, suggesting conscientiousness may compensate for lower cognitive ability. Learning approach (deep or shallow) and self-regulated learning strategies are also relevant, and have been shown to mediate between other factors (such as factors of personality and factors of motivation) and academic performance [Biggs et al., 2001; Entwhistle, 2005; Swanberg and Martinsen, 2010].

Many publications emanating from educational psychology report on statistical analysis of academic performance metrics and their correlations with, or dependencies on, a wide variety of cognitive and non-cognitive psychometric factors of learning [Dekker et al., 2009; Herzog, 2006; Robbins et al., 2004]. However, measurement and analysis of non-cognitive factors of learning has received limited attention from the learning analytics community [Buckingham Shum and Deakin Crick, 2012] with the exception of factors inferred from online behaviour (e.g. [Ali et al., 2014; Shute and Ventura, 2013]). A range of non-cognitive psychometric factors have been associated with an effective learning disposition, such as a deep learning approach, ability to self-regulate, setting learning goals, persistence, conscientiousness and sub-factors of openness, namely intellectual curiosity, creativity and open-mindednesss [Buckingham Shum and Deakin Crick, 2012; Knight et al., 2013; Tishman et al., 1993]. An effective learning disposition describes attributes and behaviour characteristic of a good learner [Buckingham Shum and Deakin Crick, 2012]. Inclusion of non-cognitive factors of learning in models of academic performance can provide informative feedback on malleable, effective learner dispositions [Knight et al., 2013].

Early modelling of students at risk of failing informs provisioning of supports and modifications to learning environment, to enable more students perform optimally [Lauria et al., 2013]. Colby [2004] identified week two as a critical point in identifying at-risk students. Milne et al. [2012] reported successful results in predicting students at risk of failing based on analysis of online behaviour in week one. This thesis details development of an algorithmic model, based on cognitive and non cognitive factors of learning, that predicts students at risk of failing based on data gathered prior to commencement of first year of study [Gray et al., 2013, 2014a,b,c, 2015].

1.2 Rationale for research in this area

Social and policy changes in Ireland over recent years has resulted in greater access to tertiary education. For example, admissions rose from 20% of 17 to 18 year olds in 1980 to 55% in 2008 [Clancy and Wall, 2000; HEA, 2008]. More recent analysis predicts a 29% increase in enrolment numbers between 2013 and 2028 [Cassells, 2015]. The fourteen Institutes of Technology (IoT) in Ireland provide 41% of the third level education [Patterson et al., 2014], with a focus on the skill needs of the community they serve (`www.ioti.ie`). Undergraduate courses are offered at level 6 (2-year certificate), level 7 (3-year ordinary degree) and level 8 (4/5-year honours degree). Level 8 degrees have equivalent learning outcomes to university level 8 degrees, as specified by the National Framework of Qualifications (NFQ, `www.qqi.ie`).

The IoT student profile differs from the university student profile. Mooney et al. [2010, Appendix A] evidenced differences in prior academic ability: The majority of IoT students attain between 200 and 400 points in the Leaving Certificate exam, the state exam at the end of secondary school; The majority of university students attain over 400 points (up to the maximum score of 600 points). The 2012-13 annual report on tertiary students in Ireland concurred with previous years regarding greater diversity in the IoT student population compared to universities, in terms of age profile, ethnicity and socio-economic background [HEA, 2013; Patterson et al., 2014]. For example, 18% of new entrants to IoTs in 2012 were mature, compared with 9% of new entrants to universities. Results of a nation wide survey of undergraduate students (n=40,463) included in Patterson et al. [2014] showed 12% of IoT students were from a minority ethnic background compared to 8% of university students. In addition, higher percentage of IoT students are from a manual skilled (15% versus 9.5%) or unskilled (3.3% versus 1.6%) socio-economic group [Patterson et al., 2014, p. 26-27]. The report also noted a higher proportion of males

(66%) in the IoTs, genders were equally represented in the university sector in the same year.

Student drop out continues to be a challenge, particularly in the IoT sector where course entry requirements are lower than corresponding courses in the university sector. For example, the average drop out rate is 11% for level 8 courses across tertiary education in Ireland, however, the average drop out rate for level 8 course in the IoT sector is 16%, while the average drop out rate for level 7 courses is 26% [Mooney et al., 2010]. Under-performance and student drop-out has adverse consequences for the student, society and the education provider [Crosling and Heagney, 2009]. Degree completion has been shown to be an important determiner of occupational status and income [Pascarella and Terenzini, 2005, p. 373]. Degree completion also improves a student's sense of self-worth and enhances their academic and social competencies [Pascarella and Terenzini, 2005, p. 214], benefiting both the student and society [Crosling and Heagney, 2009]. Additionally, retention rates have a direct financial impact on college funding; specifically, annual core funding for the Irish IoT sector is based on the number of students registered on March 1st of the previous academic year.[1]

Tinto [2012] identified structured support during the early years of college life as one of the essential components in improving retention rates. Supports may be informed by early profiling of students at risk of failing. Dekker et al. [2009] and Smith et al. [2012] agreed that data already gathered by colleges, such as prior academic performance, age, and gender, was not sufficient on its own to identify at-risk students, additional factors were needed to improve predictive accuracy. For example, Dekker et al. [2009] improved model performance by including early tertiary level assessment results. The purpose of this study was to identify and profile first year IoT students at risk of failing at the end of year one of tertiary education, based on cognitive and non-cognitive factors of learning measured prior to commencement of first year course work.

1.3 Study aim and objectives

The aim of this study was to investigate if factors of learning, measured prior to commencement of first year, could accurately predict students at risk of failing at the end of year one of tertiary education. The ensuing six study objectives were:

Ob1. To research factors predictive of academic performance in tertiary education, with a focus on cognitive and non-cognitive factors of learning that can be measured prior

[1]Further details on funding are available at `www.hea.ie/en/funding/institutional-funding`.

to commencement of first year of study.

Ob2. Review data modelling techniques prevalent in educational psychology and educational data mining.[1]

Ob3. Develop a learner profiling tool for administration during first year student induction to collect data on non-cognitive factors of learning.

Ob4. Complete statistical analysis of cognitive and non-cognitive study factors for comparison with other published studies, and so validate early measurement of factors of learning.

Ob5. Train and evaluate a range of classification models predicting students at risk of failing.

Ob6. Identify the key cognitive and non cognitive factors of learning that are predictive of first year students at risk of failing.

1.4 Study research questions

The primary research question was:

Q1. Can algorithmic student modelling accurately predict Irish IoT students at risk of failing in first year of study based on factors that can be measured prior to commencement of tertiary education?

The thesis also addressed the following secondary research questions:

Q2. Which classification algorithms are appropriate for modelling psychometric data indicative of Irish IoT students at risk of failing?

Q3. Which cognitive and non-cognitive factors of learning are indicative of Irish IoT students at risk of failing?

1.5 Uniqueness of this study

This study adds to existing knowledge in a number of ways:

[1] Educational data mining (EDM) and Learning analytics (LA) are similar in terms of the data analysed. However, LA focuses on data exploration to empower students and educators while EDM focuses on model development to support automated delivery [Baker and Salvador, 2014].

1. This is a first study to predict students at risk of failing based on learner profiling completed prior to commencement of first year of study.
2. The unique focus on the Institute of Technology sector in Ireland incorporated a diverse student population in terms of age, socio-economic background and prior academic ability.
3. A comparison of classification algorithms highlighted k-nearest neighbour had good predictive accuracy, an algorithm that has received limited attention in learning analytics to date.
4. The study evaluated a range of non-cognitive factors of learning, including learner modality which has not been cited previously as predictive of academic performance.

The following sections elaborate on each of these.

1.5.1 Early models of academic performance

Many classification models of academic performance include factors relating to online or classroom activity arising from course work (e.g. Baker et al. [2011]; Lauria et al. [2012]; Romero et al. [2013]). In addition, studies that considered non-cognitive factors of learning typically gathered data after course commencement; there is some rationale for this timing. Instruments for non-cognitive factors of learning typically phrase items in the context of course work. For example, *'I expect to do very well on this course'* may be hard to evaluate prior to course commencement. Similarly, expressions such as *'I have a regular place set aside for studying'* may be better assessed after experience of college library facilities and time to establish a study pattern. However, as will be discussed in Sections 5.2 and 6.4, correlations between factors of learning and academic performance in this study were similar to results cited in other studies where data was gathered later in the semester. Therefore, this study concluded that measurement of factors of learning prior to commencement of first year of study can generate informative feedback on learner disposition.

1.5.2 The Institute of Technology sector

A study of Irish students in tertiary education by Bergin [2006] showed that factors predictive of university students at risk of failing differed from factors predictive of Institute of Technology (IoT) students at risk of failing. Her study focused on factors important for the subject of software programming in year 1 of study; this study extended that research to focus on factors predictive of students at risk of failing across a range of disciplines in the IoT sector. Evaluation of factors relevant to the academic performance of a student profile typical of the IoT sector has received limited attention to date.

1.5.3 Evaluation of classification models

Algorithmic modelling of static student data that will be discussed in Section 2.8.2 typically cites performance based on Naïve Bayes, Decision Tree, Neural Network, Support Vector Machine, Logistic Regression and their Ensembles. k-Nearest Neighbour (k-NN) has received little attention in educational data mining studies. However, results from this study highlighted that k-NN had consistently good performance both for models of all participants and models of subgroups by course of study, gender and age. In addition, k-NN results using cross validation had similar accuracies to models applied to a new student cohort, suggesting k-NN models generalise well.

1.5.4 Non-cognitive factors of learning

As will be illustrated in Table 2.11, studies that include non-cognitive factors of learning are often limited in the number of factors analysed. In comparison, the inclusion of fifteen non-cognitive factors of learning in this study represented a comprehensive analysis of a range of non-cognitive factors of learning. Salient outcomes from this study included:

- Kinaesthetic learner modality improved classification model accuracy, but has not been previously cited as a predictor of academic performance.
- Openness also improved classification model accuracy for a number of algorithms in spite of relatively low correlations with GPA.
- The difference in model accuracy with, and without, non-cognitive factors of learning was not statistically significant. Therefore, the value of including non-cognitive factors of learning was their contribution to feedback on learner disposition rather than improvement in model accuracy.

1.6 Structure of the thesis

This section outlines the structure of the thesis, and identifies where each of the six study objectives (*Ob1 - Ob6*) and three research questions (*Q1 - Q3*) will be addressed.

Chapter 2 will review published literature on both factors predictive of academic performance in tertiary education (*Ob1*) and the statistical and data mining approaches typically used to model educational data (*Ob2*). The evidence that will be presented confirms that both cognitive and non-cognitive factors of learning are correlated with academic performance in tertiary education. An argument for greater use of empirical data analysis on psychometric data will be discussed, followed by a review of classification algorithms prevalent in educational data mining research.

Chapter 3 will detail the study methods used. Study participants will be described, including descriptive statistics for each study factor. Addressing *Ob3*, development of the study's online profiler will be discussed including items used to measure fifteen non-cognitive factors of learning. Instrument reliability test results will indicate good reliability for thirteen of the fifteen non-cognitive factors. Issues with two factors, *intrinsic goal orientation* and *study time*, will be discussed both in Chapter 3 and in Chapter 6. The two statistical approaches used to meet *Ob4* will be detailed, namely correlation analysis and linear regression. Analysis methods to identify group differences will also be discussed. The eight classification algorithms used to model the data will be described. This will include: parameter tuning done; model evaluation metrics used; and statistical tests used to compare model performances, as required for *Ob5*. Attribute subset selection techniques used to address *Ob6* will also be discussed.

Data quality will be discussed in **Chapter 4** and justification provided for preprocessing decisions made. Discreatisation of academic performance to a binary class label that identifies at-risk students will be reviewed. Results from progressive sampling will indicate the sample size (n=1,207) will be sufficient for model accuracy convergence. Solutions to address class imbalance will be evaluated, and justification for attribute scaling using a standard normal Z-transformation will be provided.

Data analysis results will be presented in **Chapter 5**. Correlation and regression results concurred with other published results, differences relating to *self-efficacy* and *study time* will be highlighted (*Ob4*). To answer the three research questions, results from a range of algorithmic models predicting students at risk of failing will be presented, including models of all participants and models of subgroups by academic discipline, age and gender. Model generalisability will be evaluated by comparing cross validation accuracies with accuracies of models trained on the 2010 and 2011 students cohorts, and tested on the 2012 student cohort (*Ob5*). The study factors used by each model will also be presented (*Ob6*).

Chapter 6 will synthesise and analyse study results. This will include: a review of classification model accuracies (*Q1* and *Q2*); an evaluation of students misclassified (*Q1*); an overview of the usefulness of study factors in predicting students at risk of failing (*Q3*); and where justified, recommendations for future work. Study key findings and main conclusions will also be presented.

1.7 Publications emanating from this study

Invited book chapter:

Pb1. Gray, G., McGuinness, C. and Owende, P. (2016: In Press) Non-cognitive factors of learning as early indicators of students at-risk of failing in tertiary education In Khine, M. S. (Eds.) *Non-cognitive Factors and Educational Attainment*, Sense Publishers, Netherlands.

Peer-reviewed journal publications:

Pb2. Gray, G., McGuinness, C., Owende, P., and Hofmann, M. (2016) Learning Factor Models of Students at Risk of Failing in the Early Stage of Tertiary Education, *Journal of Learning Analytics*, 3(2):330-372, 2016. Available online at `https://files.eric.ed.gov/fulltext/EJ1126865.pdf`

Pb3. Gray, G., McGuinness, C., Owende, P., and Carthy, A. (2014) A review of psychometric data analysis and applications in modelling of academic achievement in tertiary education. *Journal of Learning Analytics*, 1(1):75-106, 2014. Available online at `https://learning-analytics.info/index.php/JLA/article/view/3255/4013`

Conference papers:

Pb4. Gray, G., McGuinness, C., and Owende, P. (2014) An application of classification models to predict learner progression in tertiary education. *4th IEEE International Advanced Computing Conference*, pages 549-554, February 2014. Available online at `http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6779384`

Pb5. Gray, G., McGuinness, C., and Owende, P. (2014) Non-cognitive factors of learning as predictors of academic performance in tertiary education. In Gutierrez-Santos, S and Santos O. C, editors, *WSEDM 2014 co-located with the 7th International Conference on Educational Data Mining (EDM 2014)*, London, July 4-7, 2014. Available online at `http://ceur-ws.org/Vol-1183/ncfpal2014_proceedings.pdf`

Pb6. G, Gray, C. McGuinness, and P. Owende (2013) An Investigation of Psychometric Measures for Modeling Academic Performance in Tertiary Education. *6th International Conference on Educational Data Mining (EDM 2013)*, Memphis, July, 2013. Available online at `http://www.educationaldatamining.org/conferences/index.php/EDM/2013/schedConf/presentations`

Pb7. Gray, G., McGuinness, C., and Owende, P. (2013) Investigating the efficacy of algorithmic student modelling in predicting students at risk of failing in tertiary education. *Young researcher track, 6th International Conference on Educational Data Mining (EMD 2013)*, Memphis, July, 2013. Available online at `http://www.educationaldatamining.org/conferences/index.php/EDM/2013/schedConf/presentations`

The abstracts from each study publication are included in Appendix A.

Chapter 2

Literature Review

2.1 Introduction

This chapter reviews a range of psychometric factors that could be used to predict academic performance in tertiary education. Specifically, four key areas are reviewed: aptitude, temperament, motivation and learning strategies. These were chosen on the basis of being directly or indirectly related to academic performance, and can be measured prior to, or during learner enrolment in tertiary education programmes. Available evidence on correlations between individual attributes and academic achievement is outlined; unless stated otherwise, studies cited were based on tertiary education. A review of pertinent data analysis techniques is also presented, with an emphasis on empirical modelling approaches prevalent in educational data mining.

The literature reviewed furthered four of the study's objectives (*Ob1* - *Ob4*) as follows:

- The review of factors predictive of academic performance supported the inclusion of both cognitive and non-cognitive factors of learning in models predicting students at risk of failing (*Ob1*). This in turn informed the selection of non-cognitive factors for the study's learner profiling tool (*Ob3*).
- A review of published statistical analysis results pertaining to factors relevant to this study informed comparisons between results in this study and other published results (*Ob4*).
- A review of data analysis approaches prevalent in educational data mining supported adopting an algorithmic modelling approach to predict students at risk of failing, specifically the use of classification algorithms (*Ob2*).

2.2 Measurement of cognitive ability and correlation with academic performance

Cognitive ability tests were originally developed to identify low academic achievers [Jensen, 1981; Munzert, 1980]. The first such test measured general cognitive intelligence, g, as identified by Spearman [1904, 1927]. Test results for an individual across a range of cognitive measures tend to correlate providing good evidence for a single measure of intelligence [Jensen, 1981; Kuncel et al., 2004]. In addition to a general cognitive intelligence, there is widespread evidence for a multi-dimensional construct of intelligence comprising of a range of sub factors [Flanagan and McGrew, 1998]. Abilities in such sub factors vary from one individual to another, and vary within an individual across factors, in other words an individual can have higher ability in one sub factor than in another [Spearman, 1927, p. 75]. Recently the Cattell-Horn-Carroll (CHC) theory of cognitive abilities has gained recognition as a taxonomy of cognitive intelligence [McGrew, 2009]. The CHC is based on ten broad cognitive categories, summarised in Table 2.1. As illustrated in Table 2.2, cognitive ability tests vary in terms of the cognitive categories they measure, but typically include Crystallised Intelligence (Gc), Short-Term Memory (Gsm) and Visual processing (Gv) [Flanagan and McGrew, 1998].

Cognitive ability tests have been criticised on the basis of what is being measured. Sternberg [1999] asserts that intelligence tests measure a developing expertise rather than a stable attribute, and the typically high correlation between intelligence scores and academic performance is because they measure the same skill set rather than evidencing a causal relationship. In an analysis of a range of IQ studies measuring IQ trends across two generations, Flynn [1987] identified a significant rise in IQ from one generation to the next. Since the observation (Flynn effect) is unlikely to be due to genetic changes in such a short period of time, it would appear to be the result of acquired skills that improve performance in IQ tests by subjects with the same IQ as the parent generation. This view is supported by other studies that have compared children in western and non-western standards of education. These have shown that children tended to score well on tests which measured skills that are valued by their parents [Sternberg, 1999, p. 8]. It is notable that correlations between general intelligence and academic performance are stronger at secondary level than tertiary level education [Bartels et al., 2002; Cassidy, 2011; Colom and Flores-Mendoza, 2007; Eysenck, 1994; Matarazzo and Goldstein, 1972]. Therefore, prior academic performance such as high school GPA (HSGPA), and / or standardised tests like

Table 2.1: Defining broad factors of intelligence [Flanagan and McGrew, 1998]

Factor	Symbol	Description
Fluid Intelligence	Gf	Ability to solve problems independently of knowledge learned.
Crystallised intelligence	Gc	Acquiring and organising knowledge and skills, and ability to use such knowledge in solving problems.
Visual processing	Gv	Ability to process and analyse visual information
Auditory Processing	Ga	Ability to process and analyse auditory information
Processing Speed	Gs	Ability to perform automatic cognitive tasks quickly (measure in minutes)
Reaction Time / Decision speed	Gt	Speed at which an individual can react to a stimulus, or make decisions (measure in seconds).
Short-Term Memory	Gsm	Ability to hold information with immediate awareness and use it again within a few seconds
Long-Term Retrieval	Glr	Ability to store and retrieve information over a longer period of time.
Quantitative Knowledge	Gq	Ability to understand quantitative concepts and relationships, and work with numeric symbols. This is a measure of mathematical knowledge acquired, as distinct from mathematical reasoning (Gf).
Reading-Writing	Grw	Basic reading and writing skills (considered by Cattell-Horn to be part of Gc)

American College Testing (ACT)[1] scores and Scholastic Aptitude Test (SAT)[2] scores are frequently used as measures of cognitive ability when modelling academic performance in tertiary education.

Table 2.3 shows that correlations between ability and academic performance in tertiary education are consistent and relatively strong for studies of standard students. For example, a meta-analysis of 109 studies conducted by Robbins et al. [2004] found average correlation between academic performance and SAT scores was r=0.388 (90% CI [0.353, 0.424]) and correlation between academic performance and HSGPA has a marginally higher (r=0.448, 90% CI [0.409, 0.488]). Eppler and Harju [1997] found that correlations between academic performance and SAT scores were not as strong for mature students. Brady-Amoon and Fuertes [2011] attribute their lower correlations[3] (r=0.16, 90% CI* [0.061,

[1]ACT tests are based on high school curriculum in English, Mathematics, Reading and Science (`www.act.org`).

[2]SAT measures general intelligence in addition to mathematicss and verbal subscales [Frey and Detterman, 2003]. Frey and Detterman [2003] found scores were highly correlated with IQ (r=0.820, p<0.001).

[3]CI* denotes confidence intervals were not provided by the author, and were calculated in R version 3.0.2 using CIr in package *psychometric* which calculates confidence intervals based on a Fisher r-to-z transformation.

Table 2.2: Factors of intelligence measured by popular cognitive aptitude tests [Flanagan and McGrew, 1998]

Factor	Gc	Gf	Gv	Ga	Gs	Gt	Gsm	Glr	Gq	Grw
WJ-R	*	*	*	*	*		*	*	*	
WAIS	*		*		*		*		*	
SB4	*	*	*				*		*	
DAS	*	*	*				*			
K-ABC	*		*				*		*	
KAIT	*	*					*	*		

WJ-R: Woodcock-Johnson-Revised; WAIS: Wechslers Adult Intelligence Scale; SB4: Stanford Binet Intelligence Scale, 4th Ed.; DAS: Differential Ability Scale; K-ABC: Kaufman Assessment Battery for Children; KAIT: Kaufman Adolescent and Adult Intelligence Test.

Table 2.3: Correlations between cognitive ability and academic performance

Study	n	Age	Academic performance	g	SAT/ ACT	Prior ability
[Brady-Amoon and Fuertes, 2011]	271	m=21.26	GPA, self reported			0.16
[Cassidy, 2011]	97	m=23.5	GPA			0.52***
[Chamorro-Premuzic and Furnham, 2008]	158	m=19.2	GPA	0.24*		
[Conrad, 2006]	300	m=19.5	GPA, self reported		0.28*	
[Duff et al., 2004]	146	m=24.3	GPA			0.27*
[Eppler and Harju, 1997]	212	m=19.2	GPA		0.37***	
[Eppler and Harju, 1997]	25	m=29.8	GPA		0.09	
[Furnham et al., 2006]	64	[20-55]	Mean exam results	0.29*		
[Kaufman et al., 2008]	315	m=25.9	GPA			0.28***
[Kobrin et al., 2008]	151,316	18+	GPA		0.35***	0.36***
[Ning and Downing, 2010]	581	m=20.2	GPA			0.10*
[Robbins et al., 2004]	Meta-analysis		GPA		0.39	0.45

*$p<0.05$; **$p<0.01$; ***$p<0.001$; m=mean.

0.256], n=271) to the fact that study participants included a more diverse group of students from a variety of ethnic backgrounds, thereby supporting the findings of Schmitt et al. [2009, p. 34] that the interaction between prior academic ability and GPA differs for students from different ethnic groups. The lower correlations reported by Ning and Downing [2010] (r=0.1, $p<0.05$, 90% CI* [0.032, 0.167], n=581) could be attributed to their measure of prior academic performance, which was based on A level[1] scores in two subjects chosen by the student. The relatively high level of correlation reported by Cassidy [2011] could be attributed to a difference in how prior academic performance is measured. Cassidy used GPA accrued in the first year of study as a measure of prior academic performance to predict students' final degree GPA.

2.3 Measurement of temperament and correlation with academic performance

Theories of temperament focus on aspects of personality that are discernible at birth [John et al., 2008]. Historically, research linking temperament with academic achievement has lacked in well-defined referential framework for the interactions between temperament and academic performance. Studies have varied in their perspective of personality, with diverse views on the relevant traits to be considered as measures of temperament, such as factors of persistence, factors relating to motivation and/or moral factors such as honesty [de Raad and Schouwenburg, 1996]. While there are many factors associated with temperament, factor analysis by a number of researchers, working independently and using different approaches, has resulted in broad agreement of five main personality dimensions [Ackerman and Heggestad, 1997; John et al., 2008]. These are commonly referred to as the Big Five [Cattell and Mead, 2008; Goldberg, 1992, 1993; Tupes and Cristal, 1961] or the related Five-Factor Model [Costa and McCrae, 1992]. The five factors include: openness, agreeableness, extraversion, conscientiousness and neuroticism, and are described in Table 2.4. While the Big Five concept is empirical rather than a theory of personality [John and Srivastava, 1999], good reliability and consistency has been reported [de Raad and Schouwenburg, 1996; John et al., 2008].

Chamorro-Premuzic and Furnham [2004] found that personality attributes measured using the Big Five construct accounted for up to 30% of the variance in academic performance at tertiary level education. There is a consensus across studies that conscientiousness is the best personality based predictor of academic performance [O'Connor

[1] Hong Kong's secondary school termination exam. Students can select from a range of subjects.

Table 2.4: Big Five personality dimensions and their labels in three commonly used scales [Cattell and Mead, 2008; de Raad and Schouwenburg, 1996; Goldberg, 1993]

Big Five (Costa and McCrae)	16PF (Cattell)	Five-Factor Model (Goldberg)	Explanation of each factor
Extraversion	Introversion/ Extraversion	Surgency	Tendency to move towards, or away from human interaction.
Neuroticism	Low Anxiety/ High Anxiety	Emotional stability	Temperamental, moody, nervousness.
Openness	Tough-Mindedness/ Receptivity	Intellect or culture	Openness to feelings, emotions, new ideas and imagination. Curiosity. Creativity.
Agreeableness	Independence/ Accommodation	Agreeableness	Kindness, trust, warmth.
Conscientiousness	Self-Control/ Lack of Restraint	Conscientiousness or dependability	Organised, thorough, reliable, work ethic.

and Paunonen, 2007; Swanberg and Martinsen, 2010], as illustrated in Table 2.5. Many researchers have cited conscientiousness as compensating for lower cognitive intelligence (see Chamorro-Premuzic and Furnham [2004, 2008]; Trapmann et al. [2007]) and it is a consistent predictor of academic performance across assessment type [Allick and Realo, 1997; Kappe and van der Flier, 2010; Shute and Ventura, 2013].

Some significant correlations between openness and academic performance have been reported, but correlations with academic performance are not as high as conscientiousness (see Table 2.5). Openness is considered by Chamorro-Premuzic and Furnham [2008] to be a mediator between ability and academic performance. Openness in turn is mediated by learning approach, with open personalities being more likely to adopt a deep learning strategy, which in turn improves academic performance [Swanberg and Martinsen, 2010]. Sub-factors of openness, namely intellectual curiosity, creativity and open-mindedness, have been associated with effective thinking and learning dispositions [Buckingham Shum and Deakin Crick, 2012; Tishman et al., 1993]. Knight et al. [2013] argued that assessment design should nurture such dispositions. Kappe and van der Flier [2010] found that open personalities tend to do better when assessment methods are unconstrained by submission rules.

The relationship between neuroticism and academic performance is not as strong, and like openness, is influenced by assessment type. Neuroticism can have a negative impact on academic performance in stressful examination conditions such as end of year exams with time limitation [Hembree, 1988]. Where academic performance is measured under less stressful conditions such as continuous assessment work, the relationship between neu-

Table 2.5: Correlations between personality and academic performance

Study	n	Age	Academic performance	Conscientous	Open	Extrovert	Neuroticism	Agreeable
[Chamorro-Premuzic and Furnham, 2008]	158	[18,21]	GPA	0.37**	0.21**	0.16	-0.05	0.02
[Chamorro-Premuzic and Furnham, 2003]	70	[17,21]	Grades, year 1	0.33**	-0.06	0.05	-0.28**	0.34**
[Conrad, 2006]	300	m=19.48	GPA, self reported	0.35*	-0.02	0.00	-0.60	0.11
[Dollinger et al., 2008]	338	m=21.9	Exam and project	0.11*	0.14*	-0.10	0.07	-0.05
[Duff et al., 2004]	146	[17,52]	GPA	0.21	0.06	0.06	-0.13	0.12
[Gray and Watson, 2002]	300	[18,21]	GPA, self reported	0.36*	0.18*	-0.09	0.00	0.15*
[Kappe and van der Flier, 2010]$^{+}$	133	[18,22]	GPA	0.46**	-0.08	0.05	-0.06	0.14
[Kaufman et al., 2008]$^{++}$	315	m=23.5	GPA	0.18	0.12	0.03	0.07	0.06
[Komarraju et al., 2011]	308	[18,24]	GPA, self reported	0.29**	0.13*	0.07	0.00	0.22**
[O'Connor and Paunonen, 2007]	Meta-analysis		Various	0.24	0.05	-0.05	-0.03	0.06
[Trapmann et al., 2007]	Meta-analysis		GPA	0.22	0.08	0.01	-0.04	0.04

*p<0.05; **p<0.01.
$^{+}$Matched exam technique to personality type.
$^{++}$Used emotional stability, the reverse of neuroticism.

roticism and academic performance is less well defined [Chamorro-Premuzic and Furnham, 2006, p. 75]. Kappe and van der Flier [2010] found neuroticism to be positively correlated with academic performance (r=0.18, 90% CI* [0.038, 0.315], n=133) when assessment is free from time constraints and supervision.

Research is inconsistent regarding the remaining two personality dimensions of extraversion and agreeableness and their relationship with academic performance. Introverts tend to have better study habits and are less easily distracted ([Entwistle and Entwistle, 1970] cited by [Chamorro-Premuzic and Furnham, 2006, p. 78]), while extraverts tend to perform better in class participation, oral exams, seminar presentations and multi-choice style questions [Furnham and Medhurst, 1995; Kappe and van der Flier, 2010]. In their meta-analysis of a number of studies investigating personality as a predictor of academic performance, O'Connor and Paunonen [2007] concluded agreeableness is not associated with academic performance. Farsides and Woodfield [2003] found that agreeableness, while not related to academic performance, was linked to other performance indicators such as attendance record. Chamorro-Premuzic and Furnham [2003] agreed, and found high correlations between academic performance and agreeableness were not replicated in later years of the study, but agreeableness was highly correlated with absenteeism in the first year of study.

The literature reviewed suggested the two most pertinent factors of personality for learner profiling are: Conscientiousness, as the best personality based predictor of academic performance; and openness because of its association with an effective learning disposition. There are a range of assessment methods used in ITB[1]; therefore, neuroticism was less likely to be a useful predictor of academic performance in this study.

2.4 Theories of motivation that relate to academic performance

Ryan and Deci [2000] define motivation simply as being 'moved to do something'. Defining how learners are motivated to behave in a certain way, and more specifically to learn, is more complex, and is characterised by a range of complementary theories which aim to explain both the level of individual motivation and the nature of the motivation [Steel and Konig, 2006]. Current theories in turn encompass a number of factors, some of which are relevant, directly or indirectly, to academic performance [Robbins et al., 2004]. Informed by the categorisation of motivation theories relevant to academic achievement proposed by Robbins et al. [2004], the following sections discuss three such theories, relating to expectancy, goals and needs.

2.4.1 Measurement of expectancy motivation and correlation with academic performance

Expectancy models of motivation explore the extent to which a person regards outcome as being a consequence of behaviour. Levels of expectancy motivation are influenced by the extent to which a person believes they are in control of the outcome (locus of control) [Cassidy, 2011]. There are two strands of expectancy motivation [Eccles and Wigfield, 2002; Pintrich and DeGroot, 1990]:

[1] The weighting given to continuous assessment (CA) varies between modules on a course, from 20% CA to 100% CA.

1. Outcome Expectation refers to a belief that a particular behaviour will lead to a particular outcome, e.g *active engagement in class work results in better grades.*

2. Self-efficacy refers to a person's belief that they can achieve that outcome e.g. *I can actively engage in class and so I can achieve better grades.* High self-efficacy is associated with setting more challenging goals, a willingness to work hard and persistence with a task.

Table 2.6 gives a summary of correlations found between expectancy motivation and academic performance. A meta analysis of a selection of studies investigating relationships between expectancy motivation and academic performance found correlations varied between 0.38 and 0.5 [Brown et al., 2008]. A number of studies found self-efficacy specifically to be a useful predictor of academic performance [Brady-Amoon and Fuertes, 2011; Cassidy, 2011; Yusuf, 2011]. Indirect relationships between self-efficacy and academic performance mediated by either other motivational factors or learning strategies are also cited [Brown et al., 2010; Yusuf, 2011]. On the other hand, Pintrich and DeGroot [1990] found that self-efficacy was not significantly related to performance when cognitive engagement variables such as engagement in the learning process, self-regulation and learning strategies were also considered, thereby concluding that self-efficacy facilitates cognitive engagement, but cognitive engagement itself is more directly linked to academic performance. Nevertheless, study results suggest self-efficacy is an important indicator of motivation in tertiary education.

Table 2.6: Correlations between expectancy motivation and academic performance

Study	n	Age	Academic performance	Self-efficacy	Outcome expectancy
[Brady-Amoon and Fuertes, 2011]	271	m=21.3	GPA	0.22*	
[Bruinsma, 2004]	117	18	Credits, year 1	0.26**	
[Cassidy, 2011]	97	m=23.5	GPA	0.40***	0.20
[DiBenedetto and Bembenutty, 2013]	113	18+	Module grade	0.37**	0.08
[Diseth, 2011]	177	m=21.2	Specific exam	0.44**	
[Klassen et al., 2008]	261	m=23.3	GPA, self reported	0.36**	
[Komarraju and Nadler, 2013]	257	m=20.5	GPA	0.30**	0.14*
[Robbins et al., 2004]	Meta-analysis		GPA	0.50	

*p<0.05; **p<0.01; ***p<0.001.

2.4.2 Measurement of achievement goals and correlation with academic performance

High self-efficacy is associated with a student setting challenging goals in terms of their academic achievements. Achievement goals fall into two categories: performance goals where an individual is looking for favourable feedback, and learning goals where an individual is looking to increase their competency [Covington, 2000; Dweck, 1986; Dweck and Leggett, 1988; Eccles and Wigfield, 2002; Eppler and Harju, 1997]. Performance oriented goals are associated with a tendency to engage in tasks in which a student is guaranteed to excel, and avoid tasks that may highlight incompetence [Dweck, 1986]. This approach can inhibit a student from challenging and enhancing existing competencies. It is also associated with superficial cognitive processing and inefficient use of study time [Covington, 2000]. Learning goals are motivated by the need or desire to increase existing competencies and master new skills and, therefore, tend to be more challenging in nature [Covington, 2000]. Learning goals are associated with high self-efficacy, a belief that ability is dynamic, and a belief that increased effort will result in increased success (outcome expectancy). This is regarded as an important learning disposition [Buckingham Shum and Deakin Crick, 2012]. Interestingly, Dweck and Leggett [1988] found that there was no relationship between a child's academic ability (at age 14) and his or her goal orientation. Instead, goal orientation was influenced by the perception of ability as being fixed (resulting in a performance goal orientation) or dynamic (resulting in a learning goal orientation).

Studies have found learning goals to be more strongly correlated with academic performance than performance goals (see Table 2.7). A contributing factor to the exception in the study conducted by Diseth [2011] could be in how academic performance was measured. Unlike the other cited studies, it was based on an exam grade (A-F) from a single six-hour exam. Eppler and Harju [1997] found a statistically significant difference in the average GPA of students with high learning goals (some of whom also had high performance goals) and those with both low learning goals and low performance goals, with learning goals accounting for 9% of the variance in academic performance. They also found older students to be stronger in their endorsement of learning goals, while younger students tended towards performance oriented goals.

Table 2.7: Correlations between achievement goals and academic performance

Study	n	Age	Academic performance	Learning goals	Performance goals
[Diseth, 2011]	177	m=21.2	Single exam grade	0.21**	0.39**
[Dollinger et al., 2008]+	338	m=21.9	Exam performance	0.21**	
[Eppler and Harju, 1997]	212	m=19.2	GPA	0.30***	0.13
[Eppler and Harju, 1997]	50	m=29.8	GPA	0.28*	0.08
[Robbins et al., 2004]+	Meta-analysis		GPA	0.18	
[Wolters, 1998]	115	m=19.1	Average grade	0.36***	-0.21*

*$p<0.05$; **$p<0.01$; ***$p<0.001$; m=mean.
+These studies cited correlations for achievement motivation in general, rather than learning or performance goals specifically.

2.4.3 Measurement of self-determination and correlation with academic performance

Self Determination Theory (SDT) focuses on our innate psychological need for competency [Deci and Ryan, 2000], and aims to explore the difference in the types of goals learners adopt, and the justification. SDT distinguishes between intrinsic motivation, where motivation arises from enjoyment of activity, and extrinsic motivation, where the outcome is attractive [Ryan and Deci, 2000]. It has been argued that this is one factor represented as a continuum from an intrinsic, behaviour oriented state, to an extrinsic, goal oriented state [Apter, 1989; Entwhistle, 2005]. Alternatively, SDT has been viewed as two separate factors that can both be present [Dweck and Leggett, 1988; Eppler and Harju, 1997]. Individuals can alter between intrinsic or extrinsic motivation, depending on the time or situation, but will generally be predisposed to one or the other [Apter, 1989]. Cury et al. [2002] found that both performance and learning goals are associated with improving a student's level of intrinsic motivation. For more detailed discussions see Apter [1989]; Entwhistle [2005] and Ryan and Deci [2000].

Correlations with academic performance tend to be higher for intrinsic motivation than extrinsic motivation, but self-determination is not as strong, or as consistent, a predictor of academic performance as either self-efficacy or learning goals (see Table 2.8). Goodman et al. [2011] found both intrinsic and extrinsic motivation to be significantly correlated with academic performance, however the selection of participants in this study could have introduced bias. Students were invited to take part by email, with responders being entered into a prize draw. There was a 6.3% response rate. Komarraju et al. [2009] found significant correlation between intrinsic motivation and academic performance in a study of participants from a variety of disciplines (n=308). The study included three sub fac-

Table 2.8: Correlations between self-determination and academic performance

Study	n	Age	Academic performance	Intrinsic motivation	Extrinsic motivation
[Bruinsma, 2004]	117	m=18	credits, year 1	0.09	
[Goodman et al., 2011]	254	[17,29]	GPA	0.28**	0.21**
[Kaufman et al., 2008]	315	m=25.9	GPA	0.08	-0.05
[Komarraju et al., 2009]	308	[18,24]	GPA, self reported	0.20**	0.11
[Komarraju and Nadler, 2013]	257	m=20.5	GPA, self reported	0.11	0.05
[Wolters, 1998]	115	m=19.1	average grade	0.14	0.05

**p<0.01.

tors of intrinsic motivation from the Academic Motivations Scale (AMS), motivation to know (r=0.17, p<0.01, 90% CI* [0.077, 0.26]), motivation to accomplish (r=0.22, p<0.01, 90% CI* [0.129, 0.308]) and motivation to experience stimulation (r=0.13, p<0.05, 90% CI* [0.037, 0.221]). In a later study, Komarraju and Nadler [2013] found the correlation between intrinsic motivation and GPA was not significant when using a shorter 4-item scale to measure intrinsic motivation, the Motivated Strategies for Learning Questionnaire (MSLQ, Pintrich et al. [1991]). Kaufman et al. [2008] in a study of non-standard students from a diversity of ethnic backgrounds, did not find correlations to be significant, suggesting that factors impacting on academic performance can vary for different student groups.

2.4.4 Causal relationships between motivation and academic performance

While many studies cite correlations between academic performance and various measures of motivation, particularly self-efficacy, learning goals and intrinsic motivation, evidence supporting causal relationships between motivation and academic performance is less consistent, and is influenced to some extent by the selection of factors included in any specific study. For example, Chamorro-Premuzic and Furnham [2003] and Brown et al. [2010] found motivation was a mediator between conscientiousness and performance, while Komarraju et al. [2009] found conscientiousness mediated between intrinsic motivation and performance. Komarraju et al. [2009] also report that motivation did not account for any additional variance on academic performance beyond what was already explained by the Big Five. Brown et al. [2008], on the other hand, in a study not including personality factors, found self-efficacy did have a causal relationship with academic performance. Diseth [2011] and Sins et al. [2008] found learning strategy mediated between motivation (specifically self-efficacy and performance goals) and academic performance. Robbins et al. [2004] found self-efficacy and achievement motivation to be the best predictors of GPA attained by learners. A number of studies investigating both personality and motivation argue that personality based factors are a better predictor of academic performance than motivation [De Feyter et al., 2012; Komarraju et al., 2009]. However, Zuffianó et al. [2013] found that self-efficacy significantly contributed to the explained variance in academic performance over and above ability and personality. It also has a more practical value in that self-efficacy beliefs are more easily changed than ability or personality. This would suggest that while there are correlations between factors of personality and motivation, measures of personality, particularly conscientiousness, and measures of motivation, particularly self efficacy and achievement goals, each have value, and are worth further consideration in models of student learning. Therefore, it was decided to consider both factors of personality and factors of motivation in this study, specifically conscientiousness, openness, self-efficacy, learning goals and achievement goals. Factors of self-determination were not included as cited relationships with academic performance were weaker than other measures of motivation.

2.5 Learning strategies

A number of studies found the relationship between academic performance and temperament or motivation is mediated by a students approach to the learning task itself. Important factors include learning approach (e.g. Bruinsma [2004]; Chamorro-Premuzic and Furnham [2008]; Diseth [2011]; Sins et al. [2008]) and self-regulation (e.g. Nasiriyan et al. [2011]; Ning and Downing [2010]). The following sections discuss both learning approach and self-regulation.

2.5.1 Measurement of learning approach and correlation with academic performance

Learning approach has its foundations in the work of Marton and Säljö [2005] who classified learners as shallow or deep. Deep learners aim to understand content, while shallow learners aim to memorise content regardless of their level of understanding. Later studies added strategic learners [Entwhistle, 2005, p. 19], whose priority is to do well, and will adopt either a shallow or deep learning approach depending on the requisites for academic success. Both personality and self-determined motivation are indicative of personal approaches to learning. Openness, conscientiousness and intrinsic motivation are correlated with a deep learning approach, while neuroticism, agreeableness and extrinsic motivation are associated with a shallow learning approach [Busato et al., 1999; Duff et al., 2004;

Marton and Säljö, 2005].

Many studies concur with a negative correlation between a shallow learning approach and academic performance (see summary in Table 2.9). Some studies show higher correlations with a deep learning approach (e.g. Chamorro-Premuzic and Furnham [2008]; Snelgrove [2004]), while others cite marginally higher correlations with a strategic learning approach (e.g. Cassidy [2011]; Duff et al. [2004]). Volet [1996] found the importance of learning approach varied with assessment type. A lack of correlation between a deep learning approach and academic performance is in itself an insightful result, as it suggests an assessment design that fails to reward an important, malleable learning disposition [Buckingham Shum and Deakin Crick, 2012; Knight et al., 2013], and hence, may elicit secondary, follow-up actions to improve assessment design. The literature reviewed supported the inclusion of learning approach in learner profiling.

Table 2.9: Correlations between learning approach and academic performance

Study	n	Age	Academic performance	Deep	Shallow	Strategic
[Cassidy, 2011]	97	m=23.5	GPA	0.31**	-0.01	0.32**
[Chamorro-Premuzic and Furnham, 2008]	158	m=19.2	GPA	0.33**	-0.15	0.18*
[Duff et al., 2004]	146	[17,52]	GPA	0.10	-0.05	0.15
[Snelgrove, 2004]	289	18+	GPA	0.20*	-0.13	0.17*
[Swanberg and Martinsen, 2010]	687	m=24.5	Single exam	0.16	-0.25	

*p<0.05; **p<0.01.

2.5.2 Measurement of self-regulation and correlation with academic performance

Self-regulated learning is recognised as a complex concept as it overlaps with a number of other concepts including temperament, learning approach and motivation, specifically self-efficacy and goal setting [Bidjerano and Dai, 2007; Boekearts, 1996]. While many students may set goals, ability to self-regulate learning can be the difference between achieving, or not achieving, goals set [Covington, 2000]. Self-regulated learners take responsibility for setting and achieving their own learning goals. This is done by planning their learning, having effective time management, using appropriate learning strategies, continually monitoring and evaluating the quality of their own learning (metacognitive self-regulation) and altering their learning strategies when required [Schunk, 2005; Zimmerman, 1990]. Such learners regard learning as a process that they can control, but their motivation factors can vary [Pintrich and DeGroot, 1990]. To be motivated to self-regulate, a student must be confident that they are able to set goals and organise their study, and in addition be confident that the effort they spend on studying will result in good marks (high self-efficacy). Such learners must also accept delayed gratification as self-regulation requires students to focus on long-term gains for their efforts [Bembenutty, 2009; Zimmerman, 1990; Zimmerman and Kitsantas, 2005]. Volet [1996] argues that self-regulated learning is more significant in tertiary level than earlier levels of education because of the shift from a teacher-controlled environment to expected self-management of the learners own study. Furthermore, Nicol and MacfarlaneDick [2006] argues that both training and formative feedback can improve self-regulation, resulting in a more effective learning disposition.

A number of studies cite significant correlations between academic performance and factors of self-regulation, see Table 2.10 for a summary. For example, a longitudinal study of first year students (n=581) found self-test strategies (r=0.48, p<0.001, 90% CI* [0.426, 0.531]) and monitoring strategies (r=0.42, p<0.001, 90% CI* [0.426, 0.531]) were more strongly correlated with academic performance than time & effort strategies (r=0.24, p<0.01, 90% CI* [0.175, 0.303]) [Ning and Downing, 2010]. However, Komarraju and Nadler [2013] found effort management (r=0.39, p<0.01, 90% CI* [0.299, 0.474], n=257) had higher correlation with academic performance than other measures of self-regulation and found that self-regulation (monitoring and evaluating learning) did not account for any additional variance in academic performance over and above self-efficacy, but study effort and time did account for additional variance. In a longitudinal study on the causal dilemma between motivation and self-regulation, De Clercq et al. [2013] concluded that a learning goal orientation results in a deep learning approach, which in turn results in better self-regulation. A study comparing the relative importance of both learning approach (deep or shallow) and learning effort, found that learning effort had a higher impact on academic performance than learning approach [Volet, 1996].

All correlations cited in Table 2.10 were statistically significant. The available evidence supported the inclusion of effort management and time management in learner profiling in addition to factors of motivation and learning approach. Metacognitive self-regulation was also included because of the importance of self-evaluation as an effective learning disposition.

2.6 Regression models of academic performance

Table 2.11 presents examples of regression models with variance in academic performance expressed as the co-efficient of determination (R^2). R^2 was relatively high for studies

Table 2.10: Correlations between self-regulation and academic performance

Study	n	Age	Academic performance	Effort regulation	Time management	Metacognitive self-regulation
[Bidjerano and Dai, 2007]	217	m=22	GPA, self reported	0.33**	0.23**	
[Dollinger et al., 2008]	338	m=21.9	Exam performance		0.21**	
[Goodman et al., 2011]	254	[17,29]	GPA	0.28**		
[Komarraju and Nadler, 2013]	257	m=20.48	GPA	0.39**	0.31**	0.14*
[Ning and Downing, 2010]$^+$	581	m=20.24	GPA		0.24**	0.42**
[Sundre and Kitsantas, 2004]	62	[18,24]	Single essay			0.43**

*p<0.05; **p<0.01.
$^+$ This study combined time management with concentration (effort regulation).

that included factors of cognitive ability combined with either factors of personality or motivation, along with some additional factors such as age and time spent studying. Cassidy [2011] reported an adjusted co-efficient of determination[1] ($\bar{R}^2$) of 0.53 (R^2=0.56) in a regression model including prior academic performance, self-efficacy and age (n=97). However, the relatively high R^2 may be due to the measure of prior academic performance used (first year GPA). Chamorro-Premuzic and Furnham [2008] reported $\bar{R}^2$=0.40 (R^2=0.41) in a regression model that included prior academic ability, personality factors and a deep learning strategy. A similar proportion of variance was reported by Dollinger et al. [2008] (R^2=0.44) in a regression model including prior academic ability, personality factors, academic goals and study time.

Not all studies concurred with these results. Both Kaufman et al. [2008] and Swanberg and Martinsen [2010] accounted for lower levels of variance when modelling non-standard students. Kaufman et al. [2008] reported R^2=0.14 in a model with prior academic performance, personality factors and self-determined motivation, when modelling students from a variety of ethnic backgrounds. Swanberg and Martinsen [2010] reported R^2=0.21 in a model with prior academic performance, personality, learning strategy, age and gender, when modelling students with an older average age (m=24.8). Lower variances were also reported in studies not including cognitive ability. Komarraju et al. [2011] reported R^2=0.15 in a model including personality and learning approach. Eppler and Harju [1997] reported R^2=0.12 in a model including factors of motivation and work commitments,

[1] An adjusted co-efficient of determination ($\bar{R}^2$) compensates for the automatic increase in R^2 when additional dependent variables are added to the model.

Table 2.11: Regression models of academic performance with significant standardised coefficients (p<0.05)

				Ability		Personality				Motivation			Learning Approach			Self-regulation		Other factors		
Study	n	Age	R^2	g	prior	C	O	N	A	SE	IM	EM	De	Sh	St	Effort	Time	Age	Gender	Job
[Bidjerano and Dai, 2007]	217	m=22	0.18	0.28												0.27				
[Bidjerano and Dai, 2007]	217	m=22	0.11			0.14										0.31				
[Chamorro-Premuzic and Furnham, 2008]+	158	m=19.2	0.41	0.29		0.49							0.21							
[Cassidy, 2011]+	97	m=23.5	0.56		0.54					0.26								0.36		
[Dollinger et al., 2008]	338	m=21.9	0.43	0.44	0.32												0.21			-0.10
[Duff et al., 2004]	146	[17-52]	0.34		0.39	0.38		-0.21										0.31		
[Eppler and Harju, 1997]	216	m=19	0.22	0.30							0.34									-0.14
[Eppler and Harju, 1997]	243	m=21.2	0.12								0.32									-0.16
[Kaufman et al., 2008]	315	m=25.9	0.14		0.24	0.12					0.15	-0.16								
[Komarraju et al., 2011]	308	[18-24]	0.15			0.33	0.14	0.19	0.15											
[Swanberg and Martinsen, 2010]	687	m=24.5	0.21		0.30									0.15	-0.17				-0.14	

g=General cognitive intelligence; C=Conscientiousness; O=Openness; N=Neuroticism; A=Agreeableness; SE=Self-efficacy; IM=Intrinsic motivation; EM=Extrinsic Motivation; De=Deep; Sh=Shallow; St=Strategic.

+The paper cited $\bar{R}^2$ only, R^2 is reported here for consistency with other cited results.

while Bidjerano and Dai [2007] reported the same R^2 in a model including factors of personality and self-regulation. These results suggest that cognitive ability is an important determinant of academic performance, particularly in models of standards students. Authors also cited that non-cognitive variables accounted for additional variance beyond what was accounted for by prior academic performance [Cassidy, 2011; Chamorro-Premuzic and Furnham, 2008; Dollinger et al., 2008; Kaufman et al., 2008; Swanberg and Martinsen, 2010]. This evidenced supported the inclusion of both cognitive and non-cognitive factors of learning in models of students at risk of failing in year 1.

2.7 Data analysis techniques used on educational data

Statistical models have dominated data analysis in the social sciences, including educational psychology [Dekker et al., 2009; Freedman, 1987; Herzog, 2006]. For example, studies cited in previous sections primarily used correlations analysis (78% of the studies) and regression (54% of the studies), with some papers citing path analysis results (14%) and structural equation models (11%) as detailed in Table 2.12. Statistical modelling has a sound theoretical basis, allowing verifiable conclusions to be drawn from model coefficients, therefore statistical models have made, and will continue to make, a valuable contribution to the understanding of learners and the learning process. However, such models are based on assumptions, including assumptions of normality, independency, linear additivity and constant variance [Nisbet et al., 2009]. It is evident from current knowledge of the factors influencing academic performance, that such factors are interdependent [Prinsloo et al., 2012]. While each factor measures unique attributes, there are overlaps in the constructs being measured. In addition, there is evidence to suggest variance is not constant for all attributes. For example, Vancouver and Kendall [2006] found evidence that high levels of self-efficacy can lead to overconfidence regarding exam preparedness, which in turn can have a negative impact on academic performance. Similarly, Poropat [2009] cites evidence of non-linear relationships between factors of personality and academic performance, including conscientiousness and openness. Duff et al. [2004] observed that because academic performance is itself a complex measure, calculated as an aggregate of a variety of assessment types, this weakens the result of correlation analysis with other learning dimensions. While recognising the continuing importance of statistical models, Freedman [1987] and Breiman [2001] argued that alternative modeling approaches should be considered when dimensionality is high and relationships are complex such as in the social sciences. Cox, in a response to Breiman's paper, notes the importance of the probabilistic base of standard statistical modelling, but agrees with Breiman that in some circumstances, an empirical

Table 2.12: Summary of analysis techniques used in 37 educational psychology studies, published between 1998 and 2012

Statistical Technique	Frequency of Use
Correlation Analysis	78%
Regression	54%
Path Analysis	14%
Structural Equation Modelling	11%
Anova	5%

approach is better [Breiman, 2001, p. 18]. It is therefore pertinent to ask if data mining's empirical modelling approach can add value to educational data analysis, in particular their relevance to models of academic achievement.

Data mining is a relatively young field, that has evolved primarily to aid the extraction of information from the vast amounts of data accumulated in databases and data repositories in many domains [Larose, 2005]. The wide range of analytical techniques used in data mining emanate from a variety of disciplines including database systems, statistics, machine learning, visualisation, logic, spatial analysis, signal processing, image analysis, information retrieval and natural language processing, thereby making data mining itself a diverse, interdisciplinary field of study [Han and Kamber, 2006]. Data mining uses inductive reasoning to find strong evidence of a conclusion. While suited to big data analysis, it does not provide the statistical certainty offered by traditional statistical modelling [Nisbet et al., 2009].

Algorithms typically used on educational data include: clustering techniques to identify homogenous subgroups in a dataset; association analysis to identify values that frequently co-occur; classification techniques to build models that predict membership of predefined classes in a dataset; and visual analytics to facilitate human analysis via interactive visual representations of the data [Baepler and Murdoch, 2010; Romero and Ventura, 2007, 2010]. A review of mining approaches used in educational data mining by Baker and Yacef [2010] identified a predominance of classification techniques, which are reviewed next.

2.8 Classification algorithms used on educational data

2.8.1 An overview of classification models

A **Decision Tree** algorithm identifies patterns in a dataset as conditions, represented visually as a decision tree [Quinlan, 1986]. For example, the following two conditions depict a branch of depth two that capture characteristics of instances in a class $atRisk = false$: *(if Conscientiousness > 5.6 and Self-Efficacy > 6.3 then atRisk = false).* The size of the tree (rule depth) is configurable, influencing the specificity of the resulting model [Quinlan, 1986]. Simpler implementations (e.g. C5.0) limit each branch to value ranges from a single attribute, making this a linear classifier with a further restriction that each condition is an axis-parallel hyperplane [Tan et al., 2014]. Less restrictive implementations can incorporate a greater range of patterns (e.g. CART, Breiman et al. [1984]). Model interpretability makes Decision Trees a popular choice [Han and Kamber, 2006].

Rule based classifiers define class membership based on a set of *if...then...* rules. Basic implementations generate models that are similar to a Decision Tree model [Tan et al., 2014] despite the difference in search strategies used. Rule based classifiers implement a depth first search, Decision Trees implement a breath first search [Gupta and Toshniwal, 2011]. However, rule based classifiers can be extended to incorporate fuzzy rules with less precise conditions, allowing an instance to match more than one class. For example, the rule (*if Conscientiousness is 'very' good and Self-Efficacy is 'fairly' good then atRisk = False*) uses the fuzzy sets 'very' and 'fairly' instead of specific value ranges. This non-deterministic model of the data can represent more complex, non-linear class boundaries [Otero and Sánchez, 2005; Tang et al., 2012].

A **Back-propagation Neural Network** (BPNN) is an empirical classifier that can approximate any function mapping input values to an output value. Inspired by the biological neural system, a Neural Network is a network of nodes, connected by weights, which when multiplied by input values and summed, will approximate an output value [Han and Kamber, 2006]. Each node can optionally apply an activation function to its output, such as a logistic function, to model a non-linear mapping from inputs to output. Training a network involves adjusting weights to bring the calculated output closer to the actual output. The resulting model may not be optimal, particularly when the solution is non-linear [Tan et al., 2014]. Nonetheless, BPNNs performance has been found to be comparable with other approaches, particularly when approximating complex patterns based on numeric input values [Groth, 2000; Sargent, 2001].

Models based on Bayes Theory include **Naïve Bayes** and **Bayesian Networks**. Naïve Bayes builds a model of probabilities based on both the distribution of classes in a dataset, and the distribution of attribute values present in each class. It then applies Bayes theorem to estimate the probability of class membership for any given combination of attribute values [Ng and Jordon, 2001]. For example, a result could be *P(atRisk=false | gender=female and self-efficacy=0.7) = 0.063; P(atRisk=true | gender=female and self-efficacy=0.7)=0.0001.* Naïve Bayes works well with a variety of data types [Tan et al., 2014], and can converge to its optimal accuracy quickly, making it suitable for relatively small datasets [Ng and Jordon, 2001]. However Naïve Bayes simplifies the learning task by assuming all attributes are independent. If this assumption is invalid, conditional probabilities between attributes can be modelled as a Bayesian Network [Bekele and Menzel, 2005]. Bayesian Knowledge Tracing (BKT), based on a Bayesian Network, is a popular method for estimating student knowledge based on their behaviour on intelligent tutoring systems. BKT models the probability that a student has learned a skill based on the estimated likelihood that a correct answer is either a guess or knowledge learned, and an incorrect answer is either a slip or lack of knowledge [Baker et al., 2011].

A **Support Vector Machine** (SVM) models class membership by approximating a hyperplane that defines a linear boundary between two classes [Cortes and Vapnik, 1995]. In cases where the class boundary is non-linear, a kernel function can transpose the dataset to a higher number of dimensions, which may provide a linear class boundary [Nisbet et al., 2009, p. 13]. Training an SVM is a convex optimisation problem to which a globally optimal solution can be found [Tan et al., 2014]. While SVMs are limited to numeric attributes and binary classification tasks, Dixon and Brereton [2009] found SVMs outperformed other learners when modelling attributes that are not normally distributed.

Ensembles aggregate the predictions of a collection of classification models [Banfield et al., 2004; Breiman, 1996]. Individual models within an Ensemble can differ based on the subset of data used to train each model, and/or the algorithms used to build each model. There are also a variety of ways to aggregate predictions including averaging, using a voting strategy, or training a learner to identify which model to use for a given instance [Tan et al., 2014, p. 276]. While resource intensive in terms of training time, Ensembles tend to outperform individual classifiers, particularly when the accuracies of individual learners are relatively poor and their incorrect predictions are uncorrelated [Tan et al., 2014].

2.8.2 Classification model accuracies when modelling academic performance

Table 2.13 summarises a selection of educational data mining studies, classification algorithms used, and accuracies achieved. A distinction is made between models of log data capturing student actions over time, and models of static data such as prior academic performance, demographic data and non-cognitive factors of learning, measured at a point in time. Many publications on student modelling focus on log data gathered from Virtual Learning Environments (VLEs) hosting educational resources and student interactions, or Intelligent Tutoring Systems (ITS) that are aimed towards curriculum adaptation to each learner by monitoring progress and measuring skill levels [Baker and Yacef, 2010; Tempelaar et al., 2013]. Less focus has been given to modelling non-temporal data from outside virtual or online learning environments.

Both Pardos et al. [2011] and Minaei-Bidgoli et al. [2003] recommended an Ensemble to predict performance on an ITS, particularly for larger datasets. However in a comparison of Ensembles with individual classifiers to track student knowledge, Baker et al. [2011] concluded that an Ensemble was not statistically significantly better than the best individual classifier, a BKT model. Bekele and Menzel [2005], Conati et al. [2002], Jonsson et al. [2005] and Mayo and Mitrovic [2001] argue that Bayesian networks are particularly suited to student models because of the inherent uncertainty in interpreting student behaviour, and the incompleteness of any dataset attempting to capture all factors relevant to classifying students. However Yu et al. [2010] found that while Bayesian networks were suitable for modelling the temporal nature of data from an online learning tool, when data was converted into a single vector per student, more traditional classification approaches gave more accurate results, such as a Decision Tree Ensemble. Romero et al. [2008] achieved the best accuracy using fuzzy rule learning when modelling Moodle (VLE) usage data converted to a single vector per student. Similarly, Merceron and Yacef [2005] achieved high accuracy using a Decision Tree to predict exam performance based on a single student vector aggregated from their behaviour on an ITS.

In a comparison of models based on prior academic performance and demographic data, Herzog [2006] found Decision Trees and Neural Networks had similar performance to Logistic Regression when modelling datasets with little collinearity between variables, but outperformed Logistic Regression when modelling datasets with greater dependencies between variables. Additionally, both Decision Tree and Neural Network models identified significant predictor variables that had shown little statistical significance in a regression model. In a comparison of Decision Tree, Logistic Regression and Support Vector Machine, Lauria et al. [2013] reported comparable performance when modelling prior academic performance, demographic data, and ITS usage data (n=6,445). However, working with similar attributes, both Jayaprakash et al. [2014] and Lauria et al. [2012] reported Logistic Regression outperformed a Decision Tree when modelling a larger dataset (n=15,150 and n=18,968 respectively). Bergin [2006] achieved good accuracy with Naïve Bayes when modelling a dataset of prior academic performance and non-cognitive factors, and observed that while an Ensemble had marginally higher accuracy than Naïve Bayes, it did not justify the additional effort involved in compiling the Ensemble.

A limited number of educational data mining studies have investigated the role of

Table 2.13: Educational data mining models: factors used and accuracy achieved

Study	Algorithm	Accuracy	n	Class label	Demographic Data	Prior Education	Psychometric data	ITS
[Bergin, 2006]	Ensemble	82%	102	Weak/strong		x	x	
[Herzog, 2006]	Decision Tree	83%	4,564	Degree completion time	x	x		
[Jayaprakash et al., 2014]	Logistic Regression	87%	15,150	Weak/strong	x	x		x
[Dekker et al., 2009]	Decision Tree	79%	1,002	Drop out		x		
[Lauria et al., 2013]	Decision	87%	6,445	Weak/strong	x	x		x
Study	Algorithm	Accuracy	n	Class label	VLE	ITS		
[Baker et al., 2011]	Bayesian Network	AUC: 0.70	76	Next question correct		x		
[Merceron and Yacef, 2005]	Decision Tree	87%	224	Pass/fail		x		
[Minaei-Bidgoli et al., 2003]	Ensemble	94%	227	Pass/fail		x		
[Pardos et al., 2011]	Ensemble	AUC: 0.77	5,422	Performance on ITS		x		
[Romero et al., 2008]	Fuzzy Rule	62.11%	438	Module performance	x			

ITS: Intelligent Tutoring System; VLE: Virtual Learning Environment; AUC: Area under the Curve.

non-cognitive factors in models of learning [Buckingham Shum and Deakin Crick, 2012]. Bergin and Reilly [2006] found that including self-efficacy and study hours improved model accuracy, but due to a small sample size (n=82) could not draw reliable conclusions from the findings. Lauria et al. [2012] reports good model accuracy (88%) when modelling prior academic performance, demographic factors and student effort inferred from student activity on a VLE. Nelson et al. [2012] recommended including non-cognitive factors in models of learning to provide useful feedback on the learning dispositions that assessment design rewards. Buckingham Shum and Deakin Crick [2012] argue for greater recognition of learning dispositions (e.g. persistence, curiosity, awareness of learning) as important dimensions of learning that should be assessed in conjunction with discipline knowledge. Shute and Ventura [2013] concur, and observe that important competencies such as persistence, openness and self-efficacy are not currently taught or assessed, despite evidence of their importance. Furthermore, Knight et al. [2013] argued that learning analytics should be more than just generating models, it should become part of the learning process itself. For example, supporting learners in self-regulating their learning through feedback on actions taken. Such developments necessitate that analytics tools acquire data on non-cognitive factors of learning to capture learner disposition and approaches to learning task.

2.9 Conclusion

This review collated evidence on the importance of both cognitive and non-cognitive factors of learning in the modelling of academic achievement in tertiary level education. While not accounting for all of the variance in the noted academic performance, prior academic performance, personality, motivation and learning strategies have significant relationships with academic performance, and overlap with noteworthy learning dispositions. Therefore, factors measuring prior academic performance, personality, motivation and learning strategies merited consideration in this study.

To date, the complementary disciplines of learning analytics and educational data mining have focused predominantly on analysing data that has been systematically gathered in educational settings, which at tertiary level includes factors of prior academic performance, demographic data such as age and gender, and data gathered by logs recording student behaviour on online learning environments. Though both are relatively new disciplines, initial results report good model accuracy across a variety of analysis techniques. Cited studies reported comparable accuracies across a range of classification algorithms used. Consequently, this study evaluated eight classification algorithms for models predictive of at-risk students. Chapter 3 will give details on the data collected for the study, and the analysis techniques used to model the data.

Chapter 3

Methods

3.1 Introduction

The primary research question (*Q1*) was: *Can algorithmic student modelling accurately predict Irish IoT students at risk of failing in first year of study based on factors that can be measured prior to commencement of tertiary education?* To answer this question, data from a diverse student population was gathered and analysed. This chapter describes the study participants and study factors used. Some data were gathered from on online learner profiling tool developed specifically for this study (Objective *Ob3*). The tool itself and questionnaire reliability are discussed. Finally, the statistical techniques applied and classification models used are explained.

3.2 Description of study participants

The participants were first year students at Institute of Technology Blanchardstown (ITB), Ireland. The admission policy at ITB supports the integration of a diverse student population in terms of age, disability and socio-economic background, as evidenced in ITB's mission statement (Appendix B.1). As was discussed in Section 1.2, course entry requirements are generally lower than corresponding university courses [Mooney et al., 2010].

Each September 2010 to 2012 all full-time, first year students at ITB were invited to participate in the study by completing an online, self-reporting, learner profiling tool administered during first year student induction. The tool included a request for permission to use student data in this study; students also signed a paper form consenting to their details being used in this study (see Appendix B.2.1 and B.2.2). A total of 1,376 (53%) full-time, first year students completed the online questionnaire. Eliminating invalid student IDs (n=100), those who did not give permission to be included in the study (n=35), students under the age of 18 at the time of profiling (n=3) and other errors in the dataset detailed in Section 4.2.1 (n=31), resulted in 46% of first year, full time students participating in the study (n=1,207).

Participants ranged in age from 18 to 60, with a mean (m) age of 23.27 (standard deviation, s=7.3); of which 355 (29%) students were mature (23 and over[1]), 713 (59%) were male and 494 (41%) were female. Students were enrolled on a range of courses in the disciplines of Business (n=402, 33%), Humanities (n=353, 29%), Computing (n=239, 20%), Engineering (n=172, 14%) and Horticulture (n=41, 3%).

Participant age and gender by year is detailed in Table 3.1. There were variations in the relative percentages profiled from each academic discipline by year even though first year enrolment numbers in each discipline were consistent over the duration of the study, as illustrated in Table 3.2. Most notable was a drop in Engineering students profiled in 2011. This may be explained by the scheduling of engineering induction that year which took place in the late afternoon. Profiling was the final session in their induction schedule and numbers declined as the afternoon progressed. In 2010, computing and engineering students were profiled during scheduled lab sessions in the first three weeks of term which may account for increased numbers that year. Participant numbers from humanities increased each year.

Table 3.1: Age and gender of participants by year

		Age			Gender	
Year	n	Range	$m \pm s$	Over 23	Male	Female
2010	418	[18,60]	24.0 ± 7.9	133 (32%)	261 (62%)	157 (38%)
2011	353	[18,59]	23.3 ± 7.3	106 (30%)	209 (59%)	144 (41%)
2012	436	[18,53]	22.6 ± 6.6	116 (27%)	243 (56%)	193 (44%)

Table 3.2: Participant numbers by discipline by year

Year	Business	Computing	Engineering	Horticulture	Humanities
2010	139 (33%)	102 (24%)	82 (20%)	13 (3%)	82 (20%)
2011	143 (40%)	63 (18%)	28 (8%)	18 (5%)	101 (29%)
2012	120 (28%)	74 (17%)	62 (14%)	10 (2%)	170 (39%)

[1]This is a state-wide definition of a mature student. Their entry requirements are less strict.

3.3 Study factors and instruments used

The study dataset included data from three sources: student registration; the study's online learner profiler; and exam results from first year of study at ITB, supplied by the college. The following sections cover each of these in further detail. Study factor names will be italicised when referenced in text.

3.3.1 Student registration data and measurement of prior academic performance

Registration data included age, gender and prior academic performance. Access to full time college courses in Ireland is based on academic performance in the leaving certificate (or equivalent), a set of state exams at the end of secondary school. The leaving certificate includes four mandatory subjects, namely mathematics, English,[1] Irish,[2] a foreign language, and typically an additional three elective subjects. Subjects can be studied at higher or ordinary level, mathematics and Irish are also offered at foundation level. College places are offered based on CAO[3] points, an aggregate score based on points achieved in a student's best six leaving certificate subjects, range [0,600]. Table 3.3 maps exam grades to points for higher level, ordinary level and foundation level examinations. The Java code used to calculate *CAO points* is included in Appendix D.1.

The study dataset included *CAO points*, points in *mathematics* and points in *English* for each student. Points achieved in additional subjects (53 in total) were included as average points achieved by subject category. The Department of Education in Ireland groups leaving certificate subjects into six categories based on subject content, namely humanities, social, artistic, practical, science and business[4]. However, average scores based on these six categories resulted in significant subsets of students without points for categories of non-mandatory subjects, namely social, business, artistic and practical subject categories. Therefore, categories were combined to create three categories as follows: *applied* (artistic and practical categories); *humanities* (humanities and social categories); and *methodical* (science and business categories). Table 3.4 lists subjects included in each category. Electives from the *humanities* category were the most popular, for example 56% studied

[1]The Leaving Certificate English syllabus aims to develop: a mature and critical literacy; a respect and appreciation for language; and an awareness of the value of literature (`http://www.education.ie`).

[2]The Leaving Certificate Irish syllabus aims to develop the students' language skills and nurture a respect and a positive attitude towards the Irish language.

[3]CAO refers to the Central Applications Office with responsibility for processing applications for undergraduate courses in Ireland.

[4]Details of subject groups can be found at the Department of Education's National Career Guidance website: `http://www.careersportal.ie`.

Table 3.3: Leaving certificate grades and their corresponding CAO points

Mark (%)	Grade	Higher level	Ordinary level	Foundation level
$[90, 100]$	A1	100	60	20
$[85, 90)$	A2	90	50	15
$[80, 85)$	B1	85	45	10
$[75, 80)$	B2	80	40	5
$[70, 75)$	B3	75	35	0
$[65, 70)$	C1	70	30	0
$[60, 65)$	C2	65	25	0
$[55, 60)$	C3	60	20	0
$[50, 55)$	D1	55	15	0
$[45, 50)$	D2	50	10	0
$[40, 45)$	D3	45	5	0
$[0, 40)$	E,F,NG	0	0	0

Bonus points for honours mathematics introduced by the Department of Education in 2012 were not incorporated into CAO calculations.
$[x, y)$ denotes a range inclusive of x but exclusive of y.

Table 3.4: Leaving certificate subject categories

Category	Description
Applied	Art, Building Construction, Craft Design & Technology, Engineering, Graphic & Tech Design, Music, Music & Musicianship, Drama & Theatre Studies, Technical Drawing, Technology, Leaving Certificate Link Modules
Humanities	Classical Studies, Economic & Social History, English, Geography, History, Home Economics, all Languages, Religious Studies
Methodical	Accounting, Agricultural Economics, Agricultural Science, Applied Mathematics, Biology, Business (Organisation/Studies), Chemistry, Computer (Science/Studies), Economics, Mathematics, Physics, Physics with Chemistry, Science/Environment Science

geography and 51% studied French. Many science and business subjects are numerate based, however, the two most popular electives in the *methodical* category did not have a significant mathematics component, namely biology (51% of students) and business (48% of students). The next most popular elective in this category, accountancy, was taken by 12% of students. All subjects in the *applied* category had a significant practical component, the most popular was building construction, taken by 17% of students. However, 43% of participants did not have a grade for this category, limiting its usefulness.

Descriptive statistics for study factors of prior academic performance in Table 3.5 confirmed a student sample with a weaker prior academic profile compared to university students as reported in [Mooney et al., 2010]. Of particular note was the low average points in *mathematics* (m=23.8, equivalent to 55%-65% in a pass level paper) which was significantly lower than all other subject areas.

Table 3.5: Descriptive statistics for study factors of prior academic performance

Subject	n	Average Points ($m \pm s$)	Range
CAO points	1,018	259.5 ± 78.1	[0,475]
Mathematics	1,008	23.8 ± 13.9	[0,90]
English	1,015	46.4 ± 18.5	[0,90]
Humanities Average	1,016	40.0 ± 14.0	[0,80]
Methodical Average	1,016	32.1 ± 15.5	[0,83]
Applied Average	647	48.5 ± 19.5	[0,87.5]

Valid range for *CAO points* is [0,600], valid range for subjects and subject categories is [0,100].

Means (m) and standard deviations (s) were calculated based on the number of participants who had results in each category as indicated by n.

3.3.2 Measurement of non-cognitive factors of learning

The following sections discuss the fifteen non-cognitive factors of learning included in the study. Questionnaire items to measure non-cognitive factors were primarily taken from validated instruments in the public domain and administered during first year student induction using an online learner profiler developed for this study (`http://www.howilearn.ie`). The wording of some questions was changed to suit the context as illustrated in Table 3.6. Unless otherwise stated, items used a five-level Likert scale. Both positive and negative questions were used. Questionnaire length can affect the quality of response [Burisch, 1997; Galesic and Bosnjak, 2009]. Consequently, the number of items was reduced for some scales by removing similar items despite the likely negative impact on internal reliability statistics, discussed in Section 3.3.2.2. The full questionnaire is included in Appendix B.2.3.

3.3.2.1 Learner profiler questionnaire design

Informed by the discussion in Section 2.3, two personality based factors were included in the questionnaire, *conscientiousness* and *openness*. Items for both scales were taken from the International Personality Item Pool (IPIP) scales for conscientiousness and openness, available in the public domain [Goldberg et al., 2006]. Six items were selected from the Conscientiousness Big Five Domain scale, and six items were selected from the Openness to Experience, NEO Domain scale.

Motivation was assessed based on *self-efficacy*, and two achievement motivation scales, *intrinsic goal orientation* (learning goal) and *extrinsic goal orientation* (performance goal), chosen because research suggests these factors of motivation are most predictive of aca-

Table 3.6: Changes made to published questionnaire items

Category	Original text	Revised Text	Reason for change
Conscientious	Follow a schedule.	I like to do things according to a plan or schedule.	Improve clarity.
Openness	I like art.	I like art and creativity.	Improve clarity.
Extrinsic goal orientation	I want to do well in class because it is important to show my ability to my family, friends, employers and others.	I get great satisfaction from doing well which drives me to work hard.	More general.
Intrinsic goal orientation	When I have the opportunity, I choose course assessment that I can learn from even if they don't guarantee good grades.	When choosing a topic for an essay I would pick a topic I can learn from, even if it means more work.	Language adjusted for context.
Self-efficacy	I am confident I can do an excellent job on assessments and tests for this course.	I think I'll be good at completing assessment work to a high standard.	Language adjusted for context.
Self-efficacy	I believe I will receive an excellent grade in this class.	I expect to do very well on this course.	Language adjusted for context.
Metacognitive self-regulation	During class time I often miss important points because I'm thinking of other things.	During class I often miss important points because I'm thinking of something else.	Language adjusted for context.
Metacognitive self-regulation	When I become confused about something I am studying for this class, I go back and try to figure it out.	When I'm confused by something I am studying, I try to go back and figure it out.	Language adjusted for context.
Metacognitive self-regulation	I often find that I have been reading for class but don't know what it was all about.	I often find that I have been studying for class but I don't know what it was all about.	Language adjusted for context.
Metacognitive self-regulation	When I study for this class, I set goals for myself in order to direct my studies for each study period.	I set goals for each study period in order to direct my activities.	Language adjusted for context.
Metacognitive self-regulation	Before I study new course material thoroughly, I often skim it to see how its organized.	Before studying a new topic, I often skim through it first to see how its organised.	Language adjusted for context.
Study time	I attend class regularly.	I plan to attend class regularly.	Language adjusted for context.
Study Effort	I work hard to do well in this class even if I don't like what we are doing.	I would work hard to do well even if I don't like what I am doing.	Language adjusted for context.

demic performance, as discussed in Section 2.4. Scales were from the Motivation Strategies for Learning Questionnaire (MSLQ) [Pintrich et al., 1991], available in the public domain. All four items from each of the intrinsic goal orientation and extrinsic goal orientation scales were included in the questionnaire. Three of the eight items from the self-efficacy scale were included.

Three factors of self regulation were included in this study: *metacognitive self-regulation*; *study time*; and *study effort*. The scales used were from MSLQ. To facilitate administration during student enrolment, items were selected based on their relevance to prior academic experiences. Five items were included from the twelve-item metacognitive self-regulation scale, four items were included from the eight-item time & study environment scale (*study time*), and three items were included from the four-item effort regulation scale (*study effort*). Items from the *study time* scale were changed to a three-item scale of: *yes; no; I'm not sure yet*. This was done to adapt items to the context as students had not yet started their course of study. For example, *'I have a regular place set aside for study'* was asked before students were familiar with the college library facilities.

Learning approach was assessed based on the Revised two-factor Study Process Questionnaire (R-SPQ-2F) published by Biggs et al. [2001] and available in the public domain. The published questionnaire provided separate scales for shallow and deep learning approaches. Items used a five-level Likert scale. The question style was changed for this study, forcing participants to choose between a *deep*, *strategic* or *shallow learning* approach. Each item on a four-item scale asked participants to pick one of three statements: two statements, relating to deep and shallowing learning approach, were taken from R-SPQ-2F; the third statement, relating to a strategic learning approach, was compiled in collaboration with the National Learning Network Assessment Service[1] (NLN). The style of question matched the style of items on a learning styles profiler designed by NLN and used by ITB in previous years.

In agreement with NLN, scales from their learning styles questionnaire were also included. This covered learner modality (*visual, auditory* and/or *kinaesthetic* (VAK) [Fleming, 1995]) which was scored from six questions, each offering two choices of modality, resulting in four items per modality across the six questions. An additional three questions focused on preference for solo or *group work*, each offering two choices.

The choice of how many items to include in the online questionnaire was influenced by a requirement to keep the questionnaire short, while gathering data on a diverse range

[1]The National Learning Network Assessment Service provides functional strategies and support for children, adolescents and adults with specific learning difficulties. They are located on campus at ITB (`http://www.nln.ie`).

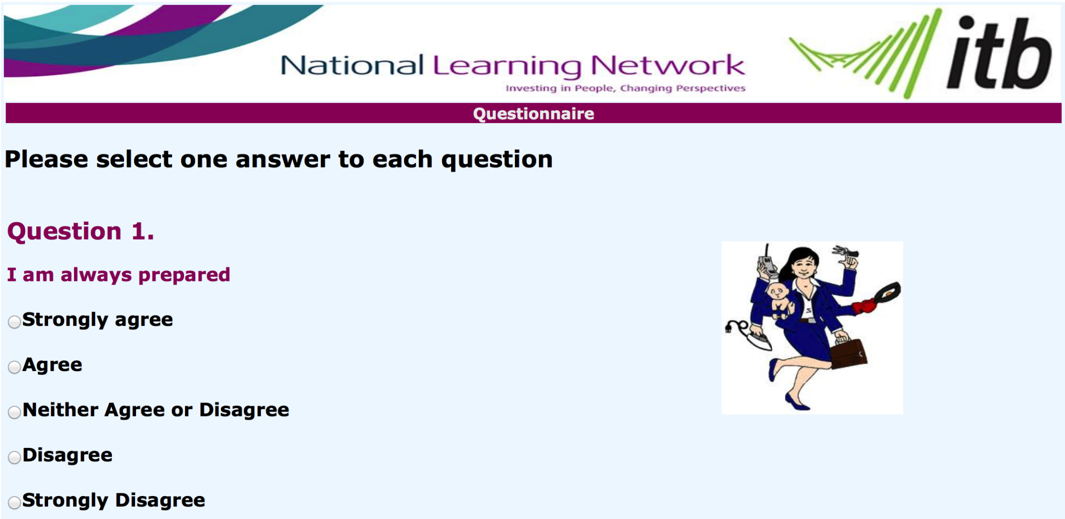

Figure 3.1: Illustration of the learner profiling tool interface used in this study

of relevant factors. Level of interest and perceived importance can reduce the negative impact on response quality when questionnaires are long [Herzog and Bachman, 1981]. Optimising interest was addressed in a number of ways: the user interface was designed to be colourful and visually appealing by the inclusion of graphical images relating to the questions being asked as illustrated in Figure 3.1; the tool gave immediate results with explanations to participants; and the questionnaire was administered as part of an information session on learning preferences and learning styles.

Table 3.7 details summary statistics for non-cognitive study factors. Attribute histograms (see Appendix C, Figure C.1) illustrated varied distributions across attributes, which is common in data relating to education and psychology [Kang and Harring, 2012; Micceri, 1989].

3.3.2.2 Learner profiler questionnaire reliability

Questionnaire validity and internal reliability were assessed using a paper-based questionnaire that included both the revised wording of questions used on the profiling tool (reduced scale), and the original questions from the published instruments (original scale). The paper questionnaire was administered during scheduled first year lectures across all academic disciplines, participation was optional. Each student completed one of five subsections of the questionnaire, comprising of questions on either one or two non-cognitive factors. Results are detailed in Table 3.8.

Pearson correlations (r) between scores calculated from the reduced scale, and scores

Table 3.7: Descriptive statistics for non-cognitive study factors

Category & Instrument	Factor	$m \pm s$	95% CI
Personality, from IPIP (`ipip.ori.org`)	Conscientiousness	5.95 ± 1.53	[5.86, 6.03]
	Openness	6.07 ± 1.29	[5.99, 6.14]
Motivation, from MSLQ [Pintrich et al., 1991]	Self-efficacy	6.85 ± 1.42	[6.77, 6.93]
	Intrinsic goal orientation	7.09 ± 1.36	[7.03, 7.17]
	Extrinsic goal orientation	7.81 ± 1.38	[7.73, 7.89]
Self-regulated learning, from MSLQ [Pintrich et al., 1991]	Metacognitive self-regulation	5.88 ± 1.36	[5.80, 5.95]
	Study effort	5.93 ± 1.77	[5.83, 6.03]
	Study time	6.17 ± 2.32	[6.04, 6.30]
Learning style, based on R-SPQ-2F [Biggs et al., 2001]	Deep learner	5.36 ± 2.91	[5.20, 5.53]
	Shallow learner	1.33 ± 1.95	[1.22, 1.44]
	Strategic learner	3.41 ± 2.48	[3.27, 3.55]
Preferred learning channel, NLN Learning Styles Questionnaire (`www.nln.ie`)	Visual	7.17 ± 2.06	[7.05, 7.28]
	Auditory	3.13 ± 2.17	[3.04, 3.29]
	Kinaesthetic	4.67 ± 2.42	[4.53, 4.80]
	Group work	6.55 ± 3.36	[6.36, 6.74]

m:mean; s:standard deviation; Valid range for each factor is [0,10].

calculated from the original scale, were high for all factors ($\geq$0.9) except *intrinsic goal orientation* (r=0.81, 95% CI [0.68,0.89][1]) and *study time* (r=0.79, 95% CI [0.65,0.87]).

Internal reliability was assessed using Cronbach's Alpha Coefficient (α). α is a measure of the split half correlation coefficients between items on a scale [Cronbach, 1951; Yu, 2001], i.e. if the items on a scale were split into two halves, and the correlation calculated between the aggregate score in each half, Cronbach's Alpha estimates the mean of all such half split correlations. A high α (>0.8) indicates good internal consistency between items on a scale [Lance et al., 2006]. The coefficient is lower for scales with less items, even when the correlation between those items is high [Cortina, 1993; Spector, 1992]. So for scales with less items, α values closer to 0.7 are acceptable [Cooper et al., 2010; Tavakol and Dennick, 2011].

All factors had acceptable reliability (>0.7) given the small number of questions per scale (between 3 and 6), with the exception again of *intrinsic goal orientation* and *study time*. On the *intrinsic goal orientation* scale, three of the four items related to work that was challenging but interesting, all three had statistically significant correlations with each other, ranging from r=0.373 (95% CI [0.08, 0.0.61]) to r=0.406 (95% CI [0.12, 0.0.63]). The fourth item focused on understanding: *'I get the most satisfaction if I understand what I am studying as thoroughly as possible'*, and had a statistically significant correlated

[1]Correlation confidence intervals are based on a Fisher r-to-z transformation [Fisher, 1915].

Table 3.8: Questionnaire validation: correlations and Cronbach alpha (α)

Factor	n	Reduced scale α	Reduced scale Items	Original scale α	Original scale Items	Correlations between reduced and original scales	Published α
Conscientiousness	42	0.69	6	0.80	10	0.95 (95% CI [0.91, 0.97])	0.79
Openness	47	0.70	6	0.84	10	0.90 (95% CI [0.82, 0.94])	0.82
Self-efficacy	48	0.82	3	0.81	7	0.93 (95% CI [0.89, 0.97])	0.94
Intrinsic goal orientation	43	0.63	4	0.53	4	0.81 (95% CI [0.68, 0.89])	0.74
Extrinsic goal orientation	48	0.69	4	0.58	4	0.90 (95% CI [0.82, 0.94])	0.62
Metacognitive self-regulation	38	0.70	5	0.70	12	0.90 (95% CI [0.81, 0.95])	0.79
Study time	48	0.55	4	0.68	8	0.79 (95% CI [0.65, 0.87])	0.76
Study effort	41	0.69	3	0.74	4	0.98 (95% CI [0.96, 0.99])	0.69
Learning style	42	0.76	4				

with just one other item: *'If choosing a topic for an essay, I would pick a topic I can learn from, even if it means more work'*, (r=0.376, 95% CI [0.06, 0.61]). The *study time* scale included items relating to: study environment; time spent studying; how that time is used; and attendance. All respondents to the paper questionnaire planned to attend regularly, however respondents were those in attendance at a lecture introducing bias to the sample. There was no correlation between this item and others in the scale. Of the remaining items on the reduced scale, the item relating to how time was used: *'When I am studying I make good use of my time'* had reasonable correlations with both the time item: *'Its hard to find time to study because of other activities'* (r=0.44, 95% CI [0.168, 0.64]), and the study environment item: *'I have a regular place set aside for studying'*, (r=0.41, 95% CI [0.14, 0.62]]). Correlation was not significant between the time and study items (r=0.26, 95% CI [-0.03, 0.51]). Similarly in the full questionnaire, the highest correlations (r=0.51 95% CI [0.27, 0.69] and r=0.47 95% CI [0.22, 0.67]) were between items about time, but correlations were weaker between items relating to time and items relating to study environment. Therefore, the low α may be a reflection of the item mix on the reduced scale, and a bias in the sample used. *Intrinsic goal orientation* and *study time* were not removed from the dataset, however, it is acknowledged that inferences based on these factors may be unreliable. Interestingly, Komarraju and Nadler [2013] reported similar difficulties with the intrinsic goal orientation scale when administered in the first week of term.

3.3.3 Measurement of year 1 academic performance

First year academic performance was measured as Grade Point Average (GPA), an aggregate score of between 10 and 12 first year modules, range [0,4]. GPA is calculated as a weighted average of grades achieved, where the weights are the number of credits per module. For this study, GPA was calculated from the results of the first sitting of each module. Appendix B.3 details the calculations done. A GPA$<$2.0, or a result of *fail* in any individual module, results in an award of *Fail* overall. Otherwise a student is awarded a *pass* result and may progress to the next academic stage. Table 3.9 shows the academic profile of study participants across three GPA bands. Of the students with GPA$\geq$2.5 (n=558, 46%), 92% passed all modules indicating a low risk group that can progress to year two. Of the students with GPA$<$2 (n=432, 36%), 91% failed three or more modules, indicating a high risk group falling well short of progression requirements. For the remaining students in the GPA band [2.0, 2.5)[1] (n=217, 18%), 35% passed all modules, 36% failed one module, 20% failed two modules, and 8% failed more than two modules. This is a less homogenous group in terms of academic results, but could be regarded as struggling academically (medium risk), either passing all modules with low grades or required to repeat one or two modules to progress.

Histograms of first year GPA did not depict normal distribution as illustrated in Figure 3.2. Therefore, a two sample Kolmogorov-Smirnov non-parametric test[2] was used to compare GPA distribution (profiled sample) with the GPA distribution of the full cohort of first year students for that year (reference sample). The recorded differences in the distribution for 2010 (D=0.032, p=0.93), 2011 (D=0.036, p=0.90) and 2012 (D=0.042, p=0.69) were not statistically significant. Profiled samples were also comparable with each other, as were reference samples. The largest difference was between the 2010 and 2012 profiled samples (D=0.063, p=0.37) and was not significant. Therefore, it is reasonable to expect models trained on first year students from one academic year would be applicable to first year students in the other admission years. The following sections detail the analysis techniques used to model the dataset.

3.4 Statistical methods used

Both statistical analysis and data mining techniques were used in this study. Details of statistical methods are included in this section. Correlation and regression analysis facilitated comparison with other studies in educational psychology (*Ob4*). Analysis of subgroup dif-

[1] [x,y) denotes a range inclusive of x but exclusive of y.
[2] Kolmogorov-Smirnov does not make assumptions on the distribution of the data.

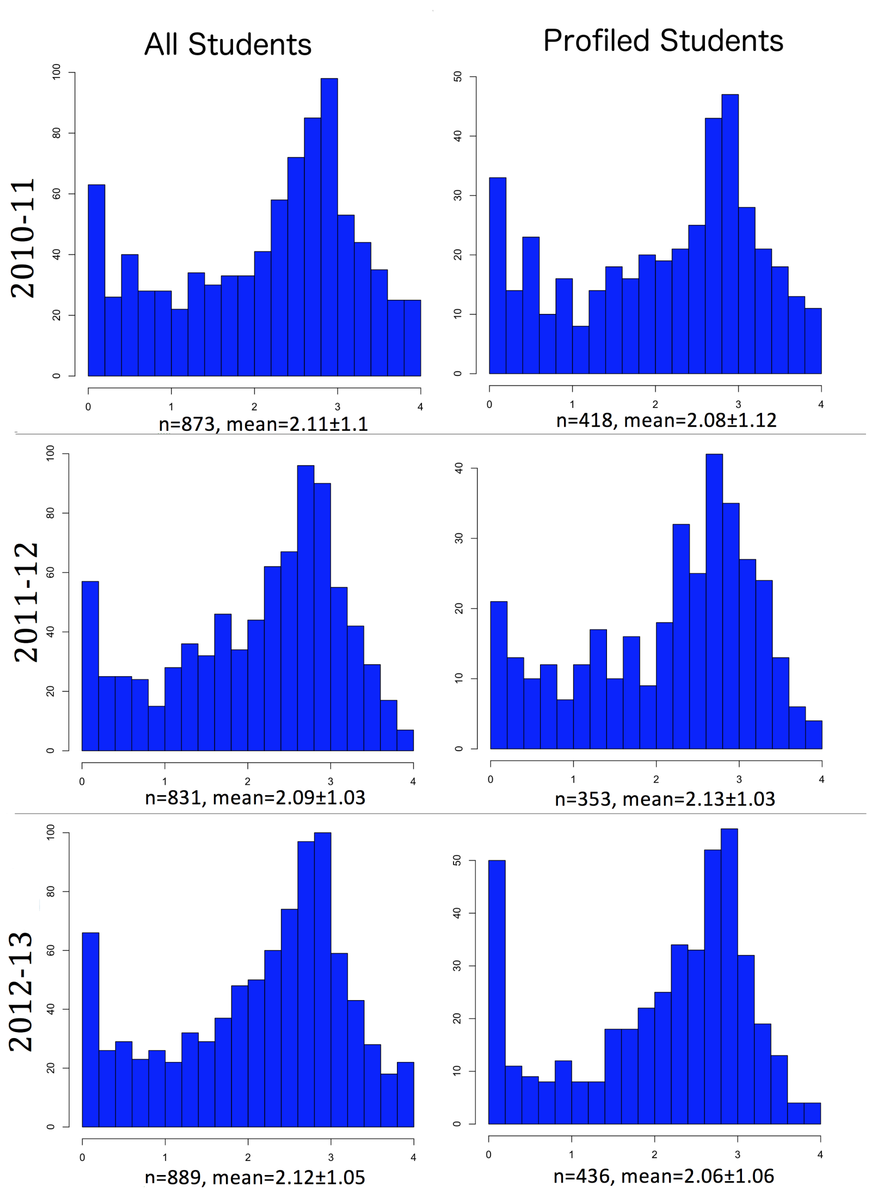

Figure 3.2: Histograms of GPA for both profiled students and all students in each academic year, including sample size (n) and GPA mean $\pm$ standard deviation.

ferences by GPA band, age group and gender aided interpretation of classification model results. Classification models will be detailed in Section 3.5.

3.4.1 Correlation analysis

Pearson product-moment correlation coefficients (r) were calculated for all study factors and GPA. The calculation is detailed in Equation 3.1 where x and y are the two variables, x_i and y_i are the values of x and y in row i respectively, and $\bar{x}$ and $\bar{y}$ are the means of x and y respectively [Chatfield, 1983, p. 187]. An assumption of calculating the statistical significance of a Pearson's correlation is that attributes are normally distributed. However, all study attributes failed a Shapiro-Wilk normality test ($p<0.05$, see Appendix C, Figure C.1 for histograms of each attribute). As was stated in Section 3.3.2.1, this is common in data relating to education and psychology [Kang and Harring, 2012; Micceri, 1989] which is likely to be skewed, have a heavy or light tail, and/or be multimodal [Smith and Wells, 2006]. Therefore, significance was verified using 1,999 Bootstrap Confidence Intervals using the Bias corrected and accelerated method (BCa) [Carpenter and Bithell, 2000] as implemented in R version 3.0.2. The code used is included in Appendix D.2.

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{3.1}$$

3.4.2 Analysis of group differences

Group differences were assessed for subgroups by GPA band (three groups), age group (three groups) and gender (two groups). For each of the eight groups, applying a Shapiro-Wilk test to attribute means of 50 bootstrap samples tested normality. A Brown-Forsythe test compared variance in attribute means across GPA bands, age categories and genders. Test results verified all attribute means were normally distributed but variances were unequal for most attributes. Therefore, results from Welch's t-test are reported for group differences by gender, although significances found concurred with the results from Student's t-test. The calculation is given in Equation 3.2 where x and y represent values in

Table 3.9: Count of participants by GPA band and number of failed modules

GPA band	n	Passed all modules	Failed 1 to 2 modules	Failed 3 to 6 modules	Failed > 6 modules
$[0.0, 2.0)$	432	1 (0.2%)	39 (9%)	146 (34%)	245 (57%)
$[2.0, 2.5)$	217	77 (35%)	122 (56%)	18 (8%)	0 (0%)
$[2.5, 4.0]$	558	515 (92%)	38 (7%)	5 (1%)	0 (0%)

two subgroups, $\bar{x}$ and $\bar{y}$ are the means of x and y respectively, s_x^2 and s_y^2 are the variances of x and y respectively, and n_x and n_y are the group size for x and y respectively.

$$\mathit{Welch's}\ t(x, y) = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} \tag{3.2}$$

For more than two groups, t-tests on all combinations of pairs increases the likelihood of a Type I error[1] (familywise error rate). A one way Analysis of Variance (ANOVA) compensates for this, and gives the same result as a t-test for two groups [Rice, 1995, p. 541]. In addition, ANOVA is robust with regard to assumptions of normality and equality of variance except in extreme cases [Hair et al., 2010, p. 458] and so was used for multiple group comparisons based on GPA bands and age groups. Post hoc comparison was done using Tukey's HSD test to compensate for familywise error. Tukey's HSD test identifies which subgroups differ significantly and is used after an ANOVA test identifies that there are significant differences between subgroups [Tukey, 1949]. Equation 3.3 gives the calculation for Tukey's HSD where MS_w is the mean square error within a group given by ANOVA, and n is group size. As ANOVA assumptions were violated, results were verified using Kruskal-Wallis, non-parametric test with post hoc Wilcoxon paired tests using Holm adjustment, a non-parametric test to compensate for familywise error [Holm, 1979; Kruskal and Wallis, 1952; Wright, 1992]. Results from ANOVA and Kruskal-Wallis tests concurred. All tests were done in R, code is included in Appendix D.3.

$$HSD(x, y) = \frac{\bar{x} - \bar{y}}{\sqrt{MS_w(\frac{1}{n})}} \tag{3.3}$$

3.4.3 Linear regression

Linear Regression models predict a continuous attribute, the dependent variable, by estimating the linear relationship mapping dataset attributes to the dependent variable. Equation 3.4 depicts this linear relationship where y is the dependent variable, a is the intercept and represents the mean value of y when all attributes equal 0, z is the number of attributes in the model, x_z represents the value of attribute z, b_z represents the coefficient for attribute z, and ϵ is the error term calculated as the difference between the estimated value of y ($\hat{y}$) and the actual value of y. Training a model involves solving for

[1]A Type I error rejects the null hypothesis when it is true [Rice, 1995, p. 300].

a and $b_1 \dots b_z$ to minimise ϵ.

$$y = a + b_1 x_1 + b_2 x_2 + \ldots + b_z x_z + \epsilon \tag{3.4}$$

Regression models predicting GPA were based on optimal attribute subsets identified using an exhaustive search as implemented in the *regsubsets* function in R V3.0.2 (leaps package V2.9, see code in Appendix D.4). All attributes were scaled to a mean of 0 and variance of 1. Two model fits are reported, adjusted coefficient of determination ($\bar{R}^2$) and mean squared error (MSE). $\bar{R}^2$ is reported to facilitate comparison with other studies. $\bar{R}^2$ compensates for the automatic increase in the coefficient of determination R^2 when additional dependent variables are added to the model. However, R^2 is influenced by variability in underlying independent variables. Consequently Achen [Achen, 1982, p. 58-61] argued prediction error is a more appropriate fitness measure for psychometric data. Therefore, MSE mean (m) and standard deviation (s) is also reported, as recommended by Pelánek [2015] for student models. Equations 3.5 to 3.7 give the calculations for R^2, $\bar{R}^2$ and MSE respectively where y_i is the actual value of y in row i, $\hat{y}_i$ is the predicted value of y in row i, $\bar{y}$ is the average value of y, n is the sample size, and m is the number of attributes in the model.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{3.5}$$

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-m-1} \tag{3.6}$$

$$MSE = \frac{1}{n}\sum_i (y_i - \hat{y}_i)^2 \tag{3.7}$$

Study results from regression models were compared using MSE mean and standard deviation. Variances in MSE were unequal, so a Welch's T-test was used (Equation 3.2). Degrees of Freedom (*df*) was estimated using Welch-Satterthwaite *df* estimate as detailed in Equation 3.8 where s is standard deviation in MSE and n is sample size [Ruxton, 2006].

$$df = \frac{\left(\frac{s_x^2}{n_x^2} + \frac{s_y^2}{n_y^2}\right)^2}{\frac{\left(\frac{s_x^4}{n_x^2}\right)}{n_x - 1} + \frac{\left(\frac{s_y^4}{n_y^2}\right)}{n_y - 1}} \tag{3.8}$$

3.5 Classification algorithms used and classification model evaluation

Classification models were used to predict students at risk of failing (Objective *Ob5*); attribute subset selection techniques were used to identify factors predictive of at-risk students (Objective *Ob6*). A binary class label of *fail* or *pass* was used, its definition will be discussed in Section 4.3.1. Classification models were evaluated for all participants and 17 subgroups by age (3 groups), gender (2 groups) and course of study (12 groups). Two training methods were compared for models of all participants: 10-fold cross validation using stratified sampling ($Model_{XVal}$); and model accuracy when trained on 2010 and 2011 participants and tested on 2012 participants ($Model_{2012}$). Subgroup models were trained on 2010 and 2011 participants and tested on 2012 participants.

Model accuracy for six classification algorithms and two ensembles were compared. Three classification algorithms were linear classifiers, namely: Decision Tree (DT), Naïve Bayes (NB) and Logistic Regression (LR). Two were non-linear: Back-propagation Neural Network (BPNN) and k-Nearest Neighbour (k-NN). A Support Vector Machine (SVM) was trained both without (linear) and with (nonlinear) a kernel function. Two ensembles were used, a Voting Ensemble and a Bagging Ensemble. RapidMiner version 5.3 (`https://rapidminer.com`) was used for classification modelling.

The following sections detail each algorithm used, parameter tuning options considered, and optimal parameter settings when modelling all participants. Approaches to attribute subset selection and model evaluation are also discussed.

3.5.1 Decision tree model

A Decision Tree (DT) represents patterns in a dataset as a simple tree structure. Each non-leaf node represents an attribute to be tested, branches represent attribute values or value ranges, and leaf nodes represent class allocation as illustrated in Figure 3.3. Dataset instances are allocated to a leaf node based on matching the corresponding branch conditions. DTs are easy to interpret, and patterns found tend to be robust provided the tree is kept small [Han and Kamber, 2006, p.304]. However, as was discussed in Section 2.8.1, DTs are limited in the type of patterns identified. Each branch represents a subgroup within the dataset, but subgroup boundaries are linear and parallel to the axis.

The DT algorithm grows the tree top down, starting with a single node that matches all instances in a dataset. If all instances matching a node are in the same class the node becomes a leaf node and is labelled with the class name. Otherwise the algorithm

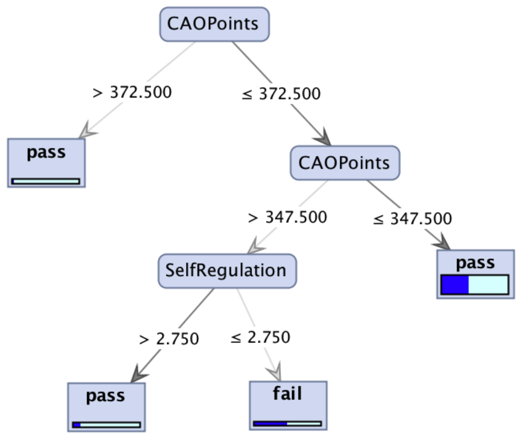

Figure 3.3: Decision Tree example

selects an attribute to split the instances matching that node and so grow the tree. For nominal attributes, a branch is created for each known value of the attribute. For numeric attributes, two disjoint ranges are selected based on a bin boundary that minimises entropy in each subgroup. This process is repeated recursively until a stopping criteria is met, determined by configurable parameters. Leaf nodes are labelled with the majority class amongst matching instances.

There are a range of heuristic measures for selecting the best attribute at each node of the tree. Preference is given to attributes that generate pure or almost pure branches, i.e. matching instance belong to the same class. Three were considered in this study: information gain (based on entropy); Gini index; and gain ratio. Information gain and Gini index are popular measures but have been criticised for bias in favour of attributes with multiple values generating many branches [Han and Kamber, 2006; Raileanu and Stoffel, 2004]. Gain ratio reduces this bias by dividing information gain by a weighting that gives preference to attributes that generate less branches [Han and Kamber, 2006]. The study dataset had numeric attributes only, limiting the tree to binary splits. However, initial tests using cross validation on the full dataset found that gain ratio achieved higher accuracy (66.1%) than information gain (61.6%), and marginally higher accuracy than Gini index (65.3%) but Gini index generated a larger tree. Therefore, gain ratio was used in this study. Equations 3.9 - 3.12 give the calculations for gain ratio where a is an attribute, k is the number of branches generated, d is the parent node before splitting, n is the sample size, t is a node on the tree, j is the number of classes in the dataset, and

$p(j|t)$ is the proportion of rows in class j at node t [Han and Kamber, 2006].

$$gainRatio(a) = \frac{InfoGain(a)}{SplitInfo(a)} \tag{3.9}$$

$$SplitInfo(a) = -\sum_{k} \frac{n_k}{n_d} log_2 \frac{n_k}{n_d} \tag{3.10}$$

$$InfoGain(a) = Entropy(d) - \sum_{k} \frac{n_k}{n_d} Entropy(k) \tag{3.11}$$

$$Entropy(t) = -\sum_{j} p(j|t)\ log_2\ p(j|t) \tag{3.12}$$

Algorithm stopping criteria influence tree size. Two stopping criteria were tuned for this study: gain ratio threshold limiting attributes considered at each branch to those that improve gain ratio by a minimum threshold value; and minimum leaf size limiting tree growth to only include leaf nodes that match a minimum number of instances. The Decision Tree required a gain ratio threshold < 0.05 to train. Therefore, models were trained for gain ratio thresholds in the range [0.0, 0.05] and minimum leaf size in the range [2, 20]. Optimal $Model_{XVal}$ accuracy used a gain ratio threshold of 0.009 and minimum leaf size of 10; optimal $Model_{2012}$ accuracy used a gain ratio threshold of 0.016 and minimum leaf size of 5.

3.5.2 Naïve Bayes model

Bayesian classifiers are based on probabilities as defined by Bayes Theorem. The probability that an instance (X) is in a particular class (C_j) is based on the distribution of attribute values in C_j as calculated from the training dataset. X is allocated to the class with the highest probability for $P(C_j|X)$. Equation 3.13 gives the formula to calculate $P(C_j|X)$ where z is the number of attributes, x_z is the value of attribute z in instance X, $P(C_j)$ represents the proportion of rows in the training dataset that are in class C_j, $P(X)$ is the probability of X which is the same for each C_j so it can be ignored, and $P(X|C_j)$ represents the combined probability that each attribute value in X could occur in class C_j. Naïve Bayes assume attributes are independent. This simplifies the calculation of $P(X|C_j)$ as the probability of all attribute values occurring together is the product of their individual probabilities. For nominal attributes, the individual probability for an attribute value (x_z) occurring in class C_j is calculated as the proportion of rows in C_j that have the value x_z. This is estimated from the training dataset. Numeric attributes are assumed to have a Gaussian distribution. The probability of x_z occurring in C_j is based on the probability distribution characterised by the mean (m_{zj}) and standard deviation (s_{zj}) of variable z in class C_j. The calculation is given in Equation 3.14 where g is the class-conditional probability defined by Equation 3.15 [Tan et al., 2014, p. 233].

$$P(C_j|X) = \frac{\sum_z P(x_z|C_j)P(C_j)}{P(X)} \tag{3.13}$$

$$P(x_z|C_j) = g(x_z, m_{zj}, s_{zj}) \tag{3.14}$$

$$g(x_j, m_{zj}, s_{zj}) = \frac{1}{\sqrt{2\pi s_{zj}}} e^{-\frac{(x-m_{zj})^2}{2s_{zj}^2}} \tag{3.15}$$

If attributes are truly independent, Naïve Bayes accuracy is optimal [Domingos and Pazzani, 1997]. Attributes are rarely independent in practice, however, empirical evidence shows Naïve Bayes can also achieve good predictive accuracy when assumptions are violated [Domingos and Pazzani, 1997].

3.5.3 Logistic regression model

As was detailed in Section 3.4.3, Linear Regression predicts a continuous dependent attribute (Equation 3.4). Similarly, Logistic Regression models the probability of an event occurring as a linear function, where an event is class membership [Han and Kamber, 2006, p. 358]. Linear regression assumes the error ϵ is normally distributed. In Logistic Regression, the error typically follows a logistic distribution. The resulting regression formula is given in Equation 3.16 where ln is natural logarithm, $\hat{p}$ is the probability that membership of a class is true; 1-$\hat{p}$ is the probability that membership of a class is false.

$$ln(\frac{\hat{p}}{1-\hat{p}}) = a + b_1x_1 + b_2x_2 + \ldots + b_zx_z + \epsilon \tag{3.16}$$

The implementation in Rapidminer is based on Shevade and Keerthi [2003]. Like SVM, their algorithm considers a complexity constants (C) representing the cost of misclassification. Best accuracy was achieved at $C \geq 1$ for both $Model_{XVal}$ and $Model_{2012}$.

3.5.4 Back-propagation neural network model

A Back-propagation Neural Network (BPNN) classification model comprises of connections between input neurons and output neurons where inputs are attribute values and outputs are classes. Connections represent weights to be applied to inputs to calculate output. Optionally there can be any number of hidden neurons representing interim values in the calculation from inputs to outputs. Figure 3.4 represents a Neural Network with three input neurons, two output neurons and four hidden neurons. Weights associated with connections between each input and the top hidden neuron are also illustrated.

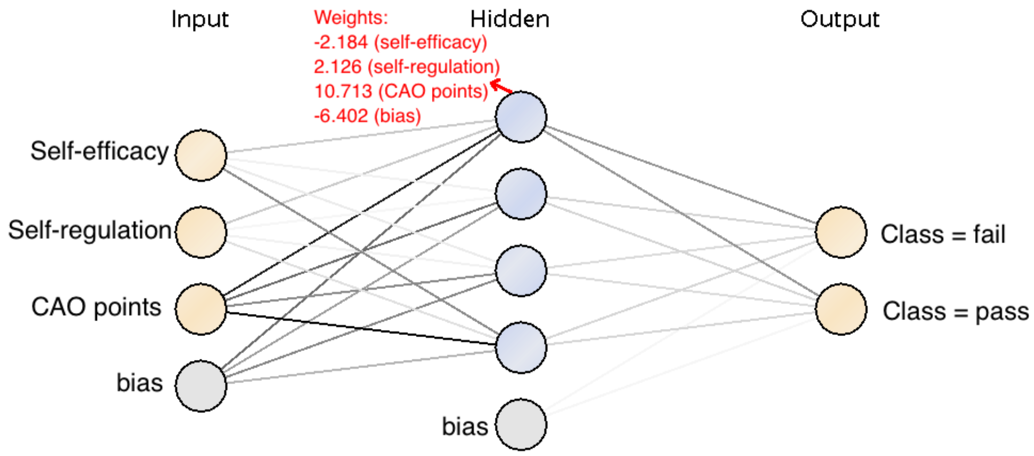

Figure 3.4: Neural Network example

The value for each input neuron is the attribute value from the row of data currently in the network. The value calculated at each hidden and output neuron is given in Equation 3.17 where y_j represents the output at neuron j, g is the activation function, i is a count of the inputs to y_j excluding the *bias*, x_i is the value of input i, and b_j is the *bias* weight for neuron j. Excluding the activation function gives the formula for a straight line or plane. An activation function, typically sigmoid, is applied to approximate a non-linear mapping from input to output. BPNNs are limited to numeric attributes. The value calculated at each output neuron, i.e. the predicted output, is also numeric. However, this can be mapped to a nominal value by comparison with a threshold value at each output neuron, for example '$y_{pass} > 0.7 \implies class = pass$'.

$$y_j = g(\sum_i w_{ij} * x_i + b_j) \tag{3.17}$$

Initially all model weights are set to random values in the range [-1,1]. Training a network involves adjusting each weight in the network so that predicted output approximates actual output. Weights are adjusted at each iteration. An iteration represents feeding one row of data into the network. Instances in a dataset are fed into the network consecutively; a single epoch is each instance fed into the network once. Training is done over a number of epochs which is a configurable parameter.

At each iteration, error is calculated as *(actual output - predicted output)*. This error term is used to adjust weights within the network, an activity termed back-propagation. Equation 3.18 illustrates weight adjustment for an output neuron (j) following iteration k where w_i^{k+1} is the weight for input i at iteration $k+1$, w_i^k is the weight for input i at iteration k, y_j is the actual output at node j, $\hat{y}_j$ is the predicted output at node j, x_i is the value of input i, and w_{adj}^k is the weight adjustment made in the previous iteration [Tan et al., 2014, p. 248]. λ is the learning weight which is a configurable parameter set in the range [0,1]. A high value for λ represents a larger weight adjustment at each iteration which may result in the network oscillating between a positive and negative error. A low value for λ represents smaller weight adjustments, and so a network that is slower to train [Larose, 2005, p. 139]. Rapidminer allows the value of λ to be decreased as predicted output converges towards actual output. α is the momentum, also a configurable parameter set in the range [0,1]. α controls the influence of previous weight adjustments, implementing exponential smoothing for weight adjustments. A higher value for α means previous weight adjustments have more influence, corresponding to a higher smoothing factor [Larose, 2005, p. 141].

$$w_i^{k+1} = w_i^k + \lambda(y_j - \hat{y}_j)x_i + \alpha w_{adj}^k \tag{3.18}$$

The weight adjustment formula is altered for hidden layer neurons as their actual output is unknown. Therefore, actual output is estimated as a proportion of the error in the subsequent layer, determined by connection weight. Thus Equation 3.18 is replaced with Equation 3.19 for weight adjustment at each hidden neuron (h) where o is the number of neurons h is connected to, err_o is the error calculated for neuron o, and w_o^k is the connection weight between j and o.

$$w_i^{k+1} = w_i^k + \lambda \sum_o (err_o * w_o^k)x_i + \alpha w_{adj}^k \tag{3.19}$$

Best accuracy for BPNN was achieved with learning rate (λ) of 0.25, momentum (α) of 0.3, 500 epochs and the default configuration of one hidden layer with number of hidden neurons set to: $1 + \frac{number\ of\ attributes\ +\ number\ of\ classes}{2}$. The activation function was sigmoid; attributes were scaled to the range [-1,1].

3.5.5 *k*-Nearest neighbour model

The k-NN model was the only lazy learner used in the study. Rather than compressing a dataset into a model representing its predictive pattern, k-NN classifies data directly from instances in the training dataset. A new instance is allocated to the majority class amongst its nearest neighbours in the training dataset, selected based on a distance measure. Neighbourhood size and distance measure are configurable.

Neighbourhood size (k) affects model performance. If k is too large, classification may be influenced by adjoining clusters in a different class; if k is too small, classification may be influenced by unusual cases not typical of the class allocation for the neighbourhood. k-NN models were trained on values of k in the range [2,50]. The most common distance measures for numeric attributes is Euclidean distance [Larose, 2005, p. 99] as given in Equation 3.20, where i and j are rows of data, x_{i1} represents attribute 1 in row i, and z is the number of attributes.

$$d_{Euclidean}(i,j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \ldots + (x_{iz} - x_{jz})^2} \tag{3.20}$$

The simpler Manhattan distance was also considered, where the difference between attribute values is not squared, reducing the influence of larger distances [Witten and Frank, 2005, p. 129]. Best $Model_{XVal}$ accuracy used Euclidean distance with k=18; best $Model_{2012}$ accuracy used Euclidean distance with k=15.

3.5.6 Support vector machine model

A Support Vector Machine (SVM) model is a linear hyperplane representing a decision boundary between two classes. An SVM algorithm solves for the optimal boundary where optimal means the decision boundary with the maximum margin between it and the classes at either side. As illustrated in Figure 3.5, a hard margin refers to a margin that does not permit misclassifications and can result in a decision boundary with a small margin that may overfit the data. A soft margin on the other hand permits some points to be misclassified resulting in a more general model with a wider decision boundary [Tan et al., 2014, p. 257-258]. A complexity constant (C) is a configurable parameter representing the cost of misclassifying an instance. Lower values for C result in softer margins.

Equation 3.21 represents the decision boundary where x is a vector on the boundary and b and w are parameters of the model. The decision boundary width can be expressed as Equation 3.22. Training an SVM involves estimating parameters b and w such that margin width is maximised and Equation 3.23 holds for each instance (x_i) in the dataset

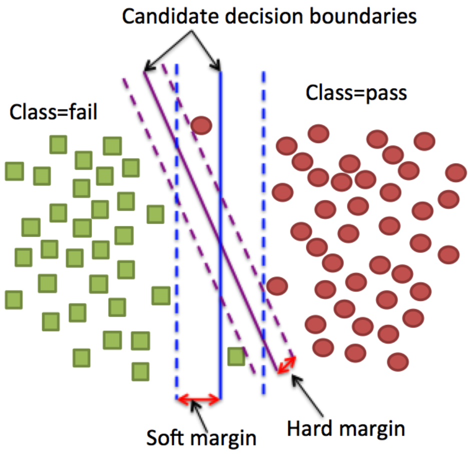

Figure 3.5: Illustration of an SVM decision boundary

where y_i is the class label for x_i. The classes of the binary class label are represented by -1 and 1 respectively.

$$w \cdot x - b = 0 \tag{3.21}$$

$$Margin\ width = \frac{2}{||w||} \tag{3.22}$$

$$y_i = \begin{cases} 1, & \text{if } \ w \cdot x_i + b \geq 1; \\ -1, & \text{if } \ w \cdot x_i + b \leq -1. \end{cases} \tag{3.23}$$

SVM is limited to numeric data, a binary class label and a linear decision boundary. If the decision boundary between classes is non-linear, a kernel functions (Φ) can be applied to the dataset. This is equivalent to mapping the dataset to a higher number of dimensions which may result in a linear class boundary. Rapidminer offers a range of kernel functions, each providing an alternative mapping to a higher dimension.

SVM models were trained on three complexity constants (C) in the range [0-5] and four kernel functions, namely dot (none), radial, polynomial and ANOVA. Best $Model_{XVal}$ accuracy used a radial kernel function and C=0; best $Model_{2012}$ accuracy used a dot kernel function and C=0.

3.5.7 Ensemble models

Ensembles combine the predictions of a number of classifiers, each referred to as a base classifier. Ensembles differ based on both the configuration of base classifiers and how their predictions are aggregated. Two types of Ensembles were used in this study, a Voting Ensemble and a Bagging Ensemble.

A Voting Ensemble combines the prediction of a number of base classifiers by a simple majority vote. Any number of base classifiers can be used. Experimenting with the six classifiers detailed in Sections 3.5.1 to 3.5.3, best $Model_{2012}$ accuracy was achieved using SVM, DT, BPNN, NB and two k-NN classifiers trained on different bootstrap samples of the training dataset, each with a sample ratio of 0.8. Including LR as a base classifier reduced overall model accuracy. Best $Model_{XVal}$ accuracy was achieved using seven learners, the same six learners used in $Model_{2012}$ and LR. In both cases, parameter settings for the constituent base algorithms were as detailed in Sections 3.5.1 to 3.5.3. A stacking ensemble was also evaluated. Rather than using a majority vote aggregate, a stacking ensemble trains a classification model to classify instances based on the predictions of base classifiers. However, a stacking ensemble failed to improve on the accuracy of a simple majority vote aggregate when tested on all participants ($Model_{2012}$).

A Bagging Ensemble uses a single base algorithm to create a number of base classifiers, each trained on a different bootstrap sample of the dataset. Results are aggregated using a simple majority vote. For each dataset modelled, the classification algorithm that achieved the highest accuracy for that dataset was selected as the base algorithm, typically k-NN. Parameter settings for the constituent base algorithm was as detailed in Sections 3.5.1 to 3.5.3. Configurable parameters include the size of each bootstrap sample (sample ratio) and the number of base classifiers used (iterations). Bagging ensembles were tested with iteration values in the range [2,15] and sample ratios in the range [0.5,1]. Optimal bagging accuracy used 8 iterations and a sample ratio of 0.9, although all iteration values gave similar results. Adaptive Boosting (AdaBoost) was also evaluated. It is an iterative form of bagging. After each iteration, instance weights are adjusted to increase the probability that misclassified instances are selected in the bootstrap sample of the next iteration. However, AdaBoost failed to improve on Bagging model accuracy when tested on all participants ($Model_{2012}$).

3.5.8 Attribute subset selection techniques used

Attribute subset selection techniques can improve model accuracy and identify relevant attributes [Hall and Homes, 2003]. Three wrapper methods were used when training each model, namely forward selection, backward selection and a genetic algorithm. The two most popular approaches, forward selection and backward selection, generally give good results [Hall and Homes, 2003] but can converge on a local optimum [Baumann, 2003]. Therefore, a genetic algorithm (GA) was also used as recommended by Yang and Honavar [1998]. In an analysis of 144 optimal $Model_{2012}$ models across algorithms and dataset subgroups, forward selection identified the best model most frequently (65% of models), followed by Genetic algorithm (33% of models). Forward selection identified 82% of the best k-NN models, where best model referred to the model with the highest overall accuracy.

3.5.9 Reporting of model accuracy

The class label was binomial (*pass* or *fail*), as will be discussed in Section 4.3.1. Table 3.10 illustrates the confusion matrix generated by a binary classification model where class recall is the percentage of instances in the class that were predicted correctly, and class precision is the number of predictions that were correct.

Table 3.10: Confusion matrix for a binary classifier

	Predicted Fail	Predicted Pass	*Recall*
Actual Fail	True Fail (TF)	False Pass (FP)	$\frac{TF}{(TF+FP)}$
Actual Pass	False Fail (FF)	True Pass (TP)	$\frac{TP}{(FF+TP)}$
Precision	$\frac{TF}{(TF+FF)}$	$\frac{TP}{(TP+FP)}$	

Two results are reported for each model, accuracy and geometric mean (GM), both calculated from the confusion matrix as illustrated by Equations 3.24 and 3.25 respectively. GM is more appropriate than accuracy for unbalanced datasets; it combines the precision and recall of each class and so compensates for the greater influence of the majority class in accuracy calculations [Akbani et al., 2004; Kubat and Matwin, 1997; Romero et al., 2008]. As will be discussed in Section 4.3.4, the degree of class imbalance varied across subgroups in the dataset, and was addressed by over-sampling the minority class. Accuracy was calculated from the confusion matrix of the balanced dataset. GM was calculated from the confusion matrix of the original, unbalanced, dataset, i.e. after removal of bootstrap replicas.

$$Accuracy = \frac{(TF+TP)}{(TF+TP+FF+FP)} \tag{3.24}$$

$$GM = \sqrt{\left(\frac{TF}{TF+FP}\right)\left(\frac{TP}{TP+FF}\right)} \tag{3.25}$$

3.5.10 Comparing classification model accuracies

Model accuracies were compared based on their confusion matrices. Two tests were used, McNemar's test and Fisher's exact test (FET). Tests were run in R, the code is included in Appendix D.5.

McNemar's test, based on chi squared (χ^2), can be used to compare the results of two classification models applied to the same dataset [Dietterich, 1998]. The contingency table to compare the two models (M_1 and M_2) is given in Table 3.11. The null hypothesis is that both models should have the same error variance, this means $n_{10} = n_{01}$ as defined in Table 3.11. McNemar's test compares the distribution of counts under the null hypothesis to the observed distribution of counts. The test statistic is given in Equation 3.26. When comparing several algorithms, p-values were adjusted using Holm correction to account for familywise error.

Table 3.11: Contingency table for McNemar's test

(n_{11}) Number of examples correctly classified by both M_1 and M_2.	(n_{10}) Number of examples correctly classified by M_1 but incorrectly classified by M_2.
(n_{01}) Number of examples incorrectly classified by M_1 but correctly classified by M_2.	(n_{00}) Number of examples incorrectly classified by both M_1 and M_2.

$$\chi^2 = \frac{(n_{01} - n_{10})^2}{n_{01} + n_{10}} \tag{3.26}$$

FET was used to compare model accuracies when applied to different datasets, for example, comparing an algorithm's performance for $Model_{XVal}$ and $Model_{2012}$. The corresponding contingency table is given in Table 3.12. The null hypothesis assumes that totals are fixed, therefore knowing any interior count allows the other numbers to be calculated. FET calculates the probability that one of the counts is equivalent to a random distribution as illustrated in Equation 3.27, calculated from the contingency table [Rice, 1995, p. 484].

Table 3.12: Contingency table for Fisher's exact test

	Correct	Incorrect	Total
XVal	n_{11}	n_{12}	$n_{1.}$
2012	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{..}$

$$p(n_{11}) = \frac{\binom{n_{1.}}{n_{11}}\binom{n_{2.}}{n_{21}}}{\binom{n_{..}}{n_{.1}}} \tag{3.27}$$

3.6 Summary

A total of 1,376 students were profiled during first year student induction over three years, 2010 through 2012, using an online learner profiling tool developed for this study. Details of the learner profiler tool and its validation were presented. Profiling data was combined with registration data from the college which included leaving certificate results, age and gender. The class label was based on an aggregate GPA calculated at the end of first year of study, which was supplied by the college. Descriptive statistics were included for all study attributes.

Three statistical approaches and eight classification algorithms were detailed. The analysis approach taken was explained, as well as details of each method used and parameter tuning done on classification algorithms. Two methods for comparing classification model accuracies were discussed, McNemar's test was used to compared models applied to the same datasets, Fisher's exact test was used to compare models applied to different datasets. Chapter 4 details the approaches used for data cleaning and preprocessing in preparation for data analysis, resulting in a study dataset of 1,207 participants.

Chapter 4

Data Cleaning and Preprocessing

4.1 Introduction

Data cleaning and preprocessing was required before completing statistical analysis and classification modelling. This chapter details data quality issues that arose and how they were handled. Justification for data preprocessing decisions is presented, including discretisation of GPA, attribute scaling, class imbalance and evaluation of sample size.

4.2 Data cleaning required

The following sections discuss assessment of data quality. Quality issues resulted in the removal of 131 instances from the dataset.

4.2.1 Errors in the dataset

The online questionnaire requested an ITB student ID, email address and course name to facilitate merging learner profiling data with college registration and results data. A total of 403 student IDs did not match with college registration data. ITB student IDs start with ‘B000’ followed by a five-digit unique code. Converting ‘b’ to ‘B’ and enforcing three preceding 0’s (‘000’) prior to the last five digits corrected 175 student IDs. For the remaining students, if a name was entered as the ID or incorporated in the email address, unique names were matched with registration data and verified by either a matching course name, or a matching student ID with two digits transposed (n=128). A remaining 113 student IDs could not be matched with registration data, their details were deleted from the dataset.

Nine participants appeared in the dataset twice. Of those, five enrolled on two different courses between 2010 and 2012. Details of their first course were deleted as learner profiling data related to their most recent year. The remaining four participants completed the online questionnaire twice using different email addresses. In two cases, answers were similar on both attempts so one row was deleted at random. In the other two cases, scores were not consistent across the two attempts so both were removed from the dataset.

Three participants had grades from two different courses in one academic year affecting GPA calculation and were deleted. In all three cases the student had changed course having failed modules on their first course, so their academic record included module grades from both courses.

Four participants had a score of 0 for all psychometric factors indicating non completion of the online questionnaire, and were removed from the dataset.

4.2.2 Missing data

Prior academic performance was unavailable for 189 (16%) of the remaining 1,207 study participants. The data was not available to the college as these students were either over 23 and so did not need a leaving certificate, or qualified under ITB's disability access scheme applied on a case by case basis. An additional 20 students (2%) had less than the required six leaving certificate subjects, so prior academic performance, measured as *CAO points*, was underestimated. Both subgroups represented non-standard students of interest to the study and remained in the dataset. However, an additional factor, *leaving cert average* (m=38.4, s=11.9, n=1,018) was added to indicate average points achieved across all leaving certificate subjects attempted. Interestingly, a ten year study of university students in Ireland found an average grade calculated across all leaving certificate subjects was more predictive of first year academic performance than *CAO points* based on the best six grades [Kelly and Marshall, 2012].

4.2.3 Outlier identification

To assess if data quality was affected by outliers, both univariate and multivariate outlier detection was used. Univariate outliers are attribute values that are unlikely to occur as part of the attribute's distribution model [Han and Kamber, 2006, p. 452]. However, study attributes had unknown distribution models (discussed in Section 3.4.1) making univariate outlier detection unreliable. A simple discordancy test of a value more than three times the standard deviation from the mean [Tan et al., 2014, p. 659] identified 21 participants with unusually low or high scores in factors of *intrinsic goal orientation*, *extrinsic goal* *orientation*, *self-efficacy* or *metacognitive self-regulation*. These four attributes had relatively high kurtosis ($\geq$3) as illustrated in Figure C.1, so a less extreme value would fail a discordancy test. A flag was added to the dataset indicating a potential outlier value to inform analysis of model results.

Multivariate outliers are rows of data with value combinations that are inconsistent with the data's model [Han and Kamber, 2006, p. 451]. Multivariate outlier detection using Local Outlier Factor (LOF) as implemented in Rapidminer V5.3 [Breunig et al., 2000; Kriegel et al., 2009] suggested the dataset did not have multi-variate outliers. LOF compares the local density of an object with the local density of its nearest neighbours and has an advantage over other outlier detection techniques because it provides a score of the degree to which a point is an outlier [Breunig et al., 2000]. Additionally, it can detect outliers in a dataset with varying cluster densities [Kriegel et al., 2009].

The calculations for LOF are explained in Equations 4.1 - 4.3 where: *MinPts* is a user defined parameter for the number of nearest neighbours that define the local neighbourhood of each object; $LOF_{MinPts^{(p)}}$ defines LOF for point p based on its *MinPts* nearest neighbours; *k-distance(p)* is the distance between p and its k^{th} nearest neighbour where $k = MinPts$; $N_{MinPts^{(p)}}$ is the number of objects who's distance from p is not greater than *k-distance(p)*. $N_{MinPts^{(p)}}$ may be greater than *MinPts* if there are two or more points at *k-distance* from p; $lrd_{MinPts^{(p)}}$ is the local reachability density of a point p as defined by Equation 4.2; and $reach\text{-}dist_{MinPts^{(p,o)}}$ is the reachable distance of object p with respect to object o and is defined as the max of *k-distance(o)* and *distance(p,o)* (Equation 4.3). *k-distance(o)* acts as a smoothing factor for distances where p is close to o. An outlier is identified as an LOF $\gg$ 1 while LOF close to 1 indicates an object with a similar density to its neighbours [Breunig et al., 2000]. Testing all *MinPts* in the range [5,50] gave a maximum LOF of 1.68. Figure 4.1 illustrates a histogram of LOF scores using the default *MinPts* range of [10,20] where the maximum LOF is returned from all values of *MinPts* in this range.

$$LOF_{MinPts^{(p)}} = \frac{\sum\limits_{o \in N_{MinPts^{(p)}}} \frac{lrd_{MinPts^{o}}}{lrd_{MinPts^{p}}}}{|N_{MinPts^{(p)}}|} \tag{4.1}$$

$$lrd_{MinPts^{(p)}} = 1 / \left(\frac{\sum\limits_{o \in N_{MinPts^{(p)}}} reach\text{-}dist_{MinPts^{(p,o)}}}{|N_{MinPts^{(p)}}|} \right) \tag{4.2}$$

$$reach\text{-}dist_{k^{(p,o)}} = max\{k\text{-}distance(o), distance(p,o)\} \tag{4.3}$$

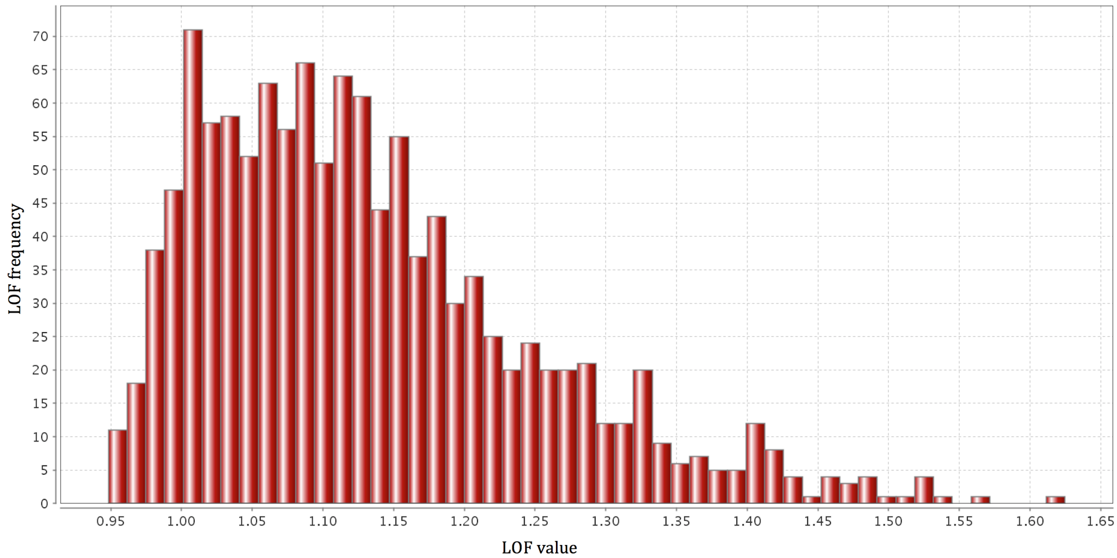

Figure 4.1: Histogram of Local Uutlier Factor (LOF) scores

4.3 Data preprocessing

4.3.1 Discretising academic performance

The class label needed to identify students at risk of failing. Experimenting with various ways to discretise GPA, Minaei-Bidgoli et al. [2003] found that higher numbers of bins (up to nine) resulted in higher errors because of low sample size in some bins (n=227). They recommended using either two (GPA$\leq$2.0 and GPA$>$2) or three (GPA$\leq$2; 2$<$GPA$<$3.5 and GPA$\geq$3.5) bins. A total of 30% of their participants achieved a GPA of 3.5 or higher. Romero et al. [2008] binned module grades into four categories based on final course mark, range [0,10]: fail [0,5), pass [5,7), good [7,9) and excellent [9,10]. However, they reported excellent and good students were frequently classified as pass, again citing lower sample sizes in those groups as a possible cause (n=438). Thai-Nghe et al. [2007] achieved best recall for failed students (64%) using two bins based on end of year GPA, range [2,4], namely fail [2,2.5) and pass [2.5,4] (n=20,492).

Both two and three GPA bins were considered for this study. To evaluate boundaries using two bins, seventeen GPA bin boundaries in the range [1.7, 2.5] were assessed using Naïve Bayes (NB)[1] with 10-fold cross validation. Optimal accuracy was achieved with a

[1]Early models of the data suggested Naïve Bayes gave comparable accuracies to other learners, con-

boundary of GPA=2.0 (accuracy: 68.5%, recall on fail: 71%) confirming a boundary between a passing and failing GPA. Models predicting three GPA bins were less successful. Sixteen models were tested using lower GPA boundary values in the range [1.0, 2.0] and upper GPA boundary values in the range [2.5, 3.25]. Models had difficulty distinguishing between medium and low risk students. The highest overall accuracy was achieved with GPA boundaries of 1.8 and 3.25 (accuracy: 53.5%, recall on fail: 64%), which was marginally better than a random guess (κ=0.3).[1] Superby et al. [2006] had similarly poor results predicting three classes. Therefore, two GPA bins were used for classification models in this study, GPA<2.0 (class=*fail*) and GPA$\geq$2 (class=*pass*). This distinguished high risk students from other students, as discussed in Section 3.3.3.

4.3.2 Evaluation of sample size

Progressive sampling indicates if combinations of attribute values likely to occur amongst study participants are sufficiently represented in a sample [Provost et al., 1999]. Each of the six classification algorithms was trained on fifty sample sizes between the sampling fractions of 0.3 (n=362) and 1 (n=1,207), using 10-fold cross validation with stratified sampling. Variance in NB model accuracy converged for sample factions > 0.75 (n=905). In addition, the slope of the LOESS regression line for NB model accuracy was approximately zero for sample fractions >0.8 (n=966) indicating convergence of model accuracy, as illustrated in Figure 4.2. NB can converge to optimal accuracy on a smaller sample size than other algorithms (Mitchell [2015]; Ng and Jordon [2001]). Variance in model accuracy for DT, BPNN and k-NN converged for sample fractions > 0.8 (n=966). Variance in SVM model accuracy converged for sample fractions > 0.85 (n=1,026) and LR model accuracy appeared to converge for sample fractions > 0.93 (n=1,123), however, a larger dataset would be needed to confirm accuracy convergence for LR.

4.3.3 Attribute scaling

Value ranges for study attributes varied from [0,10] to [0,600]. Attributes with larger ranges have a greater impact on models based on distance calculations such as k-NN [Larose, 2005, p. 100]. Two scaling options were considered, scaling to the range [0,1] and standard normal Z-transformation (m=0, s=1). Resulting improvements in model

curring with Bergin [2006] as discussed in Section 2.8.2.

[1] Cohens Kappa coefficient (κ) is a measure of the extent to which this result could have occurred by chance, range [-1,1]. κ in the range [-0.2,0.2] suggests a performance similar to a random guess [Kundel and Polansky, 2003].

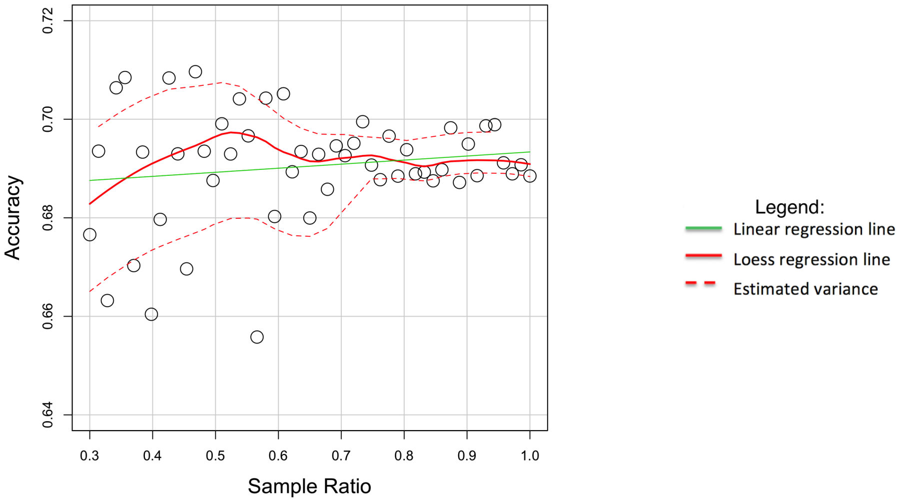

Figure 4.2: Model accuracy for progressive sampling using Naïve Bayes, generated using the scatterplot function in the R package *car*, version 2.0-21

accuracy[1] were statistically significant for Z-transformation only. For example, a k-NN model of all participants had an accuracy of 67.86% without attribute scaling when trained using 10-fold cross validation based on all study attributes. Accuracy increased to 69.27% when attributes were scaled to range [0,1] which was not statistically significant (McNemar's χ^2 (1, n=1207)=2.5128, p=0.113). Accuracy increased to (71.67%) using a Z-transformation which was statistically significantly higher than no scaling (McNemar's χ^2 (1, n=1,207)=7.42, p=0.006). Therefore, all attributes were scaled using a standard normal Z-transformation.

4.3.4 Class imbalance

A total of 38% of all participants were in class *fail*. The relative class size for *fail* varied across dataset subgroups by age, gender and course of study, ranging from 15% to 56% (see notched box plots of GPA by course in Appendix C, Figure C.3). Results from progressive sampling indicated the dataset was too small to under-sample the majority class, therefore, two class balance options were considered, basic over sampling of the minority class and synthetic minority over-sampling (SMOTE). Basic over-sampling of the minority class has been criticised for not addressing the issue of lack of data, and for over-fitting the data [Weiss, 2004]. Chawla et al. [2002] proposed SMOTE as an alternative over-sampling approach that generates synthetic instances along line segments joining nearest neighbours

[1]Accuracy was based on models of the original, unbalanced dataset. Class imbalance was considered after evaluation of attribute scaling.

in the minority class.

Both approaches were assessed by training models on the 2010 and 2011 student cohort and tested on the 2012 student cohort. This allowed training and test datasets to be balanced separately reducing the risk of overfitting, i.e. instances from the training dataset were not available when resampling test instances, and vice versa. To test SMOTE, additional instances of the minority class were generated using the SMOTE function implemented in R (package $DMwR$ version 0.4.1). The code is included in Appendix D.6.

SMOTE resulted in lower model accuracies than simple over-sampling for all learners, therefore, simple over-sampling of the minority class was used. However, it is worth noting that a comparison of model accuracies using SMOTE versus simple over-sampling showed the differences were not statistically significant. For example, Decision Tree (DT) models had the largest difference in accuracy; DT model accuracy was 63.12% using SMOTE, and 69.97% using basic over-sampling. A comparison of labelled test instances common to both models showed the difference in predictive accuracy was not statistically significant (McNemar's χ^2 (1, n=436)=3.03, p=0.082). This comparison excluded synthetic instances generated using SMOTE.[1]

4.4 Summary

Following data cleaning, 1,207 instances remained in the dataset including 209 (17%) participants for whom prior academic performance was unknown or incomplete. An additional attribute was added, *leaving cert average*, calculated as the average of all leaving certificate exams attempted. Results from both univariate and multivariate outlier detection suggested a small portion of the dataset (n=21, 2%) may contain outlier values.

Analysis of bin boundaries supported using two classes, *fail* (GPA$<$ 2) and *pass* (GPA$\geq$ 2). With the exception of Logistic Regression, n=1,026 was sufficient for classification model accuracy convergence. Class imbalance was resolved by over-sampling the minority class; in addition, attributes were scaled using a standard normal Z-transformation. Chapter 5 will detail data analysis results.

[1] A Fishers exact test (FET) comparing the same DT model accuracies indicated a statistically significant difference (p=0.02). While SMOTE's synthetic instances could be included in this comparison, FET assumes samples are independent. The number of test instances common to both datasets violated this assumption.

Chapter 5

Results

5.1 Introduction

Following data cleaning and preprocessing, the dataset was analysed using both statistical analysis and classification models as discussed in Sections 3.4 and 3.5. Results of statistical analysis exploring relationships between study factors and GPA are given to facilitate both comparison with other studies, and inform the discussion of classification model results in Chapter 6. This includes correlation analysis, analysis of group differences and linear regression models. Results from classification models predicting a binary class label of *fail* (GPA$<$ 2.0) and *pass* (GPA$\geq$ 2.0) are presented, including identification of key attributes used across classification models.

The results presented in this Chapter addressed the following three study objectives:

- Complete statistical analysis of cognitive and non-cognitive study factors for comparison with other published studies (*Ob4*).
- Train and evaluate a range of classification models predicting students at risk of failing (*Ob5*).
- Identify the key cognitive and non cognitive factors of learning that are predictive of first year students at risk of failing (*Ob6*).

5.2 Correlations between study factors, including year 1 academic performance

All measures of prior academic performance had significant correlations with each other and lower but significant correlations with GPA ($p<0.05$) as illustrated in the heat map in

Figure 5.1. *Methodical average* (r=0.302, 95% B-CI [0.24, 0.36]), *CAO points* (r=0.285, 95% B-CI [0.22, 0.34]) and *mathematics* (r=0.274, 95% B-CI [0.21, 0.33]) had highest correlations with GPA.[1] Results concurred with correlations between prior academic performance and GPA cited in other studies that included mature students [Conrad, 2006; Duff et al., 2004; Kaufman et al., 2008]. Dekker et al. [2009] also reported that a science aggregate, an overall aggregate and mathematics results were most predictive of degree completion.

	GPA	CAO Points	English	Maths	Applied Average	Humanities Average
CAO Points	0.285 [0.22, 0.34]					
English	0.169 [0.11, 0.23]	0.698 [0.67, 0.73]				
Maths	0.274 [0.21, 0.33]	0.477 [0.42, 0.53]	0.251 [0.19, 0.31]			
Applied Average	0.172 [0.10, 0.24]	0.560 [0.50, 0.61]	0.365 [0.29, 0.43]	0.173 [0.09, 0.25]		
Humanities Average	0.228 [0.17, 0.29]	0.820 [0.79, 0.84]	0.693 [0.66, 0.73]	0.263 [0.20, 0.32]	0.338 [0.26, 0.40]	
Methodical Average	0.302 [0.24, 0.36]	0.707 [0.68, 0.74]	0.432 [0.38, 0.48]	0.681 [0.65, 0.71]	0.194 [0.12, 0.26]	0.418 [0.36, 0.47]

Intervals are 95% Confidence Intervals based on 1,999 bootstrap samples. Only students with school leaving certificate results were included (n=1,018). *Applied average* results are based on the subset of students who did applied subjects (n=647, 64%).

Figure 5.1: Heat map of correlations between factors of prior academic performance, including GPA

With the exception of *visual* and *auditory modality*, all non-cognitive factors of learning were significantly correlated with GPA ($p<0.05$). Figure 5.2 is a heat map visualisation of correlations between non-cognitive study factors. Correlations with 95% confidence intervals are given in Appendix C (Figure C.2) indicating statistical significance. *Age* (r=0.25, 95% B-CI [0.2, 0.3]), a *deep learning approach* (r=0.234, 95% B-CI [0.18, 0.29]) and *study effort* (r=0.187, 95% B-CI [0.14, 0.24]) had highest correlations with GPA. *Openness* (r=0.084, 95% B-CI [0.03, 0.14]) and *group work* (r=-0.08, 95% B-CI [-0.13, -0.02]) had the weakest significant correlations with GPA. Correlations were comparable with other studies of diverse student populations with the exception of *self-efficacy* (r=0.12, 95% B-CI [0.06, 0.18])) which was lower than expected (for example Cassidy [2011]: r=0.397; Diseth [2011]: r=0.44; Komarraju and Nadler [2013]: r=0.30).

[1] B-CI: Bootstrap confidence interval as discussed in Section 3.4.1

	GPA	Personality		Motivation			Self-regulation			Learning Approach			Other			Modality	
	GPA	Con	Open	SE	EM	IM	SR	StE	StT	Deep	Stra	Shal	Group	Age	Gen	Vis	Aud
Con	0.150																
Open	0.084	0.032															
SE	0.120	0.313	0.178														
EM	0.124	0.280	0.049	0.308													
IM	0.149	0.334	0.316	0.421	0.381												
SR	0.130	0.515	0.101	0.409	0.298	0.429											
StE	0.187	0.450	0.064	0.334	0.232	0.330	0.594										
StT	0.101	0.396	0.009	0.259	0.175	0.227	0.452	0.378									
Deep	0.234	0.352	0.209	0.273	0.158	0.417	0.431	0.360	0.285								
Stra	-0.158	-0.167	-0.174	-0.158	-0.012	-0.274	-0.213	-0.133	-0.115	-0.791							
Shal	-0.146	-0.330	-0.096	-0.221	-0.234	-0.294	-0.398	-0.394	-0.290	-0.519	-0.103						
Group	-0.080	0.052	-0.042	0.056	0.059	0.027	0.113	0.094	0.084	0.020	0.037	-0.081					
Age	0.250	0.156	0.038	0.038	0.051	0.257	0.234	0.210	0.023	0.284	-0.200	-0.181	-0.022				
Gen	0.100	-0.005	0.022	-0.048	0.035	0.004	0.005	0.023	0.086	0.086	-0.001	-0.130	0.026	-0.038			
Vis	0.050	0.069	0.063	-0.024	0.041	0.054	0.024	-0.003	0.038	0.067	-0.020	-0.089	0.021	-0.038	-0.046		
Aud	0.020	0.073	0.023	-0.002	0.013	-0.016	0.065	0.039	0.081	0.077	-0.068	-0.026	-0.097	0.025	0.205	-0.347	
Kin	-0.059	-0.124	-0.074	0.022	-0.046	-0.032	-0.078	-0.033	-0.105	-0.126	0.078	0.099	0.069	-0.055	-0.144	-0.541	-0.601

Con:Conscientiousness; Open:Openness; SE:Self-efficacy; IM:Intrinsic goal orientation; EM:Extrinsic goal orientation; SR:Metacognitive self-regulation; StE:Study effort; StT:Study time; Deep:Deep learner; Shal:Shallow learner; Stra:Strategic learner; Group:Group work; Gen:Gender; Vis:Visual modality; Aud:Auditory modality; Kin:Kinaesthetic modality.

Figure 5.2: Heat map of correlations between non-cognitive factors of learning, including GPA

The relatively low internal reliability of *intrinsic goal orientation* was not reflected in correlations with other attributes. All factors of motivation were significantly correlated with each other. The highest correlation was between *intrinsic goal orientation* and *self-efficacy* (r=0.421, 95% B-CI [0.37, 0.47]), which concurred with Diseth [2011] (r=0.46) and was marginally lower than Komarraju and Nadler [2013] (r=0.53). Also of note was the significant correlation between *intrinsic* and *extrinsic goal orientation* (r=0.381, 95% B-CI [0.33, 0.43]) as correlations cited in other studies were inconsistent [Diseth, 2011; Eppler and Harju, 1997; Komarraju and Nadler, 2013]. Correlations between factors of motivation and factors relating to learning strategy were also significant, particularly *intrinsic goal orientation* and a *deep learning approach* (r=0.417, 95% B-CI [0.37, 0.47]).

Study time had relatively low internal reliability as discussed in Section 3.3.2.2. While correlations with *metacognitive self-regulation* (0.452, 95% B-CI [0.4, 0.49]) and *study effort* (r=0.378, 95% B-CI [0.33, 0.43]) were significant, they were lower than results cited in other studies, for example Bidjerano and Dai [2007] (r=0.55 and r=0.64 respectively) which was based on a similar participant profile.

The high negative correlation between a *deep* and *strategic learning approach* (r=-0.791, 95% B-CI [-0.81, -0.77]) reflected that most participants reported they were not shallow learners (m=1.3, s=1.9, range=[0,10]), selecting either deep or strategic statements. As expected, a *shallow learning approach* was negatively correlated with other non-cognitive factors. However, a *strategic learning approach* was also negatively correlated with other factors of learning and GPA (r=-0.158, 95% B-CI [-0.22, -0.10]), contradicting other studies, for example Duff et al. [2004] and Swanberg and Martinsen [2010]. The difference may be explained by their questionnaire design which facilitated selection of both strategic and deep learning approaches, resulting in a significant positive correlation between the two learning approaches.

As discussed in Section 3.2, computing and engineering students from the 2010 cohort were profiled during the first three weeks of term rather during student induction. To assess the impact of later administration of the learner profiler, correlations between study factors and GPA for the 2010 engineering students were compared with correlations for the 2010 general business students, selected because a t-test showed both cohorts had similar GPA distributions. Correlations between study factors and GPA were similar for both groups, indicating learner profiling administrated in the early weeks of semester 1 yielded similar results to learner profiling during student induction.

Key finding: With the exception of *study time*, correlations between study factors concurred with results cited in other studies. Also of note was the relatively low correlation between *self-efficacy* and GPA.

5.3 Analysis of group differences

Comparison of group means highlighted some statistically significant differences across subgroups. Results from three categories of subgroups are presented, namely GPA bands, age groups and gender.

5.3.1 Group differences by discretised year 1 academic performance

Group differences were assessed for the three GPA bands discussed in Section 3.3.3, namely high risk (GPA<2.0), medium risk ($2.0 \leq$GPA< 2.5) and low risk (GPA≥ 2.5) students.[1] A *deep learning approach* was the only attribute with statistically significant differences across the three groups (F(2, 1,204)=25.95, p<0.001). High risk students had significantly lower prior academic performance than either medium or low risk students, particularly in *methodical average* (F(2, 1,015)= 59.98, p<0.001), *CAO points* (F(2, 1,015)=50.33, p<0.001) and *mathematics* (F(2, 1,015)= 46.02, p<0.001). Low risk students were significantly different from the other two groups in some effective learning dispositions. They had higher scores in *intrinsic goal orientation* (F(2, 1,015)=50.22, p<0.001), *study effort* (F(2, 1,204)=17.76, p<0.001), *conscientiousness* (F(2, 1,204)=11.42, p<0.001) and *openness* (F(2, 1,204)=5.77, p=0.003). For the remaining non-cognitive factors, there were statistically significant differences between high and low risk students only, but medium risk students did not differ significantly from the other two groups. Groups did not differ significantly in *visual* or *auditory modality*. Group means are given in Table 5.1.

Key finding: Group differences by GPA band confirmed that both prior academic performance and non-cognitive factors associated with an effective learning disposition differentiated high risk from low risk students.

[1] Analysis of groups differences for two GPA bands corresponding to the class label (GPA<2.0; GPA$\geq$ 2.0) gave the same results in terms of statistically significant differences with two exceptions: differences in *group work* were not significant (p=0.06) but differences in *visual modality* were significant (p=0.03). Three bands are shown here to provide a more detailed picture of group differences.

Table 5.1: Group differences by GPA band, $m \pm s$

Study factor	Range	Low risk (n=558)	Medium risk (n=190)	High risk (n=459)	p
CAO points	[0,600]	223.5 ± 137.1	221.1 ± 117.6	212.2 ± 91.4	***
Leaving cert average	[0,100]	33.4 ± 20.6	33.1 ± 17.4	30.9 ± 13.5	***
Mathematics	[0,100]	21.6 ± 17.1	21.4 ± 15.4	17.1 ± 13.1	***
English	[0,100]	39.1 ± 26.3	40.6 ± 24.0	38.3 ± 21.0	***
Applied average	[0,100]	24.8 ± 29.4	26.4 ± 28.6	27.4 ± 26.3	
Methodical average	[0,100]	29.1 ± 20.8	28.8 ± 18.2	23.8 ± 14.7	***
Humanities average	[0,100]	34.2 ± 22.3	34.0 ± 19.0	33 ± 15.7	***
Conscientiousness	[0,10]	6.2 ± 1.5	5.8 ± 1.4	5.7 ± 1.6	***
Openness	[0,10]	6.2 ± 1.3	5.9 ± 1.2	6.0 ± 1.4	**
Self-efficacy	[0,10]	7.0 ± 1.4	6.9 ± 1.3	6.7 ± 1.5	***
Extrinsic goal orientation	[0,10]	8.0 ± 1.3	7.8 ± 1.3	7.6 ± 1.4	***
Intrinsic goal orientation	[0,10]	7.3 ± 1.3	6.9 ± 1.4	6.9 ± 1.4	***
Metacognitive self-regulation	[0,10]	6.0 ± 1.3	5.9 ± 1.3	5.7 ± 1.4	**
Study effort	[0,10]	6.2 ± 1.7	5.9 ± 1.7	5.6 ± 1.8	***
Study time	[0,10]	6.3 ± 2.2	6.4 ± 2.3	5.9 ± 2.4	*
Deep learner	[0,10]	6.0 ± 2.9	5.3 ± 2.8	4.7 ± 2.8	***
Shallow learner	[0,10]	1.0 ± 1.8	1.4 ± 1.9	1.7 ± 2.1	***
Strategic learner	[0,10]	3.1 ± 2.4	3.5 ± 2.5	3.8 ± 2.5	***
Gender	[0,10]	0.4 ± 0.5	0.5 ± 0.5	0.3 ± 0.5	***
Group work	[0,10]	6.2 ± 3.5	7.2 ± 2.9	6.8 ± 3.3	***
Age	[18,60]	25.1 ± 8.4	22.5 ± 6.7	21.4 ± 5.5	***
Visual	[0,10]	7.3 ± 2.0	7.3 ± 2.0	7.0 ± 2.1	
Auditory	[0,10]	3.2 ± 2.2	3.0 ± 2.0	3.1 ± 2.2	
Kinaesthetic	[0,10]	4.5 ± 2.4	4.7 ± 2.3	4.9 ± 2.4	*

*$p<0.05$,**$p<0.01$, ***$p<0.001$.

5.3.2 Group differences by age group

Study factor mean and standard deviations were initially compared for ten age categories: 18, 19, 20, 21, 22-23, 24-25, 26-28, 29-32, 33-39 and ≥40. Age groups were combined to ensure at least 60 students per group.[1] Analysis of differences in group means reduced the ten categories to three, namely: 18-23 (n=875); 24-28 (n=131); and 29-60 (n=201). This was based on a lack of statistically significant differences for study factors in age groups within these three age categories. Analysis of differences across the three age categories showed average GPA score increased significantly with *age* ($F(2, 1{,}204)=48.95$, $p<0.001$), as did many non-cognitive factors associated with an effective learning disposition, namely a *deep learning approach* ($F(2, 1{,}204)=68.54$, $p<0.001$), *intrinsic goal orientation* ($F(2, 1{,}204)=51.6$, $p<0.001$), *metacognitive self-regulation* ($F(2, 1{,}204)=39.19$, $p<0.001$), *study effort* ($F(2, 1{,}204)=32.57$, $p<0.001$), *conscientiousness* ($F(2, 1{,}204)=16.06$, $p<0.001$) and *extrinsic goal orientation* ($F(2, 1{,}204)=5.287$, $p<0.01$). As expected, *CAO points* decreased with *age* ($F(2, 1{,}015)=54.08$, $p<0.001$) as entry requirements are lower for students aged 23 and over. This difference in prior academic performance was reflected in all subject areas except *mathematics* ($F(2, 1{,}015)=0.271$, $p=0.763$). These results concur with a number of studies reporting a better learning disposition and academic performance amongst older students, for example Cassidy [2011]; Eppler and Harju [1997] and Hoskins et al. [1997]. The middle age category, age 23-28, was significantly higher than younger and older students in *openness* ($F(2, 1{,}204)=8.173$, $p<0.001$). Group means are given in Table 5.2.

Key finding: Group differences by age concurred with other studies that older students had higher GPA and a more effective learning disposition.

5.3.3 Group differences by gender

Engineering and computing courses were predominantly male and had low entry requirements. Humanities courses were predominantly female and had high entry requirements. Therefore, it was unsurprising that males had significantly lower *CAO points* than females ($t(918)= -4.077$, $p<0.001$). This difference was reflected in all subject areas, the least significant difference was in *mathematics* ($t(913)= -2.081$, $p=0.038$). GPA scores were also

[1] Applying both a Wilcoxon Rank Sum non-parametric test and a parametric t-test to a range of distributions common in psychometric data, Kang and Harring [2012] reported t-tests inflated Type I errors (incorrect finding of significance) for sample sizes less than 60 only, but performed well for larger samples.

Table 5.2: Group differences by age group, $m \pm s$

Study factor	Range	[18, 23] (n=875)	[24, 28] (n=131)	[29, 60] (n=201)	p
CAO points	[0,600]	266.0 ± 76.0	147.9 ± 127.6	59.9 ± 102.2	***
Leaving cert average	[0,100]	39.1 ± 11.8	22.3 ± 19.2	9.7 ± 16.1	***
Mathematics	[0,100]	23.4 ± 13.7	15.4 ± 17.3	7.4 ± 14.9	
English	[0,100]	47.6 ± 18.2	27.5 ± 25.4	9.3 ± 17.6	***
Applied average	[0,100]	31.5 ± 28.2	18.6 ± 26.5	7.0 ± 17.2	**
Methodical average	[0,100]	32.5 ± 15.5	19.0 ± 18.8	8.3 ± 15.5	***
Humanities average	[0,100]	41.1 ± 13.9	22.6 ± 20.1	8.8 ± 15.8	***
Conscientiousness	[0,10]	5.8 ± 1.5	6.0 ± 1.5	6.5 ± 1.4	***
Openness	[0,10]	6.0 ± 1.3	6.5 ± 1.3	6.1 ± 1.3	***
Self-efficacy	[0,10]	6.8 ± 1.4	7.0 ± 1.4	7.0 ± 1.5	
Extrinsic goal orientation	[0,10]	7.7 ± 1.4	7.9 ± 1.3	8.1 ± 1.4	**
Intrinsic goal orientation	[0,10]	6.9 ± 1.3	7.5 ± 1.2	7.8 ± 1.3	***
Metacognitive self-regulation	[0,10]	5.7 ± 1.3	6.1 ± 1.3	6.6 ± 1.2	***
Study effort	[0,10]	5.7 ± 1.8	6.2 ± 1.7	6.8 ± 1.5	***
Study time	[0,10]	6.2 ± 2.3	5.8 ± 2.4	6.5 ± 2.2	*
Deep learner	[0,10]	4.8 ± 2.8	6.5 ± 2.8	7.1 ± 2.5	***
Shallow learner	[0,10]	1.6 ± 2.1	0.8 ± 1.5	0.6 ± 1.3	***
Strategic learner	[0,10]	3.7 ± 2.5	2.8 ± 2.3	2.4 ± 2.1	***
Gender	[0,10]	0.4 ± 0.5	0.4 ± 0.5	0.4 ± 0.5	
Group work	[0,10]	6.6 ± 3.3	6.1 ± 3.6	6.5 ± 3.4	
Visual	[0,10]	7.1 ± 2.1	7.3 ± 2.0	7.3 ± 2.1	
Auditory	[0,10]	3.1 ± 2.1	3.1 ± 2.2	3.2 ± 2.2	
Kinaesthetic	[0,10]	4.7 ± 2.4	4.6 ± 2.5	4.4 ± 2.4	
GPA	[0,4]	1.9 ± 1.1	2.3 ± 1.0	2.7 ± 0.9	***

*p<0.05; **p<0.01; ***p<0.001.

significantly lower for males (t(1,158)= -3.595, p<0.001). Gender differences in academic performance were not reflected in factors of temperament or motivation. Females had higher mean scores for *study time* (t(1,065)= -2.988, p=0.003), a *deep learning approach* (t(1,107)= -3.038, p=0.002) and *auditory modality* (t(1,079)= -7.3, p<0.001) while males had higher mean score for a *shallow learning approach* (t(1,170)= 4.723, p<0.001) and *kinaesthetic learning modality* (t(1,131)= 5.175, p<0.001). Group means are given in Table 5.3.

Table 5.3: Group differences by gender, $m \pm s$

Study factor	Range	Male (n=713)	Female (n=494)	p
CAO points	[0,600]	209.9 ± 117.4	231.7 ± 119.1	***
Leaving cert average	[0,100]	31.0 ± 17.5	34.4 ± 17.8	**
Mathematics	[0,100]	19.1 ± 15.5	21.1 ± 15.6	*
English	[0,100]	36.8 ± 24.0	42.3 ± 23.6	***
Applied average	[0,100]	27.1 ± 28.5	24.5 ± 27.4	*
Methodical average	[0,100]	24.6 ± 17.6	30.5 ± 18.9	***
Humanities average	[0,100]	32.2 ± 19.3	35.9 ± 19.6	***
Conscientiousness	[0,10]	6.0 ± 1.5	5.9 ± 1.6	
Openness	[0,10]	6.0 ± 1.3	6.1 ± 1.2	
Self-efficacy	[0,10]	6.9 ± 1.4	6.8 ± 1.4	
Extrinsic goal orientation	[0,10]	7.8 ± 1.4	7.9 ± 1.3	
Intrinsic goal orientation	[0,10]	7.1 ± 1.4	7.1 ± 1.4	
Metacognitive self-regulation	[0,10]	5.9 ± 1.4	5.9 ± 1.4	
Study effort	[0,10]	5.9 ± 1.8	6.0 ± 1.7	
Study time	[0,10]	6.0 ± 2.3	6.4 ± 2.3	**
Deep learner	[0,10]	5.2 ± 3.0	5.7 ± 2.8	**
Shallow learner	[0,10]	1.5 ± 2.1	1.0 ± 1.7	***
Strategic learner	[0,10]	3.4 ± 2.5	3.4 ± 2.5	
Group work	[0,10]	6.5 ± 3.3	6.7 ± 3.4	
Age	[18,60]	23.5 ± 7.7	22.9 ± 6.8	
Visual	[0,10]	7.2 ± 2.1	7.1 ± 2.0	
Auditory	[0,10]	2.8 ± 2.2	3.7 ± 2.1	***
Kinaesthetic	[0,10]	5.0 ± 2.5	4.2 ± 2.2	***
GPA	[0,4]	2.0 ± 1.1	2.2 ± 1.0	***

*p<0.05; **p<0.01; ***p<0.001.

5.4 Regression models predicting year 1 academic performance

Regression models predicting GPA were run for three groups: all participants; participants with prior academic history (*CAO Points*>0); and younger participants (age:[18,21]). The best regression model for all participants predicting GPA ($\bar{R}^2$ = 0.186, MSE = 0.934 ± 1.235) was based on eight factors, namely *age*, *CAO points*, *mathematics, study effort, deep learner, extrinsic goal orientation, gender* and *group work*. Model fit and standardised coefficients are given in Table 5.4. Model fit remained the same when *mathematics* was replaced with *methodical average* or *CAO points* was replaced by either *leaving cert average* or *humanities average*. Prior academic performance was unavailable for some participants (n=189) but was statistically significant in the model. Modelling participants with *CAO points* > 0 improved $\bar{R}^2$ (0.237), although the improvement in MSE (0.863 ± 1.151) was not significant (t(2200)=1.402, p=0.161). $\bar{R}^2$ was comparable to other reported models of a diverse student population [Bidjerano and Dai, 2007; Kaufman et al., 2008; Komarraju et al., 2011; Swanberg and Martinsen, 2010]. Modelling younger students further improved model fit. For example, a model of participants aged [18,21], for whom *CAO points* were available (n=800), had $\bar{R}^2$=0.301 concurring with results from other studies that excluded mature students [Chamorro-Premuzic and Furnham, 2008; Dollinger et al., 2008; Robbins et al., 2004]. Improvement in MSE (0.764 ± 1.009) was not significant when compared to the model of all participants with *CAO points*>0 (t(1794)=1.951, p=0.051), but was statistically significantly better than the model of all participants (t(1919)=3.376, p<0.001).

Age was statistically significant in all three regression models. Factors of prior academic performance were also statistically significant, particularly overall performance (*CAO points* or *leaving cert average*) and *mathematics*. The most significant non-cognitive attribute was a *deep learning approach*. *Extrinsic goal orientation* was also statistically significant across regression models.

Key finding: Regression model results concurred with other studies; models of younger students had higher coefficient of determination than models that included mature students.

Table 5.4: Regression models predicting GPA

	All participants	All ages; CAO points > 0	Age [18,21]; CAO points > 0
R^2	0.191	0.243	0.306
$\bar{R}^2$	0.186	0.238	0.301
MSE	0.934 ± 1.235	0.863 ± 1.151	0.764 ± 1.009
n	1,207	1,018	800
	Standardised model co-efficients		
Intercept	0	0	0
Age	0.431***	0.315***	0.127***
CAO points	0.271***	0.343***	
Leaving cert average			0.442***
Mathematics	0.123**	0.117***	0.105**
Conscientiousness		0.070*	0.117***
Extrinsic goal orientation	0.094***	0.089**	0.075*
Study effort	0.095***		
Deep learner	0.142***	0.111***	0.086*
Group work	-0.088**	-0.048	
Gender	0.076**	0.098***	0.118***

*p<0.05; **p<0.01; ***p<0.001.

5.5 Classification models predicting students at risk of failing

The following sections detail results for classification models of all participants including model accuracy, attributes used and analysis of misclassifications. Models of subgroups by course of study, gender and age are also discussed.

5.5.1 Classification accuracy for models of all participants

Figure 5.3 gives classification model accuracies for models of all participants as a heat map. Two training methods were compared, 10 fold cross validation ($Model_{XVal}$) and models trained on the 2010 and 2011 cohorts and tested on the 2012 cohorts ($Model_{2012}$). As discussed in Section 4.3.4, the dataset was balanced by oversampling the minority class. Reported model accuracies are based on the resulting balanced dataset. However, geometric mean (GM) was calculated after removal of replicated instances from the labelled dataset as explained in Section 3.5.9. This is indicated by the sample size (n) included in Figure 5.3. Contingency tables for McNemar's test (χ^2) and Fisher's exact test (FET) also excluded replicated instances.

Best $Model_{2012}$ accuracy was k-NN (accuracy: 71.98%, GM: 70.35%). However, a comparison of $Model_{2012}$ accuracies using McNamer's test with Holm correction for familywise error indicated model performance was comparable across algorithms. The only statistically significant difference was between LR (accuracy: 65.38%, GM: 61.12%) and k-NN (χ^2 (1, n=436)=15.95, p<0.001). The difference between LR and Voting Ensemble (accuracy: 71.06%, GM: 68.89%) was χ^2 (1, n=436)=4.845 (p=0.03) which was not statistically significant following Holm adjustment for family wise error.

In contrast with $Model_{2012}$ results, best $Model_{XVal}$ accuracy was SVM (81.62%). Its lower GM (72.18%) reflected a higher recall on *pass* (88.24%) than *fail* (59.04%). Given the objective of identifying students at risk of failing, models with a high recall on *fail* are preferable. BPNN also had good accuracy (75.33%) but a lower GM (69.32%) reflecting a higher precision on *pass* (78.53%) than *fail* (57.48%). A comparison of $Model_{XVal}$ accuracies using McNamer's test with Holm correction showed SVM had statistically significantly higher accuracy than all other algorithms. In addition, LR accuracy was statistically significantly lower than other algorithms (accuracy: 66.64%, GM: 63.06%). Differences in the accuracies of the remaining six algorithms was not statistically significant.

$Model_{XVal}$ accuracy was higher than $Model_{2012}$ accuracy for each algorithm used. However, a comparison of the $Model_{2012}$ and $Model_{XVal}$ accuracy for each algorithm found a statistically significant difference for SVM only (p<0.001 FET).

Extensive search strategies were used for optimal attribute subset selection as discussed in Section 3.5.8. This can result in model overfitting [Baumann, 2003; Saeys et al., 2007]. A comparison of cross validation accuracies with ($Model_{XVal}$), and without ($Model_{all}$), attribute subset selection found the difference was statistically significant for two algorithms only, k-NN (χ^2 (1, n=1207)=20.1, p<0.001) and a k-NN Bagging Ensemble (χ^2 (1, n=1207)=6.7, p<0.01). $Model_{all}$ accuracies are included in Figure 5.3.

Key finding: Models of students at risk of failing based on factors measured prior to commencement of first year of study achieved good predictive accuracy. Model accuracies were comparable across a number of classifiers.

5.5.2 Comparison of model predictions

A review of $Model_{2012}$ predictions highlighted that models generally concurred on participant classification, particularly the four models with the highest accuracy, namely k-NN, SVM, BPNN and DT, as illustrated in Figure 5.4. LR and SVM had the highest level of agreement, 79% of their respective predictions concurred. LR and DT has the lowest level of agreement, 66% of their respective predictions concurred.

Algorithm	$Model_{2012}$ Accuracy (%) (*n*=546)	$Model_{2012}$ GM (%) (*n*=436)	$Model_{XVal}$ Accuracy (%) (*n*=1,496)	$Model_{XVal}$ GM (%) (*n*=1,207)	$Model_{All}$ Accuracy (%) (*n*=1,496)
k-NN	71.98	70.35	72.39	71.38	66.12
Voting	71.06	68.89	74.26	68.86	72.73
Bagging	70.51	68.78	71.93	70.51	67.45
SVM	70.33	67.32	81.62	72.18	81.48
DT	69.96	68.38	70.19	69.32	69.59
NB	66.30	64.73	69.52	69.14	68.78
BPNN	68.50	66.29	75.33	68.97	76.40
LR	65.38	61.12	66.64	63.06	64.91

GM: Geometric mean.

Figure 5.3: Heat map of classification model accuracies for all participants

	k-NN	SVM	DT	BPNN	NB
SVM	72				
DT	73	73			
BPNN	75	75	72		
NB	67	69	69	74	
LR	67	79	66	67	67

Figure 5.4: Concurrence between classification algorithms, measured as the percentage of instances with the same predicted class label

5.5.3 Optimal attribute subsets used

Table 5.5 illustrates optimal attribute subsets for each of the eight $Model_{2012}$ classification algorithms. The five factors most frequently used were *age*, *methodical average*, *leaving cert average*, *self-efficacy* and *kinaesthetic modality*. *English* and *auditory modality* were ignored by all algorithms while *CAO points*, *mathematics*, *conscientiousness* and *study time* were each selected by just one algorithm.

Classification algorithms differ with respect to insights provided on how instances are classified. Decision trees are relatively easy to interpret [Tan et al., 2014, p. 169], however the numeric attributes in the study dataset generated a relatively large decision tree incorporating many split points on each attribute which limited is usefulness as an descriptive tool (see Appendix C, Figure C.6). The relative magnitude of attribute weights in NN, SVM and LR models also provide insight into the influence of each attribute on how instances were classified; weights close to zero indicate redundant features while higher positive or negative weights indicate a greater influence in model predictions [Tan et al., 2014]. Table 5.5 includes attribute weights from SVM, BPNN[1] and LR; a visualisation of BPNN model weights is included in Appendix C (Figure C.7). However, correlations between study attributes meant some model attributes could be replaced without effecting model fit. For example, replacing *conscientiousness* (SVM weight=0.17) with both *study effort* and *study time* in the SVM model gave the same model accuracy and a weight of -1.3 for *study effort*. Therefore, exclusion of an attribute did not indicate a weight close to 0.

To compare the predictive accuracy of non-cognitive factors of learning with the predictive accuracy of factors available from student registration, three models were compared: a k-NN model trained on factors of prior academic performance, *age* and *gender* only ($Model_{Prior}$), a k-NN model trained on non-cognitive learning factors, *age* and *gender* only ($Model_{NCog}$) and a k-NN model trained on all attributes ($Model_{2012}$). Models were trained on 2010 and 2011 data and tested on the 2012 data. Forward selection was used for attribute subset selection and models were trained on values of k in the range [10, 30]. Accuracies were compared using McNemar's test with Holm correction. $Model_{Prior}$ accuracy (70.33%) was marginally lower than $Model_{2012}$ accuracy (71.88%), the difference was not statistically significant (p=0.44). $Model_{NCog}$ accuracy (64.10%) was statistically significantly lower than $Model_{2012}$ (p=0.04) but was not statistically significantly lower than $Model_{Prior}$ (p=0.12). A subset of four attributes was used in $Model_{Prior}$, namely *leaving cert average*, *methodical average*, *age* and *gender*. A subset of nine attributes were was in the $Model_{NCog}$, namely *conscientiousness, self-efficacy, intrinsic* and *extrinsic goal orientation, study effort, deep* and *shallow learning approaches, age* and *gender*.

Key findings:

The five attributes most predictive of students at risk of failing were *age*, *methodical average*, *leaving cert average*, *self-efficacy* and *kinaesthetic modality*.

Improvement in model accuracy attributed to non-cognitive factors of learning was not statistically significantly.

[1] Attribute weights for BPNN in Table 5.5 depict the highest weight for that attribute.

Table 5.5: Attributes used by $Model_{2012}$ models with model weights

Attribute	SVM	Vote	BPNN	k-NN	DT	LR	Bagging	NB	*Total*	Weights SVM	Weights BPNN*	Weights LR
Age	x	x	x	x	x	x	x		7	1.14	21.02	1.04
Methodical average	x	x	x	x			x	x	6	0.50	23.07	
Leaving cert average	x	x	x		x	x	x		6	0.34	20.31	1.04
Self-efficacy	x	x		x	x	x	x		6	0.09		0.05
Kinaesthetic	x	x	x	x	x	x			6	-0.07	7.61	-0.06
Humanities average	x	x	x	x		x			5	0.06	19.92	-0.31
Intrinsic goal orientation	x	x	x		x	x			5	0.19	9.65	0.18
Openness	x	x	x		x				4	0.08	4.32	
Gender	x	x	x			x			4	0.19	17.22	0.17
Deep learner		x	x			x			3		13.31	0.32
Applied average				x			x		2			
Extrinsic goal orientation	x							x	2	0.03		
Metacognitive self-regulation	x					x			2	-0.09		-0.06
Study effort		x		x					2			
Shallow learner	x		x						2	-0.04	14.87	
Group work	x						x		2	0.01		
Visual				x	x				2			
Strategic learner		x	x						2		8.21	
Mathematics					x				1			
CAO points				x					1			
Conscientiousness	x								1	0.17		
Study time					x				1			
Auditory									0			
English									0			
Total	14	12	11	9	9	9	6	2				

*Maximum weight across 8 hidden neurons; x: Attribute was included in the model; Dashed lines were added to improve readability.

5.5.4 Group differences between misclassifications and correct predictions

As discussed in Section 5.5.2 algorithm predictions generally concurred, this included participants misclassified. For example, 75% of participants misclassified by k-NN were misclassified by at least four of the eight algorithms used. Therefore, student misclassification was defined as an instance misclassified by at least four of the eight algorithms. The resulting confusion matrix identified four groups as illustrated in Table 5.6: 166 students correctly predicted as fail (*True Fail*); 186 students correctly predicted as pass (*True Pass*); 87 students incorrectly predicted as fail (*False Fail*); and 47 students incorrectly predicted as pass (*False Pass*). Group differences detailed below concurred with a similar analysis of k-NN misclassifications.

Of specific interest was each misclassified group and how they differed from correctly classified participants. A Shapiro-Wilk test of fifty bootstrap samples of each group verified group means were normally distributed for each study factor but a Levene's test found variances were unequal. Therefore, Welch's t-test was used to compare each misclassified group with correctly classified participants. Results and group means are given in Table 5.7. *False Pass* had similar mean scores to *True Pass* in all study attributes. However, *False Pass* had statistically significantly higher mean scores than *True Fail* in a number of non-cognitive factors of learning, *methodical average* and *mathematics*. *False Fail* had similar mean scores to *True Fail* in all factors except *CAO points*, *humanities average* and *English*, and had a statistically significantly lower GPA than *True Pass*.

Outliers were not a factor in misclassifications. Outlier detection suggested a small portion of the dataset (n=21, 2%) may contain outlier values as discussed in Section 4.2.3. Just one univariate outlier was misclassified, an instance with a low *self-efficacy* score was misclassified as *fail*. Misclassified instances had the the same average LOF score (1.09±0.92) as correctly classified instance (1.09±0.98). LOF calculations were detailed in Section 4.2.3.

Table 5.6: Confusion matrix illustrating misclassifications

	Predicted Fail	Predicted Pass	*Recall*
Actual Fail	116	47	71.17%
Actual Pass	87	186	68.13%
Precision	57.14%	79.83%	69.27%

Table 5.7: Group differences for misclassified participants, including group means ($m \pm s$)

Attribute	*True Fail* (n=116)	*False Fail* (n=87)	*False Pass* (n=47)	*True Pass* (n=186)	*False Pass* compared with: *True Fail*	*True Pass*	*False Fail* compared with: *True Fail*	*True Pass*
CAO points	207.80±66.27	235.98±60.94	236.40±130.92	236.27±141.43			**	
Mathematics	16.45±11.51	17.76±12.08	23.00±16.67	23.72±17.28	*			**
English	40.66±18.81	45.81±15.19	37.34±25.64	42.63±27.09			*	
Methodical average	21.41±10.50	24.28±10.59	33.15±19.61	32.97±20.96	***			***
Creative average	31.69±26.67	32.98±27.95	25.53±26.65	22.31±29.39				**
Humanities average	32.77±12.49	37.22±11.55	36.45±22.00	35.94±23.01			**	
Conscientiousness	5.89±1.68	5.90±1.45	6.38±1.55	6.47±1.42				**
Openness	5.94±1.42	5.82±1.13	6.14±1.37	6.20±1.35				*
Self-efficacy	6.72±1.35	6.83±1.39	7.32±1.30	7.27±1.27	**			*
Intrinsic goal	6.76±1.46	6.7±1.33	7.45±1.26	7.32±1.25	**			***
Extrinsic goal	7.46±1.50	7.79±1.24	8.01±1.23	8.01±1.15	*			
Metacognitive self-regulation	5.66±1.51	5.55±1.18	6.28±1.33	6.24±1.31	*			***
Study effort	5.39±1.94	5.65±1.64	6.35±1.54	6.26±1.74	**			**
Study time	6.06±2.49	6.04±2.07	6.14±2.27	6.53±2.18				
Deep learner	4.78±2.61	5.06±2.61	6.22±2.32	6.22±2.96	***			**
Strategic learner	3.73±2.43	3.79±2.44	2.71±2.07	2.81±2.51	**			**
Shallow learner	1.49±1.84	1.15±1.74	1.06±1.54	0.97±1.58				
Age	20.06±2.12	20.40±2.14	24.34±8.05	24.73±8.41	***			***
Group work	7.09±3.35	6.32±3.23	6.81±3.16	6.60±3.29				
Visual	7.13±2.04	7.27±2.14	7.18±1.78	7.16±2.08				
Auditory	3.06±2.11	3.36±2.15	3.14±2.24	3.07±2.15				
Kinaesthetic	4.81±2.38	4.37±2.70	4.68±2.31	4.77±2.49				
GPA	0.83±0.70	2.64±0.37	1.03±0.76	2.83±0.41		***	***	***

*p<0.05; **p<0.01; ***p<0.001.

Key finding: The misclassifications of most concern are *false pass*. Analysis of group differences between instances incorrectly predicted as *pass* and instances correctly predicted as *pass* showed a lack of statistically significant differences between the two groups. Conversely, differences between instances incorrectly predicted as *pass* and instances correctly predicted as *fail* were statistically significant in a number of effective learning dispositions and measures of prior academic performance.

5.5.5 Classification models of subgroups

Classification models were trained for subgroups by academic course of study (12 subgroups), gender (2 groups) and age (3 subgroups). As with the full dataset, eight classification algorithms were trained on 2010 and 2011 data and tested on 2012 data. Subgroup model accuracies were higher than models trained on the full dataset, particularly for smaller groups as illustrated in the heat map in Figure 5.5. k-NN achieved the highest accuracy, or close to the highest accuracy, for each subgroup. Optimal values for k varied, but were predominantly in the range [7,18]. As discussed in Section 4.3.4 class imbalance varied, the relative size of the minority class varied from 15% to 48%.

All subgroups had $n < 900$ and so were below the minimum sample size required to accurately model study factors as discussed in Section 4.3.2. Therefore, to assess if insufficient data contributed to high model accuracies, model accuracy for each subgroup was compared to the mean accuracy of models trained on 50 random bootstrap samples of the same size, selected from all participants. The random samples were constructed to match the class balance of their corresponding subgroup, i.e. the numbers of passes and fails matched for test datasets (n_{test}) and training datasets. Models of random samples were trained using k-NN with forward selection for attribute subset selection, and values of k in the range [2, 20]. As illustrated in Table 5.8, average model accuracies for random samples ($Model_{ran}$) were also higher than models of all participants, and in many cases were similar to their corresponding subgroup model accuracy. Fisher's exact tests[1] comparing the highest accuracy for each subgroup ($Model_{sub}$) with $Model_{ran}$ found differences were not statistically significant for any subgroup, however the power of the test was low for small sample sizes. For example, the largest difference in accuracy was between students in the age group [24,28] (89.66%, n_{test}=44) and random samples of the same size (83.07%, p=0.42 FET), however the power of this test was 0.09, indicating a 9% chance of detecting

[1]Fishers exact test (FET) assumes independent samples. It's possible that some instances from $Model_{XVal}$ were replicated in $Model_{ran}$, violating the assumptions of FET. Significance concurred with results from a selection of other statistical tests including chi-squared and Z-score probability.

Sub group	*n*	*k*-NN	Bagging	Voting	BPNN	DT	LR	SVM	NB
IT	137	88.10	85.71	73.81	80.95	83.33	69.05	69.05	71.43
CDM	102	84.48	84.48	81.03	82.76	79.31	74.14	82.76	74.14
Elec	52	96.15	96.15	92.31	88.46	88.46	80.77	88.46	88.46
EngC	73	90.00	90.00	82.00	84.00	88.00	82.00	84.00	76.00
Hort	41	100.00	100.00	85.71	92.86	92.86	85.71	92.86	78.57
BGen	183	85.94	87.50	87.50	82.50	77.50	82.50	85.00	87.50
BwIT	60	93.33	96.67	90.00	86.67	84.62	86.67	86.67	86.67
BInt	64	92.31	92.31	80.77	84.62	92.31	61.54	76.92	80.77
Sports	95	85.94	84.38	84.38	76.56	73.44	81.25	79.69	78.12
ASC	146	82.20	82.20	79.66	77.97	80.51	66.95	72.03	77.12
EC&E	80	81.25	79.69	81.25	81.25	79.69	78.12	78.12	78.12
SCD	127	81.43	75.71	75.71	74.29	72.86	75.71	75.71	75.71
Males	713	73.79	71.72	72.41	70.34	67.93	66.21	72.07	72.07
Females	494	75.00	72.27	74.22	71.48	71.09	66.80	73.44	76.56
[18,23]	875	73.86	72.84	74.11	73.60	69.80	70.05	74.87	72.59
[24,28]	131	89.66	79.31	68.97	79.31	77.59	77.59	81.03	81.03
[29,60]	210	81.91	80.85	80.85	76.60	74.14	68.09	76.60	70.21
Average accuracy		85.61	84.22	80.28	80.25	79.61	79.58	79.37	77.95

IT: Computing (IT); CDM: Creative Digital Media; Elec: Electronic and Computer Engineering; EngC: Engineering Common Entry; Hort: Horticulture; BGen: Business General; BwIT: Business with IT; BInt: International Business; Sports: Sports Management & Coaching; ASC: Applied Social Care; EC&E: Early Childcare & Education; SCD: Social and Community Development.

Figure 5.5: Heat map of classification model accuracies (%) for subgroups by course of study, gender and age

a difference.[1]

> **Key finding:** Sample sizes were too small to draw conclusions on subgroup model accuracies.

[1] Power was calculated using *power.fisher.test* in the statmod package V1.4.20 for R. A lower power increases the chance of a Type II error, i.e. failing to recognise a statistically significant difference [Chatfield, 1983, p 158].

Table 5.8: Comparison of subgroup model accuracies with random samples of the same size

Subgroup	$n_{subgroup}$	n_{test}	$Model_{sub}$ (%)	$Model_{ran}$ $m \pm s$ (%)
Computing (IT)	137	38	88.10	84.28 ± 4.54
Creative Digital Media	102	36	84.48	85.88 ± 4.61
Electronic & Computer Engineering	52	20	96.15	92.06 ± 4.83
Engineering Common Entry	73	32	90.00	89.04 ± 5.02
Horticulture	41	10	100.00	96.49 ± 4.68
Business General	183	40	87.50	81.67 ± 3.37
Business with IT	60	26	96.67	90.51 ± 4.86
International Business	64	20	92.31	89.17 ± 5.14
Sports Management & Coaching	95	43	85.94	87.37 ± 5.02
Applied Social Care	146	74	82.20	82.52 ± 3.95
Early Childcare & Education	80	40	81.25	87.70 ± 4.67
Social & Community Development	127	56	81.43	84.68 ± 4.57
Males	713	243	73.79	72.95 ± 1.83
Females	494	193	76.56	74.98 ± 3.08
Age group: [18,23]	875	334	74.87	72.98 ± 2.08
Age group: [24,28]	131	44	89.66	83.07 ± 4.09
Age group: [29,60]	210	58	81.91	80.20 ± 3.50

$n_{subgroup}$: Subgroup size (test and training dataset); n_{test}: Test dataset size.

5.6 Summary

Correlations between study attributes concurred with other studies with the exception of correlations between factors of self-regulation and the correlation between *self-efficacy* and GPA. Group differences by GPA band confirmed that both prior academic performance and non-cognitive factors associated with an effective learning disposition differentiated high risk from low risk students. Older students also had higher mean scores for factors associated with an effective learning disposition. Male participants had lower prior academic performance and a lower GPA than female participants, however, results may have been influenced by low entry points for male dominated courses. Regression models confirmed that models of younger participants (age≤21) had higher R^2 than models of all participants. Factors of prior academic performance, *age*, a *deep learning approach* and *extrinsic goal orientation* were significant in regression models.

Cross validation results indicated SVM had highest cross validation model accuracy (82%) when modelling all participants using a binary class label of *pass* or *fail*; this was statistically significantly higher than other models although recall on *fail* (59%) was relatively poor. k-NN had highest accuracy (72%) when models were trained on 2010 and 2011 student cohorts and tested on the 2012 student cohort, which was similar to its cross validation accuracy (72.4%). The most significant attributes in classification models were *age*, *methodical average*, *leaving cert average*, *self-efficacy* and *kinaesthetic modality*. A comparison of the group means of all study attributes showed participants incorrectly classified as *pass* did not differ significantly from participants correctly classified as *pass*. Participants incorrectly predicted as *fail* had a statistically significantly lower GPA than participants correctly predict as *pass*. Classification model accuracy was higher for subgroups by course of study, gender and age. However, relative accuracies for each subgroup were similar to the mean model accuracy of 50 random samples of the same size.

Key findings cross referenced to study objectives are as follows:

- *Ob4*: Correlations and regression analysis of cognitive and non-cognitive study factors concurred with other studies with the exception of correlations between factors of self-regulation, and the correlation between *self-efficacy* and GPA.

- *Ob5*: Results from training and evaluation of a range of classification models predicting students at risk of failing found that models predicting students at risk of failing, based on factors measure prior to commencement of first year of study, achieved good predictive accuracy. Participants incorrectly classified as *pass* did not differ significantly from participants correctly classified as *pass*; participants incorrectly predicted as *fail* had statistically significantly lower GPA than participants correctly predict as *pass*.

- *Ob6*: A review of key cognitive and non cognitive factors of learning that are predictive of first year students at risk of failing found both prior academic performance and non-cognitive factors associated with an effective learning disposition differentiated high risk from low risk students. *Age*, *methodical average*, *leaving cert average*, *self-efficacy* and *kinaesthetic modality* were selected by most classification models. However, improvement in model accuracy attributed to non-cognitive factors of learning was not statistically significantly.

A discussion of these results, and recommendations for future work, are presented in Chapter 6.

Chapter 6

Discussion

6.1 Introduction

The key research question this study addressed was: (*Q1*) *Can algorithmic student modelling accurately predict Irish IoT students at risk of failing in first year of study based on factors that can be measured prior to commencement of tertiary education?* Results from this study indicated that algorithmic student modelling can accurately predict students at risk of failing in first year of study based on factors that can be measured prior to commencement of tertiary education. Models trained on data collected prior to, or during, first year student induction achieved good predictive accuracy when applied to a different student cohort. The following sections discuss the salient outcomes from this study, and suggest directions for future work. The study's two secondary research questions are also addressed, namely: (*Q2*) *Which classification algorithms are appropriate for modelling psychometric data indicative of Irish IoT students at risk of failing?* and (*Q3*) *Which cognitive and non-cognitive factors of learning are indicative of Irish IoT students at risk of failing?*

6.2 Classification models of students at risk of failing

The study dataset was diverse in terms of student age, prior academic ability and course of study (illustrated in Tables 3.1, 3.2 and 3.5). Modes of assessment also varied across courses. For example, modules on humanities courses were more likely to give a higher weighting to end of term examinations, modules on other courses were more likely to give equal or higher weighting to continuous assessment work. Assessment methods can affect academic performance [Pérez-Martínez et al., 2009] and its relationships with factors such as openness and learning approach as discussed in Sections 2.3 and 2.5.1. Notwithstanding these sources of variability, classification model accuracy was high (71.98%) when applied to a different student cohort ($Model_{2012}$).

As reported in Section 2.8.2, a number of studies modelling educational data have cited comparable accuracies between classification algorithms, although there are inconsistencies regarding which algorithms achieve optimal predictive accuracy when modelling educational data. For example, Jayaprakash et al. [2014] reported SVM and LR had comparable accuracy predicting high risk students based on factors including SAT scores, enrolment data and data from an online learning environment, but DT had poorer recall (n=15,150). On the other hand, Lauria et al. [2013] reported DT had comparable accuracy with both SVM and LR when distinguishing between strong and weak students, also based on prior academic performance, demographic data and log data from an intelligent tutoring system (n=6,445). Herzog [2006] found DT and BPNN had similar performance to LR provided independent variables had little co-linearity, but LR had lower accuracy when variables with greater dependencies were included in the model (n=4,564). Bergin [2006] found NB and a Stacking Ensemble outperformed DT, BPNN, k-NN (k=3) and LR when classifying students as strong or weak based on prior academic performance and psychometric data (n=102). Results from this study concurred that a number of classification algorithms achieved similar accuracy. It was expected that some algorithms would concur on the predicted class label. For example, the implementation of LR used in this study was based on an adaption of an SVM algorithm proposed by Keerthi et al. [2005], explaining the high concurrence between LR and SVM predictions as illustrated in Figure 5.4; SVM's higher accuracy can be explained by its use of structural risk minimisation guaranteeing a globally optimal solution [Tan et al., 2014, p. 276]. On the other hand, a DT's approach to the classification task is quite different from SVM as discussed in Sections 2.8.1 and 3.5, nevertheless, both algorithms concurred on 73% of predictions when applied to a different student cohort. (Misclassifications are discussed further in Section 6.3.)

The models with the highest accuracies when applied to a different student cohort were k-NN, a Voting Ensemble, SVM, a k-NN Bagging Ensemble and a DT, as illustrated in Figure 5.3. LR had the lowest accuracy when modelling all participants. Additionally, progressive sampling failed to confirm convergence of LR model accuracy. k-NN and a Bagging ensemble had the highest average accuracy for models of subgroups, although a larger sample size was needed to confirm subgroup model accuracies.

While model accuracy estimated using 10-fold cross validation ($Model_{XVal}$) was higher than model accuracy when tested on a different student cohort ($Model_{2012}$), the increase was statistically significant for one algorithm only, SVM. As discussed in Section 3.5.6, a kernel function improved SVM model accuracy for $Model_{XVal}$, generating a more complex model than $Model_{2012}$ where a kernel function failed to improve model accuracy. In contrast, optimal parameter settings for both $Model_{2012}$ and $Model_{XVal}$ were similar for other algorithms, for example k in k-NN (15 and 18 respectively). Therefore, results indicated that cross validation provided a good estimate of model accuracy with the exception of a nonlinear SVM model.

Baumann [2003] reported that attribute subset selection techniques used in combination with cross validation resulted in model over-fitting, particularly when a large range of alternatives were assessed, but this was not observed in this study. A comparison of cross validation accuracies with and without attribute subset selection observed that increases in model accuracy were not significant with the exception of k-NN based models, as illustrated in Figure 5.3. k-NN used a Euclidean distance measure calculated from all attributes in the dataset; unlike the other classification algorithms, each attribute had equal influence on the outcome [Han and Kamber, 2006, p. 349]. This, rather than over-fitting, may explain accuracy improvement when using an attribute subset.

In response to research question *Q2*, study results suggested k-Nearest Neighbour with attribute subset selection is the most appropriate algorithm for modelling psychometric data indicative of IoT students at risk of failing. However, it must be acknowledged that with the exception of LR, other classifiers achieved comparable accuracies. Ensembles failed to improve on base model accuracies.

6.3 Analysis of misclassifications

The misclassifications of particular interest were participants incorrectly predicted as *pass*. Group comparisons of study attributes failed to identify differences between this group and those correctly predicted as *pass*. There may be a number of reasons for this. Firstly, a number of factors relevant to retention and progression arise after student induction, such as academic and social integration, change in circumstance resulting in economic pressure [Tinto, 2006], and classroom related affects on academic performance such as teaching methods [Ganyaupfu, 2013; Hake, 1998]. Such factors may explain why prior academic performance and learning disposition alone are insufficient to predict academic performance in all cases. Secondly, it could be argued that profiling learners during first year student induction is too early in the semester to accurately measure some study attributes. For example, *intrinsic* and *extrinsic goal orientation* may vary depending on the time or situation [Apter, 1989]. Similarly, students may be unsure of study expectations during the initial period of induction; Winters et al. [2008] concluded both learner and task characteristics influenced levels and methods of self-regulation. On the other hand, factor correlations in this study (discussed in Section 5.2) concurred with evidence cited in other studies where data were gathered later in the semester. Therefore, it was more likely that factors not included in the study explained incorrect predictions of *pass*. Individual differences dictate that any deterministic model of human behaviour will have some level of pure error [Eysenck and Keane, 2005, p. 5]; further work is needed to accurately determine the potential improvements in model accuracy if additional data gathered after student induction were included.

Analysis of participants incorrectly predicted as *fail* showed they had a lower GPA than students correctly predicted as *pass*. Scores in a range of effective learning dispositions and prior academic performance were also lower, characterising a group of lower academic achievers that may benefit from additional support to develop an effective learning disposition.

6.4 Impact of study attributes

The recorded significant correlations between many study attributes meant different attribute subsets achieved comparable accuracies when predicting GPA. Consequently, no common subset of attributes could be isolated for use in this project. For example, a *deep learning approach* is associated with *intrinsic learning goals* (see Section 2.5.1) and correlations between the two factors in this study were relatively good (r=0.417, 95% B-CI [0.37,0.47]). Six of the eight models used either a *deep learning approach* or *intrinsic goal orientation*, but just two models, BPNN and LR, used both. The k-NN model used neither of these, but instead used *self-efficacy*, which had a relatively high correlation with *intrinsic goal orientation* (r=0.421, 95% B-CI [0.37,0.47]), and *study effort*, which had a relatively high correlation with a *deep learning approach* (r=0.360, 95% B-CI [0.31,0.41]). How models identify patterns also impacts on attributes used. For example, in the regression models, *conscientiousness* could be replaced by *study effort* without effecting model fit (see Section 5.4). However, those attributes were not interchangeable in a k-NN model as their respective effects on distance calculations were not equivalent. Nevertheless, it is clear from reviewing attribute subsets used (Table 5.5) that some attributes achieved higher accuracy in prediction of students at risk of failing than others. The following sections outline the study factors and assess their relative usefulness in predicting students at risk of failing, responding to research question *Q3*.

6.4.1 Age and gender

Existence of correlations between *age* and academic performance is well cited in literature (Cassidy [2011]; Hoskins et al. [1997]; Wigfield et al. [1996]), and is also supported by evidence from this study. It is evident from analysis of group differences discussed in Section 5.3 that older students have a more effective learning disposition: they are more likely to adopt a deeper learning approach, set learning goals and regulate their learning. Both classification and regression models concurred that *age* was a good predictor of academic performance in first year of study (see Table 5.4 and Figure 5.3).

A number of studies reported that *gender* is not a significant factor in predicting academic performance in tertiary education ([Dollinger et al., 2008; Hoskins et al., 1997; Naderi et al., 2009]). Four of the eight algorithms included *gender*, in spite of its relatively low correlation with GPA (r=0.1, 95% B-CI [0.05,0.15]). Gender group differences highlighted males as having lower prior academic performance, lower GPA, and lower scores in a number of effective learning dispositions. However, it should be noted that the dataset contained bias: courses that were predominantly male had lower entry requirements than courses that were predominantly female. Therefore, further work is needed to assess if the significance of *gender* in classification models was reflective of their course of study rather than actual gender differences.

6.4.2 Prior academic performance

Aggregate scores of prior academic performance, particularly *methodical average* and overall average (*leaving cert average*), were found to be more predictive of students at risk of failing first year of study than individual school leaving certificate grades in either *English*[1] or *mathematics*. *Methodical average* had the highest correlation with GPA, and was used by most classification models. While *methodical average* could replace *mathematics* in regression models without changing model fit (see Section 5.4), the same replacement in classification models reduced model accuracy. For example, k-NN model accuracy dropped from 71.98% to 66.29% when *methodical average* was replaced by *mathematics* (McNemar's χ^2 (1, n=436)=7.33, p<0.01). Therefore, while both factors displayed comparable correlations with GPA, *methodical average*, an aggregate covering results in mathematics, science and business subjects, achieved significantly higher accuracy statistically, when predicting students at risk of failing, than *mathematics* alone.

[1]As explained in Section 3.3.1, *English* refers to grades in a subject aimed at developing: a mature and critical literacy; a respect and appreciation for language; and an awareness of the value of literature (`www.education.ie`).

A number of classification models included the factor *humanities average*, particularly models with higher accuracies such as k-NN, Voting Ensemble, SVM and BPNN. Analysis of correlations with GPA by course of study showed *humanities average* to be more predictive of GPA in business and humanities courses than courses in engineering and computing (see Appendix C, Figures C.4 and C.5). Correlations between leaving certificate *English* and GPA also varied by discipline. *English* had a statistically significant positive correlation with GPA for humanities courses and some business courses, but had a statistically significant negative correlation with GPA for Computing (IT). Correlations between *English* and GPA for other technical disciplines were not statistically significant ($p>0.05$). Therefore, it was unsurprising that leaving certificate *English* was not a significant factor in classification or regression models of all participants.

An aggregate of points attained in school leaving cert examinations (*CAO points*) dictate entry requirements for most tertiary courses in Ireland. Results from this study suggested other prior academic performance aggregates were better predictor of students at risk of failing in year 1. Further work is needed to investigate which aggregates are most appropriate for predicting at-risk students. The study sample size was too small to draw conclusions on prior academic performance predictors for subgroups by course of study. However, correlation results by course of study (Figure C.5) indicated an investigation of appropriate aggregates by course of study is also warranted.

6.4.3 Factors of personality

As discussed in Section 2.3, conscientiousness is the best personality based predictor of academic performance in tertiary education particularly for younger students [Allick and Realo, 1997; Chamorro-Premuzic and Furnham, 2008; Kappe and van der Flier, 2010; Kaufman et al., 2008]. While correlation results from this study concurred with the available evidence, all classification models except SVM ignored the factor *conscientiousness*. This suggested that other study factors accounted for conscientiousness, and there was no additional predictive value in measuring conscientiousness specifically.

Openness was used by four classification models (k-NN, BPNN, SVM and voting), suggesting it was a useful predictor of students at risk of failing as evidenced by statistically significant group differences in *openness* for low risk students (GPA$>$ 2.5). The mix of assessment methods used across courses and within courses may explain the low correlation between *openness* and GPA and its absence in regression models. Openness is the most controversial of the Big Five personality factors, in terms of defining both meaning and sub-factors [de Raad and Schouwenburg, 1996, p. 321]. The six-question scale used in this study covered four sub-factors: creativity (2 questions), intellect (2 questions), imagination and openness to new experience (1 question each). Creativity, specifically, is frequently cited as an effective learning disposition that is to be encouraged and promoted in assessment design [Buckingham Shum and Deakin Crick, 2012]. Further work is required to investigate if sub-factors inherent in openness may be more appropriate predictors of academic performance than generalised openness itself.

6.4.4 Factors of motivation

Results from this study support findings by Robbins et al. [2004] regarding the importance of *self-efficacy*; it was the single most predictive non-cognitive factor of students at risk of failing. Interestingly, Deakin Crick and Goldspink [2014] observed that being able to express confidence in learning ability was also a strong indicator of an effective learning disposition. However, while *self-efficacy* was important in classification models, in contrast with other studies (e.g. Bergin and Reilly [2005]) it had lower than expected correlations with GPA and was not significant in regression models. *Self-efficacy* is partly informed by prior academic achievements [Diseth, 2011]. Participants in this study represented a student cohort that had relatively weak prior academic performance as discussed in Section 3.3.1. Further work is needed to determine if the participant profile explains the insignificance of *self-efficacy* in statistical models.

Intrinsic goal orientation was also indicative of good academic performance, however, inferences on its relative importance must consider the poor reliability for that factor as discussed in Section 3.3.2.2.

6.4.5 Learning strategies

Classification models largely ignored factors of self-regulation, although *study effort* was significant in a regression model of all participants, as illustrated in Table 5.4. The importance of self-regulated learning is well cited (see Zimmerman [1990]), however, self-regulation is related to a number of other factors of learning. For example, in a longitudinal study on the causal dilemma between motivation and self-regulation, De Clercq et al. [2013] concluded that a learning goal orientation resulted in a deep learning approach, which in turn resulted in better self-regulation. Self-regulation, as measured in this study, failed to improve classification model performance over and above factors of motivation and approaches to learning. In addition, the poor reliability for *study time* discussed in Section 3.3.2.2 may have diminished its usefulness as a predictor of students at risk of failing.

A *deep learning approach* yielded a higher correlation with GPA than other non-cognitive study factors (see Figure 5.2) and was the only non-cognitive factor of learning with a statistically significant difference in mean score across all three GPA bands of high risk, medium risk and low risk participants (Table 5.1). However, only half of classification models used learning approach: BPNN, SVM, LR and a Voting Ensemble (Table 5.5). Volet [1996] found that goal setting influences self-regulation, which in turn influences learning approach adopted. Similarly in this study, a *deep learning approach* had strongest correlations with *intrinsic goal orientation* (r=0.417), *metacognitive self-regulation* (r=0.431) and *study effort* (r=0.360) as illustrated in Figures 5.2 and C.2. With the exception of a Bagging Ensemble, all models that found approaches to learning to be unimportant, showed that either *goal orientation* or *study effort* were significant. Additionally, apart from a Voting Ensemble, no model used both learning approaches and *study effort*, although a number of models used learning approaches with learning goals.

6.4.6 Learner modality

While awareness of learner modality by student and lecturer can improve student-learning experience [Duffin and Gray, 2009a; Gilakjani, 2012], there is no evidence to suggest that learner modality is predictive of academic performance [Gilakjani, 2012; Kablan, 2014]. Correlation results (Figure 5.1) and analysis of group differences (Table 5.1) concurred with this observation for both *visual* and *auditory modality*. However *kinaesthetic modality* (learn by doing) had a weak but statistically significant correlation with GPA (r=-0.059, p <0.05). In addition, there were statistically significant group differences for *kinaesthetic modality* by GPA band (p<0.05). *Kinaesthetic modality* was significant in six of the eight classification models. Kinaesthetic learners were more likely to be male, more likely to have registered poor prior academic performance and displayed a weak but negative correlation with factors related to an effective learning disposition. Further work is needed to determine the real importance of *kinaesthetic modality*, i.e. if it was used as a proxy for a poor learning disposition, or was itself indicative of failure. In addition, study results justify a review of potential improvements in academic performance achievable from greater use of ‘learn by doing’ activities in the classroom.

6.5 Malleable learner dispositions

Comparison of models with and without non-cognitive factors of learning suggested that the addition of non-cognitive factors of learning provided limited improvement in predictive accuracy in spite of their significant correlations with GPA (see Section 5.5.3).

Therefore, their value in learner profiling at student induction merits consideration. Several studies have reported that learning disposition is malleable. For example, Miller-Reilly [2006] evidenced that teaching approaches can change adult learners' self-efficacy in mathematics. Similarly, a meta analysis of studies on self-regulation reported improvements in self-regulation, and consequently learning goals, following self-regulation training and support [Winters et al., 2008]. It could be argued that all students at risk of failing require further support in developing effective learning dispositions regardless of individual profiles. However, a profiling tool facilitates feedback to both students and lecturers that may support other interventions. The profiler used in this study gave immediate feedback to students on their learner profile. Duffin and Gray [2009b] found that 56% of students understood their learning profile based on online feedback, and this rose to 83% when profiling was followed up by explanatory workshops. Jayaprakash et al. [2014] found that simply making students aware that they may be at risk of failing significantly increased the numbers passing and number of withdrawals; however, provision of further course supports did not effect additional change in either outcome. Therefore, further work is needed to assess the impact of timely feedback on learner disposition, specifically on subsequent optimal use of that feedback.

6.6 Conclusion

Models of learning developed in this study predicted students at risk of failing in first year of study with an accuracy of 72% when applied to a different student cohort. The dataset was diverse in terms of age, academic discipline and assessment strategies used (n=1,207). Informed by a review of factors predictive of academic performance in tertiary education, study factors related to prior academic performance, personality, motivation, learning strategies, learner modality, age and gender. The twenty-four study factors were measured prior to commencement of first year of study. Correlations between study factors were similar to results reported in other studies with the exception of *study time*. In addition, *self-efficacy* had a lower correlation with GPA compared with other studies.

A review of educational data mining studies highlighted a predominance of classification algorithms, although there was a lack of consensus on which algorithms yielded the best accuracy when modelling non-temporal datasets. Consequently, eight classification algorithms were evaluated for this study, namely: k-Nearest Neighbour (k-NN); a Voting Ensemble; Support Vector Machine (SVM); k-NN Bagging Ensemble; Decision Tree; Back-Propagation Neural Network; Naïve Bayes and Logistic Regression (LR). k-NN achieved the highest accuracy when models were applied to a difference student cohort, although accuracies achieved by other algorithms were comparable; LR achieved the lowest accuracy. As expected, 10-fold cross validation model accuracies were higher than models applied to a different student cohort, however, the increase in model accuracy was statistically significant for SVM only.

Analysis of misclassifications showed that *fails* misclassified as *pass* did not differ in learning disposition or prior academic performance, from those correctly classified as *pass*. Factors measurable later in the semester, such as academic and social integration, economic pressures, and teaching methods, may explain misclassifications. Further work is needed to determine potential improvements in model accuracy for an Institute of Technology student cohort if data gathered after student induction were included in the model.

Attributes that were significant for accurate prediction of students at risk of failing across a range of courses included:

- Age. The study sample had an age range of [18,60]. Younger students had a greater risk of failing in first year of study.
- Aggregates of prior academic performance. In particular, an aggregate of mathematics, science and business related subjects was found to be a stronger predictor of year 1 students at risk of failing compared to other prior academic performance aggregates.
- Factors of motivation, particularly self-efficacy and intrinsic goal orientation.
- Learning approach. A shallow or a strategic learning approach was indicative of students at risk of failing.
- Openness, indicating a creative, inquisitive temperament, was indicative of a passing grade.
- A kinaesthetic modality (preference for learn by doing) was indicative of students at risk of failing.

It was found that self-regulation factors were not significant once learning goals and approaches to learning were considered. Similarly, conscientiousness did not improve model accuracy over and above other factors of learning. Respective correlations between study factors and GPA, and significance in regression models predicting GPA, were not indicative of factors significant in classification models of students at risk of failing. For example, both openness and kinaesthetic learner modality were significant in a number of classification models but had relatively weak correlations with GPA. Conclusions from this study that openness and kinaesthetic learner modality were significant predictors of at-risk students were not widely observed in other studies, particularly the significance of learner modality. Therefore, further work is needed to determine if their importance in models of learning, as reported in this study, generalises to other student cohorts.

The primary value of non-cognitive factors of learning in this study was to distinguish the learning profile of students at risk of failing from the learning profile of students that passed, rather than provide improvement in model predictive accuracy. It has been argued that non-cognitive factors of motivation, self-regulation and approaches to learning are malleable, and key to an effective learning disposition, which in turn should be a valued learning outcome of courses in tertiary education. Further work is needed to evaluate subsequent benefits of profiling non-cognitive factors of learning during student induction, both for the student, and for first year mentoring and support programmes.

References

Achen, C. (1982). *Interpreting and Using Regression.* Number 07-029 in Quantitative Applications in the Social Sciences. Sage Publications, Inc. 49

Ackerman, P. L. and E. D. Heggestad (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin 121*(2), 219–245. 15

ACT (2012). 2012 retention completion summary tables. Technical report, www.act.org. 1

Akbani, R., S. Kwek, and N. Japkowicz (2004, September 20-24). Applying support vector machines to imbalnced dataset. In *15th European Conference on Machine Learning (ECML)*, pp. 39–50. Springer Berlin Heidelberg. 59

Ali, L., H. Marek, P. H. Winne, and D. Gašević (2014). Leveraging mslq data for predicting student achievement goal orientation. *Journal of Learning Analytics 1*(3), 157–160. 2

Allick, J. and A. Realo (1997). Intelligence, academic abilities, and personality. *Personality and Individual Differences 23 (5)*, 809–814. 1, 16, 95

Apter, M. J. (1989). *Reversal Theory: motivation, emotion and personality.* Taylor & Frances/Routledge. 21, 92

Baepler, P. and C. J. Murdoch (2010). Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching and Learning 4*(2), 1–9. 29

Baker, R. S. J. D., Z. A. Pardos, S. M. Gowda, B. B. Nooraei, and N. T. Heffernan (2011, July). Ensembling predictions of student knowledge within intelligent tutoring systems. In *Proceedings of 19th International Conference on User Modeling, Adaptation, and Personalization*, Girona, Spain, pp. 13–24. Springer-Verlag Berlin. 6, 31, 32, 33

Baker, R. S. J. D. and P. I. Salvador (2014). *Learning analytics*, Chapter Educational data mining and learning analytics, pp. 61–75. Springer. 5

Baker, R. S. J. D. and K. Yacef (2010). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining 1*(1), 3–17. 29, 32

Banfield, R. E., L. O. Hall, K. W. Bowyer, and D. Bhadoria (2004). A comparison of ensemble creation techniques. *Multiple Classifier Systems, Lectures Notes in Computer Science 3077*, 223–232. 31

Bartels, M., M. J. H. Rietveld, G. C. Van Baal, and D. I. Boomsma (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research 5*, 544–553. 12

Baumann, K. (2003). Cross-validation as the objective function for variable-selection techniques. *Trends in Analytical Chemistry 22*(6), 395–406. 58, 80, 92

Bayer, J., H. Bydzovská, J. Géryk, T. Obsívac, and L. Popelínsky (2012, June). Predicting drop-out from social behaviour of students. In *Proceedings of the Fifth International Conference on Educational Data Mining*, Chania, Greece, pp. 103–109. 2

Bekele, R. and W. Menzel (2005). A bayesian approach to predict performance of a student (BAPPS): A case with Ethiopian students. In *Proceedings of the International Conference on Artificial Intelligence and Applications*, Vienna, Austria. 31, 32

Bembenutty, H. (2009). Academic delay of gratification, self-regulation of learning, gender differences, and expectancy-value. *Personality and Individual Differences 26*, 347–352. 25

Bergin, S. (2006). *Statistical and machine learning models to predict programming performance.* Ph. D. thesis, Computer Science, National University of Ireland, Maynooth. 6, 32, 33, 66, 91

Bergin, S. and R. Reilly (2005). The influence of motivation and comfort-level on learning to program. In P. Romero, J. Good, E. Acosta Chaparro, and S. Bryant (Eds.), *Proceedings of the 17th Workshop on Psychology of Programming, PPIG05*, pp. 293–304. 96

Bergin, S. and R. Reilly (2006). Predicting introductory programming performance: A multi-institutional multivariate study. *Computer Science Education 16*(4), 303–323. 33

Bidjerano, T. and D. Y. Dai (2007). The relationship between the big-five model of personality and self-regulated learning strategies. *Learning and Individual Differences 17*, 69 – 81. 24, 26, 27, 28, 72, 78

Biggs, J., D. Kember, and D. Leung (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Education Psychology 71*, 133–149. 2, 41, 43

Boekaerts, M. (2001). Bringing about change in the classroom: strengths and weaknesses of the self-regulated learning approach. In *EARLI Presidential Address*, Centre for the Study of Education and Instruction, P.O. Box 9555, Leiden 2300 RB, The Netherlands. 2

Boekearts, M. (1996). Self-regulated learning at the junction of cognition and motivation. *European Psychologist 1 (2)*, 100–112. 24

Brady-Amoon, P. and J. N. Fuertes (2011). Self-efficacy, self-rated abilities, adjustment and academic performance. *Journal of Counseling and Development 89*(4), 431–438. 13, 14, 19

Breiman, L. (1996). Bagging predictors. *Machine Learning 24*, 123–140. 31

Breiman, L. (2001). Statistical modelling: The two cultures. *Statistical Science 16(3)*, 199–231. 28, 29

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen (1984). *Classification and Regression Trees*. Chapman and Hall/CRC. 30

Breunig, M. M., H.-P. Kriegel, R. T. Ng, and J. Sander (2000). LOF: Identifying density-based local outliers. In *ACM SIGMOD 2000 International Conference on Management of Data*, Dallas, TX. 64

Brown, S. D., R. W. Lent, K. Telander, and S. Tramayne (2010). Social cognitive career theory, conscientiousness, and work performance: A meta-analytic path analysis. *Journal of Vocational Behaviour 79*(1), 81–90. 19, 22

Brown, S. D., S. Tramayne, D. Hoxha, K. Telander, X. Fan, and R. W. Lent (2008). Social cognitive predictors of college students' academic performance and persistence: A meta-analytic path analysis. *Journal of Vocational Behaviour 72*, 298–308. 19, 22

Bruinsma, M. (2004). Motivation, cognitive processing and achievement in higher education. *Learning and Instruction 14*, 549–568. 19, 22, 23

Buckingham Shum, S. and R. Deakin Crick (2012). Learning dispositions and transferable competencies. Pedagogy, modelling and learning analytics. In *2nd International Conference on Learning Analytics and Knowledge*, Vancouver, BC, Canada, pp. 92–101. 1, 2, 16, 20, 24, 33, 96

Burisch, M. (1997). Test length and validity revisited. *European Journal of Personality 11*(4), 303–315. 39

Busato, V. V., F. J. Prins, J. J. Elshout, and C. Hamaker (1999). The relation between learning styles, the big five personality traits and achievement motivation in higher education. *Personality and Individual Differences 26*, 129–140. 23

Carpenter, J. and J. Bithell (2000). Bootstrap confidence intervals - when, which, what? A practical guide for medical statisticians. *Statistics in Medicine 19*, 1141–1164. 47

Cassells, P. (2015, January). The role, value and scale of higher education in Ireland. Technical report, Discussion Paper for Stakeholder Consultation, Expert Group on Future Funding for Higher Education. 3

Cassidy, S. (2011). Exploring individual differences as determining factors in student academic achievement in higher education. *Studies in Higher Education 37*(7), 1–18. 2, 12, 14, 15, 18, 19, 24, 26, 27, 28, 70, 75, 94

Cattell, H. E. P. and A. D. Mead (2008). *The SAGE Handbook of Personality Theory and Assessment*, Volume 2 Personality Measurement and Testing, Chapter 7. The Sixteen Personality Factor Questionnaire (16PF), pp. 135–159. SAGE publications Ltd. 15, 16

Chamorro-Premuzic, T. and A. Furnham (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality 37*, 319–338. 17, 18, 22

Chamorro-Premuzic, T. and A. Furnham (2004). A possible model for understanding the personality – intelligence interface. *British Journal of Psychology 95*, 249–264. 2, 15, 16

Chamorro-Premuzic, T. and A. Furnham (2006). *Personality and intellectual competence* (1st ed.). Sussex, UK: Psychology Press. 17

Chamorro-Premuzic, T. and A. Furnham (2008). Personality, intelligence and approaches to learning as predictors of academic performance. *Personality and Individual Differences 44*, 1596–1603. 2, 14, 16, 17, 23, 24, 26, 27, 28, 78, 95

Chatfield, C. (1983, June). *Statistics for technology. A course in applied statistics.* Chapman and Hall/CRC. 47, 87

Chatti, M. A., A. L. Dychhoff, U. Schroeder, and H. Thüs (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning. Special Issue on State of the Art in TEL 4*(5/6), 318–331. 1

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kelelmeyer (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research 16*, 321–357. 67

Clancy, P. and J. Wall (2000). Review of social backgrounds of higher education entrants. Technical report, Higher Education Authority, Dublin. 3

Colby, J. (2004, 31 August–2 September). Attendance and attainment - a comparative study. *Innovation in Teaching and Learning in Information and Computer Sciences 4*(2). 3

Colom, R. and C. Flores-Mendoza (2007). Intelligence predicts scholastic achievement irrespective of SES factors: Evidence from Brazil. *Intelligence 35*, 243–251. 12

Conati, C., A. Gertner, and K. Vanlehn (2002). Using bayesian networks to manage uncertainty in student modeling. *Use Modeling and User-Adapted Interaction 12*, 371–417. 32

Conrad, M. A. (2006). Aptitude is not enough: How personality and behavior predict academic performance. *Journal of Research in Personality 40*, 339–346. 14, 17, 70

Cooper, A. J., L. D. Smillie, and P. J. Corr (2010). A confirmatory factor analysis of the mini-ipip five-factor model personality scale. *Personality and Individual Differences 48*(5), 688–691. 43

Cortes, C. and V. Vapnik (1995). Support-vector networks. *Machine Learning 20*, 273–297. 31

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *American Psychological Association 78*(1), 98–104. 43

Costa, P. T. J. and R. R. McCrae (1992). *Revised NEO personality inventory (NEO-PI-R) and NEO Five-Factor inventory (NEO-FFI): Professional Manual.* Odessa Psychological Assessment Resources. 15

Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology 51*, 171–200. 20, 24

Cronbach, L. (1951, Sept). Coefficient alpha and the internal structure of tests. *Psychometrika 16*(3), 297–334. 43

Crosling, G. and M. Heagney (2009). Improving student retention in higher education. *Australian University Review 51*(2), 9–18. 4

Cury, F., A. Elliot, P. Sarrazin, D. Da Finseca, and M. Rufo (2002). The trichotomous achievement goal model and intrinsic motivation: A sequential mediational analysis. *Journal of Experimental Social Psychology 38*, 473–481. 21

De Clercq, M., B. Galand, and M. Frenay (2013). Chicken or the egg: Longitudinal analysis of the causal dilemma between goal orientation, self-regulation and cognitive processing strategies in higher education. *Studies in Educational Evaluation 39*, 4–13. 25, 96

De Feyter, T., R. Caers, C. Vigna, and D. Berings (2012). Unraveling the impact of the big five personality traits on academic performance: The moderating and mediating effects of self-efficacy and academic motivation. *Learning and Individual Differences 22*, 439–448. 2, 23 de Raad, B. and H. C. Schouwenburg (1996). Personality in learning and education: A review. *European Journal of Personality 10*, 303–336. 15, 16, 95

Deakin Crick, R. and C. Goldspink (2014). Learning dispositions, self-theories and student engagement. *British Journal of Educational Studies 62*(1), 19–35. 96

Deci, E. L. and R. M. Ryan (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry 11(4)*, 227–268. 21

Dekker, G., M. Pechenizkiy, and J. Vleeshouwers (2009). Predicting students drop out: A case study. In T. Barnes, M. C. Desmarais, C. Romero, and S. Ventura (Eds.), *Proceedings of the 2nd International Conference on Educational Data Mining*, Cordoba, Spain, pp. 41–50. 2, 4, 28, 33, 70

DiBenedetto, M. K. and H. Bembenutty (2013). Within the pipeline, self-regulated learning, self-efficacy, and socialization among college students in science courses. *Learning and Individual Differences 23*, 218–224. 19

Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Journal of Neural Computation 10*(7), 1895–1923. 60

Diseth, Á. (2011). Self-efficacy, goal orientations and learning strategies as mediators between preceding and subsequent academic achievement. *Learning and Individual Differences 21*, 191–195. 19, 20, 21, 22, 23, 70, 72, 96

Dixon, S. and R. Brereton (2009). Comparison of the performance of five common classifiers represented as boundary methods: Euclidean distance to centroids, linear discriminant analysis, quadratic discriminant analysis, learning vector quantization and support vector machines, as dependent on data structure. *Chemometrics and Intelligent Laboratory Systems 95*, 1–17. 31

Dollinger, S. J., A. M. Matyja, and J. L. Huber (2008). Which factors best account for academic success: Those which college students can control or those they cannot? *Journal of Research in Personality 42*, 872–885. 17, 21, 26, 27, 28, 78, 94

Domingos, P. and M. Pazzani (1997). On the optimality of the simple bayesian classifier under zero-one loss. *Machine Learning 29*, 103–130. 53

Drachsler, H. and W. Greller (2012, 29 April- 2 May). The pulse of learning analytics understandings and expectations from the stakeholders. In *2nd International Conference on Learning Analytics and Knowledge*, Vancouver, BC, Canada, pp. 120–129. ACM. 1

Duff, A., E. Boyle, K. Dunleavy, and J. Ferguson (2004). The relationship between personality, approach to learning and academic performance. *Personality and Individual Differences 36*, 1907–1920. 14, 17, 23, 24, 27, 28, 70, 72

Duffin, D. and G. Gray (2009a, March). Accommodating learner diversity in the classroom. In *3rd International Technology, Education and Development Conference (INTED)*, Valencia, Spain, pp. 4629–4635. 97

Duffin, D. and G. Gray (2009b, August 31 - September 2). Using ICT to enable inclusive teaching practices in higher education. *Assisstive Technology Research Series 25*, 640–645. 98

Dweck, C. S. (October 1986). Motivational processes affecting learning. *American Psychologist 41 (10)*, 1040–1048. 20

Dweck, C. S. and E. L. Leggett (1988). A social-cognitive approach to motivation and personality. *Psychological Review 95 (2)*, 256–273. 20, 21

Eccles, J. S. and A. Wigfield (2002). Motivation beliefs, values and goals. *Annual Review of Psychology 53*, 109–132. 18, 20

Entwhistle, N. (2005). Contrasting perspectives in learning. In F. Marton, D. Hounsell, and N. Entwhistle (Eds.), *The Experience of Learning* (3rd ed.)., pp. 3–22. Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment. 2, 21, 23

Entwistle, N. and D. Entwistle (1970). The relationship between personality, study methods and academic performance. *British Journal of Educational Psychology 40*, 132–143. 17

Eppler, M. A. and B. L. Harju (1997). Achievement motivation goals in relation to academic performance in traditional and nontraditional college students. *Research in Higher Education 38 (5)*, 557–573. 13, 14, 20, 21, 26, 27, 72, 75

Eysenck, H. (1994). *Test Your IQ*. Thorsons. 12

Eysenck, M. W. and M. T. Keane (2005). *Cognitive Psychology: A student handbook* (5th ed.). Psychology Press. 93

Farsides, T. and R. Woodfield (2003). Individual differences and undergraduate academic success: the roles of personality, intelligence, and application. *Personality and Individual Differences 34*, 1225–1243. 1, 18

Fisher, R. A. (1915, May). Frequency distribution of the values of the correlation coefficient in samples from an indefinitel large population. *Biometrika 10*(4), 507–521. 43

Flanagan, D. P. and K. S. McGrew (1998). Interpreting intelligence tests from contemporary gf-gc theory: Joint confirmatory factor analysis of the wj - r and kait in a non - white sample. *Journal of School Psychology Vol. 36, No. 2*, 151 – 182. 12, 13, 14

Fleming, N. D. (1995). I'm different, not dumb. Modes of presentation (VARK) in the tertiary classroom. In *Research and Development in Higher Education, Proceedings of the 1995 Annual Conference of the Higher Education and Research Development Society of Australasia (HERDSA)*, Volume 18, pp. 308–313. 41

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *American Psychological Association 101:2*, 171–191. 12

Freedman, D. (1987). As others see us: A case study in path analysis. *Journal of Educational Statistics 12(2)*, 101–128. 28

Frey, M. C. and D. K. Detterman (2003). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science 15(6)*, 373–378. 13

Furnham, A. and S. Medhurst (1995). Personality correlates of academic seminar behaviour: A study of four instruments. *Personality and Individual Differences 19(2)*, 197–208. 17

Furnham, A., J. Zhang, and T. Chamorro-Premuzic (2006). The relationship between psychometric and self-estimated intelligence, creativity, personality and academic achievement. *Imagination, Cognition and Personality 25(2)*, 119–145. 14

Galesic, M. and M. Bosnjak (2009). Effect of questionniare lenght on participation and indicators of response qualirt in a web survey. *Public Opinion Quarterly on Topics in Survey Measurement and Public Opinion 73*(2), 349–360. 39

Ganyaupfu, E. M. (2013, September). Teaching methods and students' academic performance. *International Journal of Humanities and Social Science Invention 2*(9), 29–35. 92

Gilakjani, A. P. (2012, December). A match or mismatch between learning styles of the learners and teaching styles of the teachers. *International Journal of Modern Education and Computer Science 11*, 51–60. 97

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment 4 (1)*, 26–42. 15

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist 48 (1)*, 26–34. 15, 16

Goldberg, L. R., J. A. Johnson, H. W. Eber, R. Hogan, M. C. Ashton, C. R. Cloninger, and H. C. Gough (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality 40*, 84–96. 39

Goodman, S., T. Jaffer, M. Keresztesi, F. Mamdani, D. Mokgatle, M. Musariri, J. Pires, and A. Schlechter (2011). An investigation of the relationship between students' motivation and academic performance as mediated by effort. *Psychological Journal of South Africa 41(3)*, 373–385. 21, 22, 26

Gray, E. K. and D. Watson (2002). General and specific traits of personality and their relationship to sleep and academic performance. *Journal of Personality 70(2)*, 177–206. 17

Gray, G., C. McGuinness, and P. Owende (2013, July 6-9). An investigation of psychometric measures for modelling academic performance in tertiary education. In S. K. D'Mello, R. A. Calvo, and A. Olney (Eds.), *Sixth International Conference on Educational Data Mining*, Memphis, Tennessee, pp. 240–243. 3

Gray, G., C. McGuinness, and P. Owende (2014a, February). An application of classification models to predict learner progression in tertiary education. In *Advance Computing Conference (IACC), 2014 IEEE International*, Gurgaon, India, pp. 549–554. 3

Gray, G., C. McGuinness, and P. Owende (2014b, July). Non-cognitive factors of learning as predictors of academic performance in tertiary education. In S. Ritter and S. Fancsali (Eds.), *Workshop on Non-Cognitive Factors and Personalization for Adaptive Learning (NCFPAL) IN EDM 2014 Extended Proceedings*, London, pp. 107–114. CEUR Workshop Proceedings. 3

Gray, G., C. McGuinness, and P. Owende (2014c). A review of psychometric data analysis and applications in modelling of academic achievement in tertiary education. *Journal of Learning Analytics 1*(1), 75–106. 3

Gray, G., C. McGuinness, P. Owende, and M. Hofmann (2015). Learning factor models of students ar risk of failing in early stages of tertiary education. *Journal of Learning Analytics in review.* 3

Groth, R. (2000). *Data Mining Building Competitive Advantage*. Prentice Hall PTR. 30

Gupta, P. and D. Toshniwal (2011). Performance comparison of rule based classification algorithms. *International Journal of Computer Science and Informatics 1*(2), 37–42. 30

Hair, J. F. J., W. C. Black, B. J. Babin, and R. E. Anderson (2010). *Multivariate Data Analysis A Global Perspective, 2nd Ed.* Pearson Education, Inc. 48

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Association of Physics Teachers 66*, 64–74. 92

Hall, M. A. and G. Homes (2003, Nov-Dec). Benchmarking attributes selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data Engineering 15*(6), 1437–1447. 58

Han, J. and M. Kamber (2006). *Data Mining Concepts and Techniques*. Burlington, MA: Morgan Kaufmann. 29, 30, 50, 51, 52, 53, 63, 64, 92

HEA (2008). National plan for equity of access to higher education: 2008 - 2013. Technical report, Higher Education Authority. 3

HEA (2013, January). Higher education key facts and figures 2011-12. Technical report, Higher Education Authority. 3

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Education Research 58*, 47–77. 1, 16

Herzog, A. R. and J. G. Bachman (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly 45*, 549–559. 42

Herzog, S. (2006, Fall). Estimating student retention and degree-completion time: Decision trees and neural networks vis-à-vis regression. *New Directions For Institutional Research 131*, 17–33. 2, 28, 32, 33, 91

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics 6*(2), 65–70. 48

Hoskins, S. L., S. E. Newstead, and I. Dennis (1997). Degree performance as a function of age, gender, prior qualifications and discipline studied. *Assessment and Evaluation in Higher Education 22(3)*, 317–328. 75, 94

Jayaprakash, S. M., E. W. Moody, E. J. M. Lauria, J. R. Regan, and J. D. Baron (2014). Early alert of academically at-risk students: An opensources analytics initiative. *Journal of Learning Analytics 1*(1), 6–47. 2, 32, 33, 91, 98

Jensen, A. R. (1981). *Straight Talk about Mental Tests*. The Free Press (MacMillian). 12

John, O. P., L. P. Naumann, and C. J. Soto (2008). *Handbook of Personality: Theory and Research* (3rd ed.)., Chapter 4 Paradigm Shift to the Integrative Big Five Trait Taxonomy: History, Measurement and Conceptual Issues, pp. 114–158. Guildford Press, NY. 15

John, O. P. and S. Srivastava (1999, June). *Handbook of personality: Theory and research* (2nd ed.)., Chapter 4 The Big Five trait taxonomy: History, measurement, and theoretical perspectives., pp. 102–138. The Guilford Press. 15

Jonsson, A., J. Johns, H. Mehranian, I. Arroyo, B. Woolf, A. Barto, D. Fisher, and S. Mahadevan (2005). Evaluating the feasibility of learning student models from data. Technical report, American Association for Artificial Intelligence. 32

Kablan, Z. (2014). The effect of manipulatives on mathematics achievement across different learning styles. *Educational Psychology ahead-of-print*, 1–20. 97

Kang, Y. and J. R. Harring (2012). Reexamining the impact of non-normality in two-group comparison procedures. *Journal of Experimental Education 83*(2), 147–174. 42, 47, 75

Kappe, R. and H. van der Flier (2010). Using multiple and specific criteria to assess the predictive validity of the big five personality factors on academic performance. *Journal of Research in Personality 44*, 142–145. 16, 17, 95

Kaufman, J. C., M. D. Agars, and M. C. Lopez-Wagner (2008). The role of personality and motivation in predicting early college academic success in non-traditional students at a hispanic-serving institution. *Learning and Individual Differences 18*, 492 – 496. 14, 17, 22, 26, 27, 28, 70, 78, 95

Keerthi, S. S., K. B. Duan, S. K. Shevade, and A. N. Poo (2005). A fast dual algorithm for kernel logistic regression. *Machine Learning 61*, 151–165. 91

Kelly, K. and C. Marshall (2012, August). Prediction of engineering student progression from entrance data. In *29th International Manufacturing Conference*, Belfast. 63

Klassen, R. M., L. L. Krawchuk, and S. Rajani (2008). Academic procrastination of undergraduates: Low self-efficacy to self-regulate predicts higher levels of procrastination. *Contemporary Educational Psychology 33*, 915–931. 19

Knight, S., S. Buckingham Shum, and K. Littleton (2013, April). Epistemology, pedagogy, assessment and learning analytics. In *Third Conference on Learning Analytics and Knowledge (LAK 2013)*, Leuven, Belgium, pp. 75–84. 2, 3, 16, 24, 33

Kobrin, J., B. F. Patterson, E. J. Shaw, K. D. Mattern, and S. M. Barbuti (2008). Validity of the SAT for predicting first year college grade point average. Technical Report 2008-5, College Board Research Report. 14

Komarraju, M., S. J. Karau, and R. R. Schmeck (2009). Role of the big five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences 19*, 47–52. 21, 22, 23

Komarraju, M., S. J. Karau, R. R. Schmeck, and A. Avdic (2011). The big five personality traits, learning styles, and academic achievement. *Personality and Individual Differences 51*, 472–477. 17, 26, 27, 78

Komarraju, M. and D. Nadler (2013). Self-efficacy and academic achievement: Why do implicit beliefs, goals, and effort regulation matter? *Learning and Individual Differences 25*, 67–72. 2, 19, 22, 25, 26, 44, 70, 72

Kriegel, H.-P., P. Kröger, and A. Zimek (2009). Outlier detection techniques. In *Tutorial at the 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Bangkok. 64

Kruskal, W. H. and W. A. Wallis (1952). Use of ranks in one-criterion variance analysis. *Journal of the Americal Statistics Association 47*(260), 583–621. 48

Kubat, M. and S. Matwin (1997, July 8-12). Addressing the curse of imbalanced trainined sets: One-sided selection. In *International Conference on Machine Learning*, Nashville, TN, USA, pp. 179–186. 59

Kuncel, N. R., S. A. Hezlett, and D. S. Ones (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology 86 (1)*, 148–161. 12

Kundel, H. L. and M. Polansky (2003). Measurement of observer agreement. *Radiology 228*(2), 303–308. 66

Lance, C. E., M. M. Butts, and L. C. Michels (2006). The sources of four commonly reported cutoff criteria. *Organisational Research Methods 9*(2), 202–220. 43

Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining.* John Wiley and Sons, Inc. 29, 55, 56, 66

Lauria, E. J. M., J. D. Baron, and M. Devireddy (2012, 29 April-2 May). Mining academic data to improve college student retention: An open source perspective. In *2nd International Conference on Learning Analytics and Knowledge (LAK 2012)*, Vancouver, British Columbia, Canada. ACM. 6, 32, 33

Lauria, E. J. M., E. W. Moody, S. M. Jayaprakash, N. Jonnalagadda, and J. D. Baron (2013, April). Open academic analytics initiative: Initial research findings. In *Third Conference on Learning Analytics and Knowledge (LAK 2013)*, Leuven, Belgium. ACM. 3, 32, 33, 91

Lent, R. W., S. D. Brown, and G. Hackett (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behaviour 45*, 79–122. 1

López, M. I., J. M. Luna, C. Romero, and S. Ventura (2012, June). Classification via clustering for predicting final marks based on student participation in forums. In *Proceedings of the Fifth International Conference on Educational Data Mining*, Chania, Greece, pp. 148–151. 2

Marton, F. and R. Säljö (2005). Approaches to learning. In F. Marton, D. Hounsell, and N. Entwhistle (Eds.), *The Experience of Learning: Implications for teaching and studying in higher education, 3rd (Internet) Edition*, pp. 36–58. Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment. 23, 24

Matarazzo, J. D. and S. G. Goldstein (1972). The intellectual caliber of medical students. *Journal of Medical Education 47:2*, 10. 12

Mayo, M. and A. Mitrovic (2001). Optimising its behaviour with bayesian networks and decision theory. *International Journal of Artificial Intelligence in Education 12*, 124–153. 32

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence 37*, 1–10. 12

Merceron, A. and K. Yacef (2005, July). Educational data mining: A case study. In *Proceedings of the 12th International Conference on Artificial Intelligence in Education (AIED)*, Amsterdam, pp. 467–474. 32, 33

Micceri, T. (1989). The unicorn, the normal curve, and other improbably creatures. *Psychological Bulletin 105*(1), 156–166. 42, 47

Miller-Reilly, B. (2006). *Affective change in adult students in second chance mathematics courses: Three different teaching approaches.* Ph. D. thesis, University of Auckland. 98

Milne, J., L. M. Jeffrey, G. Suddaby, and A. Higgins (2012, 25-28 November). Early identification of students at risk of failing. In *ASCILITE-Australian Society for Computers in Learning in Tertiary Education Annual Conference*, Volume 1, Wellington, New Zealand. 3

Minaei-Bidgoli, B., D. A. Kashy, G. Kortemeyer, and W. F. Punch (2003, November). Predicting student performance: An application of data mining methods with educational web-based system lon-capa. In *Proceedings of the 33rd ASEE/IEEE Frontiers in Education Conference*, pp. 13–18. IEEE. 32, 33, 65

Mirriahi, N., D. Gašević, P. Long, and S. Dawson (2014). Scientometrics as an important tool for the growth of the field of learning analytics. *Journal of Learning Analytics 1*(2), 1–4. 1

Mislevy, R. J., J. T. Behrens, and K. E. Dicerbo (2012). Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *Journal of Educational Data Mining 4 (1)*, 11–48. 1

Mitchell, T, M. (2015). *Machine Learning* (2nd ed.)., Chapter 3, pp. 55–85. McGraw Hill. 66

Mooney, O., V. Patterson, M. O'Connor, and A. Chantler (2010, October). A study of progression in higher education: A report by the higher education authority. Technical report, Higher Education Authority, Ireland. 1, 3, 4, 35, 38

Moran, M. A. and M. J. Crowley (1979). The leaving certificate and first year university performance. *Journal of Statistical and Social Enquiry in Ireland XXIV, part 1*, 231–266. 1

Munzert, A. W. (1980). *Test Your IQ.* MacMillian, USA. 12

Naderi, H., H. T. Abdullah, J. Sharir, and V. Kumar (2009). Creativity, age and gender as predictors of academic achievement among undergraduate students. *Journal of American Science 5(5)*, 101–112. 2, 94

Nandagopal, K. and K. Anders Ericsson (2012). An expert performance approach to the study of individual differences in self-regulated learning activities in upper-level college students. *Learning and Individual Differences 22*(5), 597–609. 2

Nasiriyan, A., H. K. Azar, A. Noruzy, and M. R. Dalvand (2011). A model of self-efficacy, task value, achievement goals, effort and mathematics achievement. *International Journal of Academic Research 3(2)*, 612–618. 23

Nelson, B., R. Nugent, and A. A. Rupp (2012). On instructional utility, statistical methodology, and the added value of ECD: Lessons learned from the special issue. *Journal of Educational Data Mining 4 (1)*, 227–233. 33

Ng, A. Y. and M. I. Jordon (2001). On discriminative vs generative classifiers: A comparison of logistic regression and naïve bayes. *Advances in Neural Information Processing Systems (NIPS) 14*, 841–848. 30, 31, 66

Nicol, D. J. and D. MacfarlaneDick (2006). Formative assessment and selfregulated learning: A model and seven principles of good feedback practice. *Studies in higher education 31*(2), 199–218. 25

Ning, H. K. and K. Downing (2010). The reciprocal relationship between motivation and self-regulation: A longitudinal study on academic performance. *Learning and Individual Differences 20*, 682–686. 14, 15, 23, 25, 26

Nisbet, R., J. Elder, and G. Miner (2009). *Statistical Analysis and Data Mining Applications.* Academic Press. 28, 29, 31

O'Connor, M. C. and S. V. Paunonen (2007). Big five personality predictors of post-secondary academic performance. *Personality and Individual Differences 43*, 971–990. 15, 17

OECD (2013). Education at a glance 2013: OECD indicators. Technical report, OECD Publishing. 1

Otero, J. and L. Sánchez (2005). Induction of descriptive fuzzy classifiers with the logitboost algorithm. *Soft Computing 10*, 825–835. 30

Pardos, Z. A., R. S. J. D. Baker, S. M. Gowda, and N. T. Heffernan (2011). The sum is greater than the parts: Ensembling models of student knowledge in educational software. *ACM SIGKDD explorations newsletter 13*(2), 37–44. 2, 32, 33

Pascarella, E. T. and P. T. Terenzini (2005). *How College Affects Students*, Volume 2. John Wiley and Sons, Inc. 4

Patterson, V., D. Carroll, and V. Harvey (2014). Key facts and figures, higher education 2012/13. Technical report, Higher Education Authority. 1, 3

Pelánek, R. (2015). Metrics for evaluation of student models. *Journal of Educational Data Mining 7*(2), 1–19. 49

Pérez-Martínez, J. E., M. J. García-García, W. H. Perdemo, and M. J. Villamide-Díaz (2009, July). Analysis of the results of the continuous assessment in the adaptation of the universidad politécnica de Madrid to the european higher education area. In R. Hadgraft and L. Mann (Eds.), *Proceedings of the Research in Engineering Education Symposium*, Palm Cove, Queensland, Australia. 90

Pintrich, P. and E. DeGroot (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal and Educational Psychology 82*, 33–40. 18, 19, 25

Pintrich, P., D. Smith, T. Garcia, and W. McKeachie (1991). A manual for the use of the motivated strategies for learning questionnaire. Technical Report 91-B-004, The Regents of the University of Michigan. 22, 41, 43

Poropat, A. E. (2009). A meta-analysis of the five-factor model or personality and academic performance. *Psychological Bulletin 135*(2), 322–338. 28

Powell, J. L. (1973). *Selection For University In Scotland.* The Scottish Council for Research in Education. 1

Prinsloo, P., S. Slade, and F. Galpin (2012, 29 April- 2 May). Learning analytics - challenges, paradoxes and opportunities for mega open distance learning institutions. In *2nd International Conference on Learning Analytics and Knowledge*, Vancouver, British Columbia, Canada, pp. 130–133. ACM. 28

Provost, F., D. Jensen, and T. Oates (1999). Efficient progressive sampling. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, San Diego, CA, pp. 23–31. ACM. 66

Quinlan, J. R. (1986). Simplifying decision trees. AI Memo 930, Massachusetts Institute of Technology Artificial Intelligence Laboratory. 29, 30

Raileanu, L. E. and K. Stoffel (2004). Theoretical comparison between the gini index and information gain criteria. *Annals of Mathematics and Artificial Intelligence 41*, 77–93. 51

Rice, J. A. (1995). *Mathematical Statistics and Data Analysis* (2nd ed.). Duxbury Press, Belmont, CA. 48, 60

Robbins, S. B., K. Lauver, H. Le, D. Davis, and R. Langley (2004). Do psychosocial and study skill factors predict college outcomes? A meta analysis. *Psychological Bulletin 130 (2)*, 261–288. 2, 13, 14, 18, 19, 21, 22, 78, 96

Romero, C., J. L. Olmo, and S. Ventura (2013, July 6-9). A meta-learning approach for recommneding a subset of white-box classification algorithms for moodle dataset. In S. K. D'Mello, R. A. Calvo, and A. Olney (Eds.), *Sixth International Conference on Educational Data Mining*, Memphis, Tennessee, pp. 268–271. 6

Romero, C. and S. Ventura (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications 33*, 135–146. 29

Romero, C. and S. Ventura (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. 29

Romero, C., S. Ventura, P. G. Espejo, and C. Hervás (2008). Data mining algorithms to classify students. In *Proceedings of the 1st International Conference on Educational Data Mining*, Montreal, Quebec, Canada, pp. 8–17. 32, 33, 59, 65

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney u test. *Behavioral Ecology 17*(4), 688–690. 49

Ryan, R. M. and E. L. Deci (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology 25*, 54–67. 18, 21

Sachin, B. R. and S. M. Vijay (2012, Jan). A survey and future vision of data mining in educational field. In *Second International Conference on Advanced Computing Communication Technologies (ACCT)*, pp. 96–100. 1

Saeys, Y., I. Inaki, and P. Larranage (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics 23*(19), 2507–2517. 80

Sargent, D. J. (2001). Comparison of artificial neural networks with other statistical approaches. *Conference on Prognostic Factors and Staging Cancer Management 91*(8), 1636–1642. 30

Schmitt, N., F. L. Oswald, T. Pleskac, R. Sinha, and M. Zorzie (2009, November). Prediction of four-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology 94*(6), 1479–97. 15

Schunk, D. H. (2005). Commentary on self-regulation in school contexts. *Learning and Instruction 15*, 173–177. 24

Shevade, S. K. and S. S. Keerthi (2003). A simple and efficient algorithm for gene selection using sparse logistic regression. *Bioinformatics 19*(17), 2246–2253. 53

Shute, V. and M. Ventura (2013). *Stealth Assessment. Measuring and Supporting Learning in Video Games*. The John D. and Catherine T. MacArthur Foundation Reports on Digital Media and Learning. MIT Press. 2, 16, 33

Siemens, G. (2012). Learning analytics: Envisioning a research discipline and a domain of practice. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, Vancouver, BC, Canada, pp. 4–8. 1

Siemens, G. and R. S. J. D. Baker (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, Vancouver, BC, Canada, pp. 252–254. 1

Sins, P. H. M., W. R. van Joolingen, E. R. Savelsbergh, and B. van Hout-Wolters (2008). Motivation and performance within a collaborative computer-based modeling task: Relations between students' achievement goal orientation, self-efficacy, cognitive processing, and achievement. *Contemporary Educational Psychology 33*, 58–77. 22, 23

Smith, M., L. Therry, and J. Whale (2012). Developing a model fo identifying students as risk of failure in a first year accounting unit. *Higher Education Studies 2*(4), 91–102. 4

Smith, Z. R. and C. S. Wells (2006, October 18-20). Central limit theorem and sample size. In *Annual Meeting of the Northeastern Educational Research Association*, New York. 47

Snelgrove, S. (2004). Approaches to learning of student nurses. *Nurse Education Today 24*, 605–614. 24

Spearman, C. (1904). General intelligence, objectively determined and measured. *Journal of Psychology 15*, 201–293. 12

Spearman, C. (1927). *The abilities of man*. MacMillian. 12

Spector, P. E. (1992). *Summated Rating Scale Construction. An Introduction.* Sage University Papers Series. Quantitice Applications in the Social Sciences, 07-082. Sage Publications, Inc. 43

Steel, P. and C. J. Konig (2006). Integrating theories of motivation. *Academy of Management Review 31 (4)*, 889–913. 18

Sternberg, R. (1999). Intelligence as developing expertise. *Contemporary Educational Psychology 24*, 359 – 375. 12

Sundre, D. L. and A. Kitsantas (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology 29*, 6–26. 26

Superby, J. F., J. P. Vandamme, and N. Meskens (2006). Determination of factors influencing the achievement of the first-year university students using data mining methods. In *Proceedings of the Workshop on Educational Data Mining at the 8th International Conference on Intelligence Tutoring Systems*, Taiwan, pp. 1–8. 66

Swanberg, A. B. and Ø. L. Martinsen (2010). Personality, approaches to learning and achievement. *Educational Psychology 30(1)*, 75–88. 2, 16, 24, 26, 27, 28, 72, 78

Tan, P.-N., M. Steinbach, and V. Kumar (2014). *Introduction to Data Mining* (International Edition ed.). Boston, MA: Pearson Education, Inc. 30, 31, 53, 55, 56, 63, 81, 91

Tang, M., X. Chen, W. Hu, and W. Yu (2012). Generation of a probabilistic fuzzy rule base by learning from examples. *Information Sciences 217*, 21–30. 30

Tavakol, M. and R. Dennick (2011). Making sense of Cronbachs alpha. *International Journal of Medical Education 2*, 53–55. 43

Tempelaar, D. T., H. Cuypers, E. van de Vrie, A. Heck, and H. van der Kooij (2013). Formative assessment and learning analytics. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13)*, New York, NY, USA, pp. 205–209. ACM. 1, 2, 32

Thai-Nghe, N., P. Janecek, and P. Haddawy (2007, October 10-13, Milwaukee, WI). A comparative analysis of techniques for predicting academic performance. In *37th ASEE/IEEE Frontiers in Education Conference*, pp. T2G7–T2G12. 65

Tinto, V. (2006). Research and practice of student retention: What next? *Journal of College Student Retention 8*(1), 1–19. 92

Tinto, V. (2012). *Completing College: Rethinking Institutional Action.* University of Chicago Press. 4

Tishman, S., E. Jay, and D. N. Perkins (1993). Teaching thinking disposition: From transmission to enculturation. *Theory into Practice 32*, 147–153. 2, 16

Trapmann, S., B. Hell, J.-O. W. Hirn, and H. Schuler (2007). Meta-analysis of the relationship between the big five and academic success at university. *Zeitschrift fur Psychologie 215*(2), 132–151. 16, 17

Tukey, J. W. (1949). Comparing individual means int he analysis of variance. *Biometrics 5*(2), 99–114. 48

Tupes, E. C. and R. E. Cristal (1961). Recurrent personality factors based on trait ratings. Personnel laboratory, aeronautical systems division, airforce systems command, United States Air Force, Lackland Air Force Base, Texas. 15

Vancouver, J. B. and L. N. Kendall (2006). When self-efficacy negatively relates to motivation and performance in a learning context. *Journal of Applied Psychology 91*(5), 1146–53. 28

Volet, S. E. (1996). Cognitive and affective variables in academic learning: The significance of direction and effort in students' goals. *Learning and Instruction 7(3)*, 235–254. 24, 25, 97

Weiss, G. M. (2004). Mining with rarity: A unifying framework. *SIGKDD Explorations 6*(1), 7–19. 67

Wigfield, A., J. S. Eccles, and P. Pintrich (1996). *Handbook of Educational Psychology*, Chapter 7. Development between the ages of 11 and 25, pp. 148–185. Simon & Schuster Macmillan. 94

Winters, F. I., J. A. Greene, and C. M. Costich (2008, July). Self-regulated learning within computer-based learning environments: A critical analysis. *Educational Psychology Review 20*, 429–444. 93, 98

Witten, I. H. and E. Frank (2005). *Data Mining Practical Machine Learning Tools and Techniques* (2nd ed.). Data Management Systems. Morgan Kaufmann. 56

Wolters, C. A. (1998). Self-regulated learning and college students' regulation of motivation. *American Psychological Association 90(2)*, 224–235. 21, 22

Wright, P. S. (1992). Adjusted p-values for simultaneous inference. *Biometrics 48*, 1005–1013. 48

Yang, J. and V. Honavar (1998, April/May). Feature subset selection using a genetic algorithm. *IEEE Intellegent Systems 13*(2), 44–49. 59

Yu, C. H. (2001, April). An introduction to computing and intrepreting cronbach coefficient alpha in sas. In *Proceedings of the Twenty-Six Annual SAS Users Group International Conference*, Long Beach, CA. 43

Yu, H.-F., H.-Y. Lo, H.-P. Hsieh, J.-K. Lou, T. G. McKenzie, J.-W. Chou, P.-H. Chung, C.-H. Ho, C.-F. Chang, Y.-H. Wei, J.-Y. Weng, E.-S. Yan, C.-W. Chang, T.-T. Kuo, Y.-C. Lo, P. T. Chang, C. Po, C.-Y. Wang, Y.-H. Huang, C.-W. Hung, Y.-X. Ruan, Y.-S. Lin, S.-d. Lin, H.-T. Lin, and C.-J. Lin (2010). Feature engineering and classifier ensemble for KDD cup 2010. In *JMLR Workshop and Conference Proceedings*, Volume 1, pp. 1–16. 32

Yusuf, M. (2011). The impact of self-efficacy, achievement motivation, and self-regulated learning strategies on students' academic achievement. *Procedia Social and Behavioral Sciences 15*, 2623–2626. 19

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist 25(1)*, 3–17. 24, 25, 96

Zimmerman, B. J. and A. Kitsantas (2005). Homework practices and academic achievement: The mediating role of self-efficacy and perceived responsibility beliefs. *Contemporary Educational Psychology 30*, 397–417. 25

Zuffianó, A., G. Alessandri, M. Gerbino, B. P. L. Kanacri, L. Di Giunta, M. Milioni, and G. V. Caprara (2013). Academic achievement: The unique contribution of self-efficacy beliefs in self-regulated learning beyond intelligence, personality traits and self-esteem. *Learning and Individual Differences 23*, 158–162. 23

Appendix A

Abstracts of Publications Emanating from this Study

Section 1.7 listed the publications emanating from this study, specifically: an invited book chapter (*Pb1*); two journal publications (*Pb2*, *Pb3*) and four conference papers (*Pb4* - *Pb7*). This appendix contains the abstract from each of these publications.

***Pb1*. Gray, G., McGuinness, C. and Owende, P. (2015) Non-cognitive factors of learning as early indicators of students at-risk of failing in tertiary education In Khine, M. S. (Eds.) *Non-cognitive Factors and Educational Attainment*, Sense Publishers.**

Following from an invitation to contribute a chapter in the forthcoming book *Non-cognitive factors and Educational Attainment*, the following abstract has been accepted. The book is due for publication in late 2015 by an international publisher. Editor: Dr. Myint Swe Khine, Adjunct Professor, Science and Mathematics Education Centre, Curtin University, Perth, Australia.

Abstract: Non-cognitive factors of learning have been associated with an effective learning disposition, describing attributes and behaviour that are characteristic of a good learner. This chapter explores fifteen non-cognitive factors of learning relating to personality, motivation, self-regulation and learning strategies. These were chosen on the basis of being directly or indirectly related to academic performance in tertiary education. Data on these non-cognitive factors of learning was gathered using an online learner profiler compiled from validated instruments in the public domain, and administered during first-year student induction. Analysis was conducted on three years of data (n=1,207) covering a diverse student population in terms of age, prior academic performance and course of study.
The efficacy of learner profiling during first-year student induction to tertiary education is reviewed, with a focus on identifying early indicators of students at risk of failing in the first year of study. Relationships between each of the non-cognitive factors of learning and academic performance are presented, along with variations found when analysing subgroups by course of study, age and gender. Significant differences between the learning profile of students that failed and the learning profile of students that passed are considered.

***Pb2*. Gray, G., McGuinness, C., Owende, P., and Hofmann, M. (2016) Learning Factor Models of Students at Risk of Failing in the Early Stage of Tertiary Education, *Journal of Learning Analytics*, 3(2): 330-372, 2016**

This manuscript was submitted to the journal of learning analytics in March 2015 in response to a call for papers for a special section on multimodal learning analytics.

Abstract: This paper reports on a study to predict students at risk of failing based on data available prior to commencement of first year of study. The study was conducted over three years, 2010 through 2012, on a student population from a range of academic disciplines (n=1,207). Data was gathered from both student enrolment data maintained by college administration, and an online, self-reporting, learner profiling tool administered during induction sessions for students enrolling into the first year of study. Factors considered included prior academic performance, personality, motivation, self-regulation, learning approaches, learner modality, age and gender. Models were trained on data from the 2010 and 2011 student cohort, and tested on data from the 2012 student cohort. A comparison of eight classification algorithms found k-Nearest Neighbour achieved best model accuracy when applied to a different student cohort (72%), but accuracies achieved by other algorithms were similar, including a voting ensemble (71%) and SVM (70%). Accuracies estimated using cross validation were higher. Correlations between study factors and GPA were not indicative of factors significant in classification models of students at risk of failing. Results indicated that early modelling of first year students yielded informative, generalisable models that identified students at risk of failing.

***Pb3*: Gray, G., McGuinness, C., Owende, P., and Carthy, A. (2014) A review of psychometric data analysis and applications in modelling of academic achievement in tertiary education. *Journal of Learning Analytics*, 1(1):75-106, 2014.**

This manuscript was accepted for publication in the inaugural issue of the Journal of Learning Analytics.

Abstract: Increasing college participation rates, and diversity in student population, is posing a challenge to colleges in their attempts to facilitate learners achieve their full academic potential. Learning analytics is an evolving discipline with capability for educational data analysis that could enable better understanding of learning process, and therefore mitigate these challenges. The outcome from such data analysis will be dependent on the range, type and quality of available data, and the type of analysis performed. This study reviewed factors that could be used to predict academic performance, but which are currently not systematically measured in tertiary education. It focused on psychometric factors of ability, personality, motivation and learning strategies. Their respective relationships with academic performance are enumerated and discussed. A case is made for their increased use in learning analytics to enhance the performance of existing student models. It is noted that lack of independence, linear additivity and constant variance in the relationships between psychometric factors and academic performance suggests increasing relevance of data mining techniques, which could be used to provide useful insights on the role of such factors in the modelling of learning process.

***Pb4*. Gray, G., McGuinness, C., and Owende, P. (2014) An application of classification models to predict learner progression in tertiary education. *4^{th} IEEE International Advanced Computing Conference*, Gurgaon, India, pp 549-554, February 2014.**

This paper was presented online at that 4^{th} IEEE International Advanced Computing Conference, India.

Abstract: This paper reports on an application of classification models to identify college students at risk of failing in the first year of study. Data was gathered from three student cohorts in the academic years 2010 through 2012. Students within the cohorts were sampled from a range of academic disciplines (n=1,074), and were diverse in their academic backgrounds and abilities. Metrics used included data that are typically available to colleges such as age, gender and prior academic performance. The study also considered psychometric indicators that can be assessed in the early stages after enrolment, specifically, personality, motivation and learning strategies. Six classification algorithms were considered. Model accuracy was assessed using cross validation and was compared to outcomes when models were applied to a subsequent academic year. It was found that mature students were more complex to model than younger students. Furthermore, 10-fold cross validation accurately estimated model performance when modelling younger students only, but over-estimated model accuracy when modelling mature students.

***Pb5*. Gray, G., McGuinness, C., and Owende, P. (2014) Non-cognitive factors of learning as predictors of academic performance in tertiary education. In Gutierrez-Santos, S and Santos O. C, editors, *WSEDM 2014 co-located with the 7th International Conference on Educational Data Mining (EDM 2014)*, London, July 4-7, 2014.**

This paper was presented at the *Non-Cognitive Factors & Personalization for Adaptive Learning* workshop at the 7^{th} International Conference on Educational Data Mining.

Abstract: This paper reports on an application of classification and regression models to identify college students at risk of failing in first year of study. Data was gathered from three student cohorts in the academic years 2010 through 2012 (n=1,207). Students were sampled from fourteen academic courses in five disciplines, and were diverse in their academic backgrounds and abilities. Metrics used included non-cognitive psychometric indicators that can be assessed in the early stages after enrolment, specifically factors of personality, motivation, self regulation and approaches to learning. Models were trained on students from the 2010 and 2011 cohorts, and tested on students from the 2012 cohort. Is was found that classification models identifying students at risk of failing had good predictive accuracy ($>$ 79%) on courses that had a significant proportion of high risk students (over 30%).

***Pb6*. G, Gray, C. McGuinness, and P. Owende (2013) An Investigation of Psychometric Measures for Modeling Academic Performance in Tertiary Education. *6th International Conference on Educational Data Mining (EDM 2013)*, Memphis, July, 2013.**

This paper was presented at the 6th International Conference on Educational Data Mining.

Abstract: Increasing college participation rates, and a more diverse student population, is posing a challenge for colleges in facilitating all learners achieve their potential. This paper reports on a study to investigate the usefulness of data mining techniques in the analysis of factors deemed to be significant to academic performance in first year of college. Measures used include data typically available to colleges at the start of first year such as age, gender and prior academic performance. The study also explores the usefulness of additional psychometric measures that can be assessed early in semester one, specifically, measures of personality, motivation and learning strategies. A variety of data mining models are compared to assess the relative accuracy of each.

Pb7. **Gray, G., McGuinness, C., and Owende, P. (2013) Investigating the efficacy of algorithmic student modelling in predicting students at risk of failing in tertiary education.** ***Young researcher track, 6th International Conference on Educational Data Mining (EMD 2013)*****, Memphis, July, 2013.**

This paper was presented at the young researcher track at the 6th International Conference on Educational Data Mining.

Abstract: The increasing numbers enrolling for college courses, and increased diversity in the classroom, poses a challenge for colleges in enabling all students achieve their potential. This paper reports on a study to model factors, using data mining techniques, that are predictive of college academic performance, and can be measured during first year enrolment. Data was gathered over three years, and focused on a diverse student population of first year students from a range of academic disciplines ($n \approx$ 1,100). Initial models generated on two years of data (n=713) demonstrate high accuracy. Advice is sought on additional analysis approaches to consider.

Appendix B

The Study Dataset

This appendix augments the details provided in Chapter 3 on the study participants and the online learner profiler. The mission statement in Section B.1 was referenced in Section 3.2 as evidence of ITB's commitment to encouragement of a diverse student population. Section 3.2 also described how permission to use student data in this study was requested; Section B.2.1 gives the paper consent form used and Section B.2.2 has the text from the introductory page of the online profiler, which also requested consent to use participant data in this study. Factors measured by the online profiler were discussed in Section 3.3.2; Section B.2.3 lists the questionnaire items used. Finally Section B.3 illustrates how first year GPA is calculated, as discussed in Section 3.3.3.

B.1 ITB mission statement

The mission of the Institute is to serve its students and the community by meeting the skills needs in the economy and increasing the level of participation in third-level education and training, particularly in Dublin North-West and its environs.

The Institute will do this:

- by achieving consistently high standards of relevance and quality in teaching, research, development and consultancy.
- by offering a welcoming and supportive environment to students from all educational and social backgrounds and to adults wishing to increase or update their level of technical skills.

The Institute is adopting admissions and student support policies to ensure that a relatively high proportion of its students are 'non-standard entrants' such as mature students:

- applicants without Leaving Certificate qualifications who can meet entry requirements in other ways
- students with disabilities
- students from disadvantaged socio-economic backgrounds

B.2 Online learner profiler

B.2.1 Consent form

The following paper based consent form was given to all students before completing the online questionnaire:

The Learning Styles questionnaire asks a range of questions based on factors that affect how you study and how you learn. In addition, current research at the Institute of Technology Blanchardstown is investigating if there is a link between these factors and academic performance. This research is investigating group trends only, and will not be investigating the profiles or academic performance of individual students.

I give permission for my details to be used anonymously for research into links between learning styles and academic performance.

Signed: ____________________ Date: ____________
Printed Name: ___________________

I also give permission for any relevant information to be seen by appropriate members of the Assessment Services.
Signed: ____________________ Date: ____________
Printed Name: ___________________

B.2.2 Learner profiler: introductory pages

The following text is from the online profiler (`www.howilearn.ie`; Access code: 12345). *The profiler was available online for all students at ITB to use at any time. The study dataset included profiling done during first year induction only.*

This tool will ask you a number of questions, following which you will be told what your learning style is, in other words how you learn best.

The tool does NOT record the answers you give to each question, just your learning style which is determined from the answers you give.

The data gathered by this system is used to generate summary reports for each class, informing lecturers of the range of learning styles in each class. Lecturers do not have access to an individual's profile. The data gathered will only be used for research into group trends.

Please enter your e-mail: ________________________
Please enter Student Number: _________________
Please enter your date of birth ____________
Please select your gender : Male _ Female _
What category of student are you? *(Options: Apprentice, Full time, Part time, Postgraduate, Other)*
Which year/phase are you in? *(Options: Year 1, Year 2, Year 3, Year 4, Year 5)*
What month does your course start?: *(Options: Jan, Mar, Jun, Sep)*
Please select your course: *(Options: all ITB courses were listed in a drop down box)*

Current research at ITB is investigating if there is a link between learning profiles and academic performance. Do we have your permission to compare your learning profile with your end of year Grade Point Average (GPA)? This is an analysis into group trends only. Individuals will not be identified: Yes _ No _

The following questionnaire contains 69 questions. Each question has at least two answers. Select one answer for each question. Do not spend too much time thinking about each question. The correct answer is generally your initial instinct. When you are finished, you will get feedback on your learning style under various headings.

Click here to start the Test

B.2.3 Learner profiler: questionnaire items

This section includes the 69 items included in the online profiler. Table B.1 maps items to their corresponding study factor.

Table B.1: Mapping learner profiler questions to study factors

Question range	Factor
1-6	Conscientiousness
12-15	Extrinsic goal orientation
16-19	Group work
20-23	Intrinsic goal orientation
25-28	Learning approach
29-40	Learner modality
41-46	Openness
52-55	Self-efficacy
56-60	Meta cognitive self-regulation
61-63	Study effort
66-69	Study time

Please select one answer to each question

Question 1. I am always prepared
☐ Strongly agree
☐ Agree
☐ Neither Agree or Disagree
☐ Disagree
☐ Strongly Disagree

Question 2. I find it difficult to get down to work
☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 3. I often forget to put things back in their proper place

☐ Strongly agree

☐ Agree

☐ Neither Agree nor Disagree

☐ Disagree

☐ Strongly Disagree

Question 4. I leave my belongings around.

☐ Strongly agree

☐ Agree

☐ Neither Agree nor Disagree

☐ Disagree

☐ Strongly Disagree

Question 5. I pay attention to detail

☐ Strongly agree

☐ Agree

☐ Neither Agree nor Disagree

☐ Disagree

☐ Strongly Disagree

Question 6. I like to do things according to a plan or schedule

☐ Strongly agree

☐ Agree

☐ Neither Agree nor Disagree

☐ Disagree

☐ Strongly Disagree

Question 12. I get the most satisfaction if I get good grades.

☐ Strongly agree

☐ Agree

☐ Neither Agree nor Disagree

☐ Disagree

☐ Strongly Disagree

Question 13. I get great satisfaction from doing well, which drives me to work hard.

☐ Strongly agree

☐ Agree

☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 14. It is important to show my family and friends that I can do well.
☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 15. If I can, I would like to get better grades than most of the people in my class.
☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 16. The idea of doing a work presentation in groups
☐ appeals to you?
☐ does not appeal to you?

Question 17. Do you prefer to
☐ spend time on your own or with one other person?
☐ spend time in the company of others?

Question 18. Do you like to
☐ share ideas with other people?
☐ or work best on your own?

Question 19. Do you prefer to
☐ work as a part of a team?
☐ to get on with a piece of work on your own?

Question 20. I like classes where the material arouses my curiosity, even if it is more difficult.
☐ Strongly agree
☐ Agree

☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 21. I prefer work that is challenging so I can learn new things.
☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 22. If choosing a topic for an essay, I would pick a topic I can learn from, even if it means more work.
☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 23. I get the most satisfaction if I thoroughly understand what I am studying.
☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 25. When preparing for an exam, would you
☐ hope to pass the exam by learning off the key points.
☐ make sure you understand the topic completely.
☐ use past papers to decide if you need to understand the material or just learn it off.

Question 26. Which of the following is more important to you?
☐ Try hard to understand course material.
☐ Try to pass the course while doing as little work as possible.
☐ Do well in assessments and exams.

Question 27. Which of the following describes you best?
☐ I will come to class interested to learn.
☐ I will find most classes boring.

☐ I will come to class to find out how to get a good grade.

Question 28. Which of the following would you agree with most?
☐ I would only do extra reading if it was required to pass an exam.
☐ Extra reading is a waste of time and can be confusing
☐ I like spending time reading up on topics that interest me.

Question 29. Do you tend to remember
☐ faces?
☐ names?

Question 30. Do you prefer
☐ to get on with a practical task and try it out?
☐ read up about it first, so you know what you need to do?

Question 31. Do you prefer to get information in
☐ pictures, diagrams, graphs, or maps?
☐ a written format?

Question 32. Do you prefer working out solutions to problems
☐ by doing the task and then seeing how it works?
☐ by talking about the task first?

Question 33. Do you remember best
☐ what you see?
☐ what you hear?

Question 34. Do you prefer to
☐ listen to the music/radio/TV?
☐ play sport/go for a walk?

Question 36. Do you in your home or work area
☐ move your furniture several times in a year?
☐ like to keep the same arrangement?

Question 37. Do you have
☐ a place for everything and like everything to be in its place?
☐ tend to put things where they land?

Question 38. Would you tend to, if you were to hang a picture on a wall,
☐ carefully measure to be sure it is centered and straight?

☐ put it where it looks right and move it if necessary?

Question 39. Do you tend
☐ to scan read when reading a new article or piece of work?
☐ to read every word carefully?

Question 40. Do you when making notes
☐ write everything out?
☐ make bullet points/key points?

Question 41. I love to daydream.
☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 42. I get excited by new ideas.
☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 43. I avoid philosophical discussions.
☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 44. I do not like poetry.
☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 45. I like art and creativity.

☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 46. I rarely look for deeper meaning in things.
☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 53. I think I will be good at completing assessment work to a high standard.
☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 54. I am confident I can understand the more complex material that will be taught on this course.
☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 55. I expect to do very well on this course.
☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 56. I set goals for each study period in order to direct my activities.
☐ Strongly agree
☐ Agree

☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 57. During class, I often miss important points because I am thinking of something else.

☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 58. When I am confused by something I am studying, I would try to go back and figure it out.

☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 59. I often find that I have been studying for class but I dont know what it was all about.

☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 60. Before studying a new topic, I often skim through it first to see how it is organised.

☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 61. I am often so bored by what I am studying that I quit before I am finished.

☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 62. I would work hard to do well even if I do not like what I am doing.
☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 63. When course work is difficult, I give up or just study the easy parts.
☐ Strongly agree
☐ Agree
☐ Neither Agree nor Disagree
☐ Disagree
☐ Strongly Disagree

Question 66. When I am studying, I make good use of my time
☐ Yes
☐ No
☐ I don't know yet

Question 67. It is hard to find time to study because of other activities
☐ Yes
☐ No
☐ I don't know yet

Question 68. I plan to attend class regularly.
☐ Yes
☐ No
☐ I don't know yet

Question 69. I have a regular place set aside for studying.
☐ Yes
☐ No
☐ I don't know yet

B.3 Procedure for calculating GPA

The following is from ITB policy document 4RAS04: Grade Point Average (GPA) calculations:

The aggregate performance of an individual student is represented by the student's Grade Point Average (GPA). In order to determine the GPA for a particular student, the following calculation is carried out.

(A) A Grade Point Value is assigned to the alphabetic grade a student has gained for each subject as illustrated in Table B.2.

(B) The Grade Point Value is multiplied by the Credits (available from the Approved Course Schedule) to arrive at a Grade Credit Score for each subject/module.

(C) The Grade Credit Scores are then added together and divided by the credits for the stage or semester to arrive at the GPA.

(D) Credits gained as a result of being awarded an X or a G in a module are not included in the calculation of the GPA.

To pass a candidate must have 60 credits (unless exemption granted) and a GPA$\geq$2.00. Table B.3 illustrates a worked example of GPA calculations using steps A to D above.

Table B.2: Module grades and their corresponding grade point value

Grade	Percentage band	Grade Point Value	Credits awarded
A	80 - 100	4.00	Yes
B+	70 - 79	3.50	Yes
B	60 - 69	3.00	Yes
B-	55 - 59	2.75	Yes
C+	50 - 54	2.50	Yes
C	40 - 49	2.00	Yes
D	35 - 39	1.50	Yes
F	<35	0.00	No

One or more D grades can be compensated for with C+ or higher grades in other subjects.

Table B.3: Example of calculating grade point average (GPA)

Module	A Credits	B Grade	C Grade Point Value	D (A x B) Grade Point Score
Construction Technology	10	B	3.0	30
Site Management	5	C	2.0	10
Civil Engineering Design	10	B	3.0	30
Mathematics	5	B	3.0	15
Advanced Surveying	10	B	3.0	30
Quality Management	5	C	2.0	10
Project	15	B	3.0	45
Total	60			170
GPA (170/60):	2.83 Merit			

Appendix C

Additional Data Exploration

This appendix includes four figures relating to study factors and two figures depicting classification models.

1. Section 3.3.2.1 discussed that lack of normality amongst study factor distributions. This is illustrated in histograms of student factors in Figure C.1. Figure C.1 also gives the skewness and kurtosis of each factor, referenced in Section 4.2.3 to aid the interpretation of outlier detection results.

2. Section 5.2 detailed correlations between study factors. Figure 5.2 included the Pearson correlation coefficient only, confidence intervals for each correlation are given in Figure C.2, indicating statistical significance.

3. The notched box plots of GPA by course in Figure C.3 were referenced in Section 4.3.4 as an illustration of variations in class imbalance found in subgroups by course.

4. Section 6.4 discussed the impact of study attributes on model predictive accuracy. While model performances by subgroup were unreliable due to a small sample size, correlation between GPA and study factor detailed in Figure C.5 informed suggestions for future work detailed in Section 6.4.2.

5. Section 5.5.3 discussed classification models in terms of insights offered on how participants were classified, and referenced the Decision Tree model ($Model_{2012}$) in Figure C.6 and the Back-propagation Neural Network model ($Model_{2012}$) in Figure C.7.

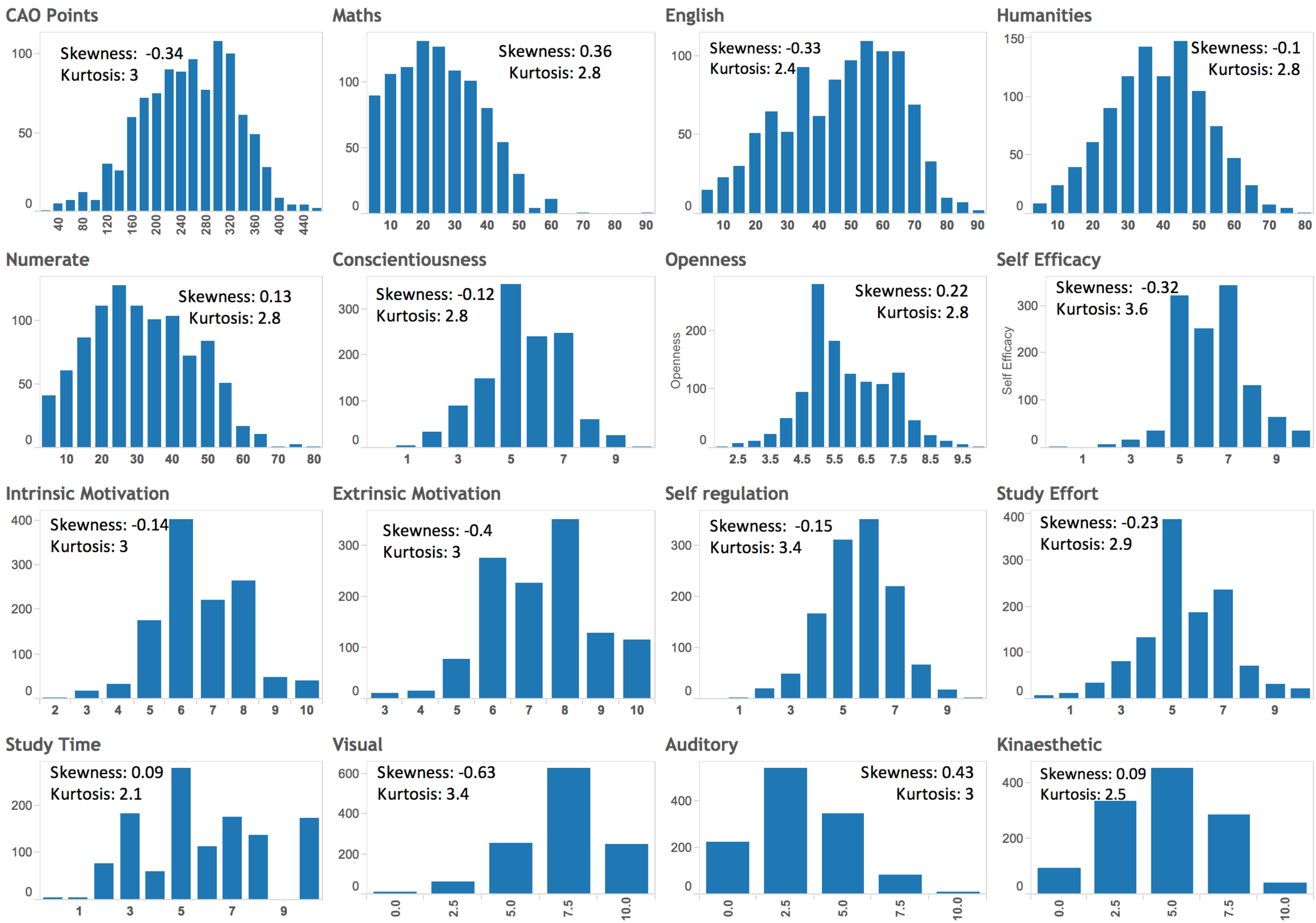

Figure C.1: Histograms for each study factor

	GPA	Conscientiousness	Openness	Self-efficacy	Extrinsic goal
Conscientiousness	0.150 [0.09, 0.21]				
Openness	0.084 [0.03, 0.14]	0.032 [-0.03, 0.09]			
Self-efficacy	0.120 [0.06, 0.18]	0.313 [0.26, 0.36]	0.178 [0.12, 0.23]		
Extrinsic goal	0.124 [0.07, 0.18]	0.280 [0.22, 0.33]	0.049 [-0.01, 0.11]	0.308 [0.25, 0.36]	
Intrinsic goal	0.149 [0.09, 0.21]	0.334 [0.28, 0.39]	0.316 [0.26, 0.37]	0.421 [0.37, 0.47]	0.381 [0.33, 0.43]
MC self-regulation	0.130 [0.08, 0.18]	0.515 [0.47, 0.56]	0.101 [0.04, 0.16]	0.409 [0.36, 0.46]	0.298 [0.24, 0.35]
Study effort	0.187 [0.14, 0.24]	0.450 [0.39, 0.50]	0.064 [0.01, 0.13]	0.334 [0.28, 0.39]	0.232 [0.17, 0.29]
Study time	0.101 [0.04, 0.16]	0.396 [0.35, 0.44]	0.009 [-0.05, 0.06]	0.259 [0.21, 0.31]	0.175 [0.12, 0.23]
Deep learner	0.234 [0.18, 0.29]	0.352 [0.30, 0.40]	0.209 [0.15, 0.26]	0.273 [0.22, 0.32]	0.158 [0.10, 0.21]
Strategic learner	-0.158 [-0.22, -0.1]	-0.167 [-0.22, -0.11]	-0.174 [-0.23, -0.12]	-0.158 [-0.21, -0.1]	-0.012 [-0.06, 0.04]
Shallow learner	-0.146 [-0.21, -0.09]	-0.330 [-0.38, -0.28]	-0.096 [-0.15, -0.04]	-0.221 [-0.28, -0.16]	-0.234 [-0.29, -0.17]
Group work	-0.008 [-0.13, -0.02]	0.052 [-0.01, 0.11]	-0.042 [-0.10, 0.02]	0.056 [0.00, 0.11]	0.059 [0.00, 0.12]
Age	0.250 [0.20, 0.30]	0.156 [0.11, 0.20]	0.038 [-0.02, 0.09]	0.038 [-0.02, 0.09]	0.051 [-0.01, 0.10]
Gender	0.100 [0.05, 0.15]	-0.005 [-0.06, 0.05]	0.022 [-0.03, 0.08]	-0.048 [-0.10, 0.00]	0.035 [-0.02, 0.09]
Visual	0.005 [-0.01, 0.11]	0.069 [0.01, 0.13]	0.063 [0.00, 0.12]	-0.024 [-0.08, 0.03]	0.041 [-0.02, 0.10]
Auditory	0.002 [-0.04, 0.08]	0.073 [0.02, 0.13]	0.023 [-0.03, 0.08]	-0.002 [-0.06, 0.05]	0.013 [-0.04, 0.07]
Kinaesthetic	-0.059 [-0.11, 0.00]	-0.124 [-0.18, -0.07]	-0.074 [-0.13, -0.02]	0.022 [-0.03, 0.07]	-0.046 [-0.10, 0.02]

	Intrinsic goal	MC self-regulation	Study effort	Study time	Deep learner
MC self-regulation	0.429 [0.38, 0.48]				
Study effort	0.330 [0.27, 0.38]	0.594 [0.55, 0.63]			
Study time	0.227 [0.17, 0.28]	0.452 [0.40, 0.49]	0.378 [0.33, 0.43]		
Deep learner	0.417 [0.37, 0.47]	0.431 [0.38, 0.48]	0.360 [0.31, 0.41]	0.285 [0.23, 0.34]	
Strategic learner	-0.274 [-0.33, -0.22]	-0.213 [-0.26, -0.16]	-0.133 [-0.18, -0.07]	-0.115 [-0.17, -0.06]	-0.791 [-0.81, -0.77]
Shallow learner	-0.294 [-0.36, -0.24]	-0.398 [-0.44, -0.35]	-0.394 [-0.44, -0.34]	-0.290 [-0.34, -0.24]	-0.519 [-0.56, -0.48]
Group work	0.027 [-0.03, 0.09]	0.113 [0.06, 0.17]	0.094 [0.04, 0.15]	0.084 [0.03, 0.14]	0.020 [-0.03, 0.07]
Age	0.257 [0.21, 0.30]	0.234 [0.18, 0.28]	0.210 [0.16, 0.26]	0.023 [-0.03, 0.08]	0.284 [0.23, 0.33]
Gender	0.004 [-0.05, 0.06]	0.005 [-0.05, 0.06]	0.023 [-0.03, 0.08]	0.086 [0.03, 0.14]	0.086 [0.03, 0.14]
Visual	0.054 [-0.01, 0.11]	0.024 [-0.04, 0.08]	-0.003 [-0.06, 0.06]	0.038 [-0.02, 0.09]	0.067 [0.01, 0.13]
Auditory	-0.016 [-0.07, 0.04]	0.065 [0.011, 0.12]	0.039 [-0.01, 0.10]	0.081 [0.03, 0.14]	0.077 [0.02, 0.13]
Kinaesthetic	-0.032 [-0.09, 0.02]	-0.078 [-0.13, -0.02]	-0.033 [-0.09, 0.02]	-0.105 [-0.16, -0.05]	-0.126 [-0.18, -0.07]

	Strategic learner	Shallow learner	Group work	Age	Gender
Shallow learner	-0.103 [-0.15, -0.06]				
Group work	0.037 [-0.02, 0.09]	-0.081 [-0.14, -0.02]			
Age	-0.200 [-0.25, -0.15]	-0.181 [-0.22, -0.14]	-0.022 [-0.08, 0.03]		
Gender	-0.001 [-0.06, 0.06]	-0.130 [-0.18, -0.07]	0.026 [-0.03, 0.08]	-0.038 [-0.09, 0.02]	
Visual	-0.020 [-0.08, 0.04]	-0.089 [-0.15, -0.03]	0.021 [-0.04, 0.08]	-0.038 [-0.09, 0.02]	-0.046 [-0.10, 0.01]
Auditory	-0.068 [-0.13, -0.01]	-0.026 [-0.09, 0.03]	-0.097 [-0.15, -0.04]	0.025 [-0.03, 0.08]	0.205 [0.15, 0.26]
Kinaesthetic	0.078 [0.02, 0.13]	0.099 [0.04, 0.15]	0.069 [0.01, 0.12]	-0.055 [-0.11, 0.00]	-0.144 [-0.20, -0.09]

	Visual	Auditory			
Auditory	-0.347 [-0.40, -0.29]				
Kinaesthetic	-0.541 [-0.58, -0.50]	-0.601 [-0.64, -0.57]			

MC: Metacognitive; Intervals are 95% Confidence Intervals based on 1,999 bootstrap samples

Figure C.2: Heat map of correlations with confidence intervals for non-cognitive factors, all participants

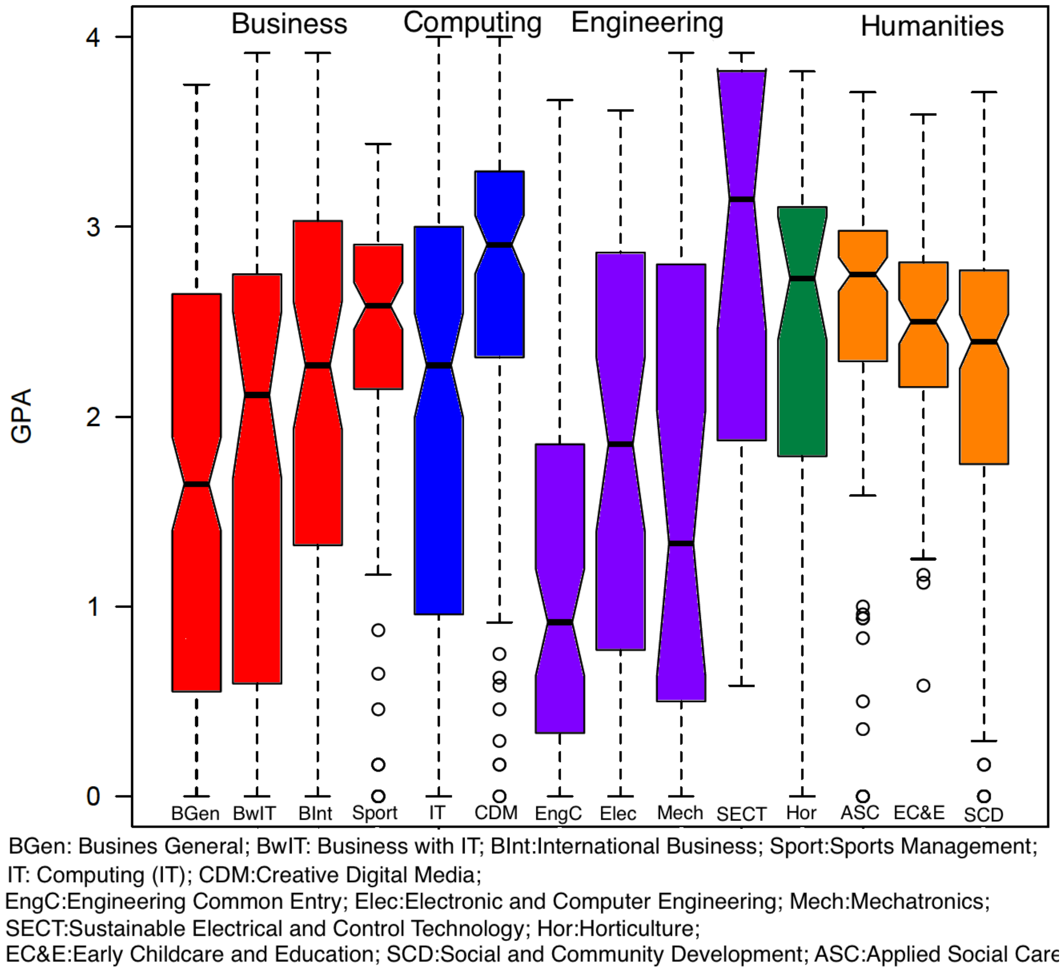

Figure C.3: Notched box plots of GPA by course of study

Sub-	Mean		Prior academic performance						Personality		Motivation			Self-regulation			Learning Approach			Other			Modality		
group	age	*n*	CAO	Maths	English	Numer	Human	Applie	Con	Open	SE	EM	IM	SR	StE	StT	Deep	Stra	Shal	Group	Age	Gen	Vis	Aud	Kin
all	23	1207	0.29	0.27	0.17	0.30	0.23	0.17	0.15	0.08	0.12	0.12	0.15	0.13	0.19	0.10	0.23	-0.15	-0.14	-0.08	0.25	0.10	0.05	0.02	-0.05
B-all	21	402	0.35	0.31	0.26	0.30	0.33	0.25	0.21	-0.02	0.08	0.15	0.08	0.11	0.19	0.12	0.18	-0.14	-0.05	-0.03	0.21	-0.02	0.08	-0.04	-0.03
BGen	21	183	0.26	0.23	0.25	0.22	0.26	0.17	0.23	0.09	0.16	0.21	0.12	0.17	0.20	0.10	0.19	-0.08	-0.16	-0.05	0.14	0.10	0.15	0.04	-0.17
BwIT	22	60	0.55	0.52	0.31	0.51	0.34	0.33	0.26	-0.11	-0.01	0.32	-0.03	0.21	0.46	0.13	0.20	-0.16	-0.09	-0.15	0.34	-0.07	-0.05	-0.15	0.16
BInt	21	64	0.36	0.26	0.09	0.22	0.38	0.15	0.18	-0.38	-0.08	0.09	0.07	0.09	0.12	0.04	0.26	-0.24	-0.07	-0.04	0.29	-0.12	-0.01	-0.05	0.05
Sports	23	95	0.26	0.26	0.30	0.26	0.24	0.22	0.17	0.02	0.05	-0.02	0.05	-0.05	0.05	0.28	0.00	-0.08	0.11	0.05	0.10	0.05	0.04	0.02	-0.06
C-all	24	239	0.18	0.24	-0.03	0.22	0.06	0.10	0.18	0.15	0.12	0.13	0.21	0.14	0.16	0.12	0.29	-0.13	-0.29	-0.09	0.28	0.14	0.09	0.01	-0.08
IT	24	137	-0.02	0.16	-0.27	0.11	-0.12	-0.02	0.28	0.18	0.18	0.18	0.34	0.29	0.21	0.16	0.37	-0.19	-0.32	-0.14	0.42	0.13	0.21	-0.06	-0.12
CDM	23	102	0.09	0.26	-0.08	0.19	-0.01	0.09	0.10	-0.01	0.02	0.18	0.09	0.10	0.21	0.08	0.22	-0.11	-0.19	0.01	0.09	0.00	-0.07	0.08	0.00
E-all	22	172	0.19	0.29	-0.03	0.26	0.14	0.08	0.14	0.10	0.26	0.14	0.29	0.21	0.25	0.11	0.37	-0.28	-0.23	-0.19	0.51	0.04	0.05	0.04	-0.07
EngC	20	73	0.26	0.13	0.03	0.11	0.14	0.20	0.18	0.05	0.21	0.26	0.24	0.22	0.15	0.08	0.36	-0.28	-0.18	0.06	0.41	0.11	0.24	0.10	-0.26
Elec	22	52	0.32	0.44	-0.08	0.43	0.28	0.00	0.35	0.28	0.42	0.21	0.48	0.31	0.35	0.23	0.48	-0.47	-0.24	-0.06	0.39	0.09	0.03	0.01	-0.04
Mech	21	27	0.10	0.30	-0.01	0.21	0.06	-0.04	0.24	0.02	0.42	0.06	0.19	0.15	0.30	0.27	0.20	-0.03	-0.27	-0.16	0.27	0.08	-0.21	0.06	0.14
SECT	27	20	0.23	0.49	0.15	0.38	0.19	0.12	0.08	0.25	0.63	0.34	0.60	0.50	0.55	0.34	0.65	-0.33	-0.58	-0.17	0.68	-0.23	0.06	0.05	-0.11
Hort	28	41	0.07	0.15	-0.09	0.19	0.01	0.24	-0.21	-0.02	0.02	0.09	0.12	0.00	-0.03	-0.15	0.05	-0.15	0.13	-0.14	0.19	0.18	0.05	-0.39	0.30
H-all	25	353	0.25	0.21	0.24	0.26	0.27	0.12	0.05	0.02	0.05	0.09	0.01	0.08	0.15	0.06	0.07	-0.05	-0.04	0.01	0.03	-0.03	0.05	0.00	-0.04
ASC	28	146	0.22	0.28	0.28	0.22	0.23	0.20	-0.09	-0.06	0.04	0.00	-0.08	-0.07	0.07	0.00	0.05	-0.04	-0.03	-0.11	-0.01	0.03	0.02	-0.05	0.02
EC&E	22	80	0.44	0.30	0.37	0.32	0.43	0.23	0.19	0.03	0.09	0.19	0.14	0.18	0.15	0.26	-0.06	0.14	-0.10	-0.04	-0.19	-0.02	-0.07	0.08	-0.03
SCD	25	127	0.27	0.17	0.21	0.31	0.29	0.02	0.14	0.02	0.04	0.14	0.03	0.18	0.22	0.09	0.14	-0.14	-0.03	0.09	0.05	-0.07	0.15	0.01	-0.16
[18,23]	20	852	0.44	0.31	0.29	0.38	0.37	0.03	0.14	0.08	0.10	0.12	0.10	0.08	0.14	0.10	0.19	-0.10	-0.13	-0.06	0.12	0.19	0.00	0.05	-0.05
[24,28]	25	154	0.16	0.26	0.07	0.26	0.06	0.06	-0.02	-0.04	0.13	0.01	0.04	0.06	0.07	0.10	0.14	-0.14	-0.03	-0.09	0.00	0.08	0.14	-0.02	-0.10
[29,60]	38	201	0.17	0.05	0.07	0.00	0.04	0.30	0.06	0.11	0.12	0.13	0.05	0.02	0.19	0.10	0.06	-0.09	0.04	-0.16	-0.08	-0.19	0.12	-0.12	0.01
Male	24	713	0.26	0.27	0.12	0.28	0.20	0.02	0.19	0.14	0.23	0.15	0.25	0.19	0.25	0.12	0.33	-0.23	-0.20	-0.07	0.33		0.09	-0.01	-0.07
Female	23	494	0.30	0.27	0.21	0.29	0.25	0.05	0.10	-0.01	-0.06	0.07	-0.02	0.03	0.08	0.04	0.05	-0.05	0.00	-0.11	0.12		-0.01	0.02	-0.01

Con:Conscientiousness; Open:Openness; SE:Self-efficacy; IM:Intrinsic goal orientation; EM:Extrinsic goal orientation; SR: Metacognitive self-regulation; StE: Study effort; StT:Study time; Deep: Deep learner; Shal: Shallow learner; Stra: Strategic learner; Group:Likes to work in groups; Gen=Gender; Vis:Visual modality; Aud:Auditory modality; Kin:Kinaesthetic modality; B-all:all Business students; BGen:Business General; BwIT:Business with IT; BInt: International Business; Sport:Sports Management; C=all:all Computing students; IT:Computing(IT); CDM: Creative Digital Media; Eng=all:all Engineering students; EngC:Engineering Common Entry; Elec:Electrical and Computer Engineering; Mech:Mechatronics; SECT:Sustainable Electrical and Control Technology; Hor:Horticulture; H-all:all Humanities students; ACS:Applied Social Care; EC&E:Early Childcare and Education; SCD:Social and Community Development. [18,23], [24,28], [29,60]: age ranges.

Figure C.4: Heat map of correlations between study factors and GPA, for subgroups by age, gender and course of study

	All *n*=1207	Business General *n*=183	Business with IT *n*=60	International Business *n*=64	Sports Management & Coaching *n*=95
Conscientiousness	0.150 [0.09, 0.21]	0.226 [0.06, 0.37]	0.259 [-0.02, 0.48]	0.175 [-0.05, 0.39]	0.169 [-0.03, 0.36]
Openness	0.084 [0.03, 0.14]	0.090 [-0.05, 0.22]	-0.113 [-0.34, 0.12]	-0.379 [-0.58, -0.10]	0.023 [-0.24, 0.22]
Self-efficacy	0.120 [0.06, 0.18]	0.162 [0.01, 0.30]	-0.012 [-0.25, 0.20]	-0.081 [-0.35, 0.15]	0.053 [-0.15, 0.24]
Extrinsic goal	0.124 [0.07, 0.18]	0.212 [0.05, 0.35]	0.319 [0.04, 0.55]	0.091 [-0.12, 0.33]	-0.018 [-0.21, 0.17]
Intrinsic goal	0.149 [0.09, 0.21]	0.124 [-0.04, 0.28]	-0.028 [-0.28, 0.24]	0.072 [-0.19, 0.31]	0.046 [-0.12, 0.21]
Self-regulation	0.130 [0.08, 0.18]	0.173 [0.03, 0.31]	0.214 [-0.03, 0.44]	0.094 [-0.13, 0.35]	-0.050 [-0.21, 0.14]
Study effort	0.187 [0.14, 0.24]	0.203 [0.05, 0.34]	0.461 [0.22, 0.64]	0.118 [-0.09, 0.32]	0.051 [-0.13, 0.23]
Study time	0.101 [0.04, 0.16]	0.099 [-0.04, 0.25]	0.128 [-0.11, 0.38]	0.044 [-0.21, 0.32]	0.278 [0.11, 0.44]
Deep learner	0.234 [0.18, 0.29]	0.189 [0.05, 0.34]	0.202 [-0.07, 0.42]	0.259 [-0.02, 0.48]	0.004 [-0.16, 0.15]
Strategic learner	-0.158 [-0.22, -0.10]	-0.075 [-0.21, 0.07]	-0.160 [-0.40, 0.11]	-0.239 [-0.47, -0.01]	-0.078 [-0.24, 0.09]
Shallow learner	-0.146 [-0.21, -0.09]	-0.162 [-0.30, -0.02]	-0.087 [-0.34, 0.14]	-0.071 [-0.35, 0.21]	0.107 [-0.03, 0.24]
Group work	-0.080 [-0.13, -0.02]	-0.051 [-0.20, 0.09]	-0.148 [-0.39, 0.11]	-0.042 [-0.31, 0.24]	0.046 [-0.17, 0.28]
Age	0.250 [0.20, 0.30]	0.135 [-0.07, 0.29]	0.337 [0.13, 0.54]	0.287 [-0.01, 0.49]	0.098 [-0.09, 0.26]
Gender	0.100 [0.05, 0.15]	0.104 [-0.05, 0.24]	-0.073 [-0.32, 0.18]	-0.121 [-0.36, 0.14]	0.052 [-0.23, 0.23]
Visual	0.050 [-0.01, 0.11]	0.148 [0.00, 0.28]	-0.049 [-0.28, 0.19]	-0.007 [-0.24, 0.24]	0.044 [-0.14, 0.24]
Auditory	0.020 [-0.04, 0.08]	0.038 [-0.12, 0.16]	-0.150 [-0.41, 0.13]	-0.045 [-0.29, 0.18]	0.024 [-0.22, 0.22]
Kinaesthetic	-0.059 [-0.11, 0.00]	-0.168 [-0.31,-0.03]	0.164 [-0.08, 0.38]	0.048 [-0.19, 0.28]	-0.060 [-0.24, 0.14]

	Computing (IT) *n*=137	Creative Digital Media *n*=102	Applied Social Care *n*=146	Early childcare & Education *n*=80	Social & Community Development *n*=127
Conscientiousness	0.279 [0.11, 0.43]	0.099 [-0.13, 0.33]	-0.085 [-0.24 , 0.07]	0.190 [-0.03, 0.38]	0.136 [-0.03, 0.31]
Openness	0.178 [0.03, 0.33]	-0.012 [-0.24, 0.21]	-0.064 [-0.20, 0.08]	0.029 [-0.26, 0.28]	0.024 [-0.18, 0.20]
Self-efficacy	0.175 [0.00, 0.32]	0.024 [-0.19, 0.22]	0.039 [-0.17, 0.25]	0.092 [-0.15, 0.31]	0.036 [-0.14, 0.21]
Extrinsic goal	0.183 [0.02, 0.35]	0.180 [0.01, 0.37]	0.004 [-0.19, 0.21]	0.186 [-0.03, 0.38]	0.139 [-0.04, 0.31]
Intrinsic goal	0.341 [0.18, 0.47]	0.094 [-0.11, 0.30]	-0.079 [-0.27, 0.12]	0.135 [-0.15, 0.34]	0.029 [-0.13, 0.18]
Self-regulation	0.286 [0.13, 0.43]	0.102 [-0.08, 0.29]	-0.065 [-0.27, 0.17]	0.182 [-0.02, 0.37]	0.175 [0.02, 0.32]
Study effort	0.205 [0.05, 0.35]	0.213 [0.02, 0.41]	0.067 [-0.11, 0.30]	0.147 [-0.05, 0.33]	0.222 [0.04, 0.38]
Study time	0.164 [-0.02, 0.31]	0.075 [-0.18, 0.28]	-0.003 [-0.18, 0.17]	0.258 [-0.01, 0.46]	0.092 [-0.10, 0.27]
Deep learner	0.371 [0.22, 0.52]	0.217 [0.02, 0.40]	0.045 [-0.10, 0.18]	-0.062 [-0.30, 0.17]	0.144 [-0.05, 0.32]
Strategic learner	-0.191 [-0.34, -0.02]	-0.112 [-0.33, 0.07]	-0.040 [-0.21, 0.09]	0.142 [-0.07, 0.37]	-0.141 [-0.34, 0.04]
Shallow learner	-0.324 [-0.49, -0.13]	-0.186 [-0.36, -0.01]	-0.030 [-0.20, 0.11]	-0.097 [-0.29, 0.08]	-0.028 [-0.19, 0.12]
Group work	-0.137 [-0.29, 0.03]	0.013 [-0.19, 0.21]	-0.109 [-0.24, 0.06]	-0.037 [-0.27, 0.20]	0.093 [-0.09, 0.27]
Age	0.418 [0.28, 0.52]	0.091 [-0.22, 0.32]	-0.006 [-0.14, 0.12]	-0.192 [-0.45, 0.14]	0.051 [-0.20, 0.25]
Gender	0.127 [-0.04, 0.24]	0.002 [-0.21, 0.19]	0.027 [-0.12, 0.20]	-0.020 [-0.10, 0.13]	-0.074 [-0.23, 0.13]
Visual	0.210 [0.06, 0.36]	-0.067 [-0.25, 0.14]	0.024 [-0.12, 0.15]	-0.071 [-0.32, 0.13]	0.154 [-0.04, 0.32]
Auditory	-0.063 [-0.21, 0.09]	0.077 [-0.11, 0.24]	-0.047 [-0.21, 0.11]	0.079 [-0.15, 0.26]	0.011 [-0.18, 0.21]
Kinaesthetic	-0.121 [-0.28, 0.05]	-0.002 [-0.18, 0.22]	0.021 [-0.13, 0.16]	-0.025 [-0.23, 0.19]	-0.160 [-0.30, -0.02]

	Engineering, common entry *n*=73	Electronic & Computer Engineering *n*=52	Mechatronics *n*=27	Sustainable Electrical & Control Technology *n*=20	Horticulture *n*=41
Conscientiousness	0.183 [-0.03, 0.38]	0.347 [0.12, 0.54]	0.235 [-0.15, 0.56]	0.077 [-0.30, 0.48]	-0.213 [-0.49, 0.12]
Openness	0.052 [-0.18, 0.26]	0.279 [0.04, 0.48]	0.023 [-0.37, 0.47]	0.253 [-0.24, 0.66]	-0.019 [-0.24, 0.24]
Self-efficacy	0.209 [0.00, 0.38]	0.421 [0.17, 0.60]	0.419 [-0.02, 0.72]	0.630 [0.37, 0.78]	0.020 [-0.29, 0.32]
Extrinsic goal	0.255 [0.01, 0.43]	0.211 [-0.08, 0.46]	0.058 [-0.39, 0.43]	0.337 [-0.16, 0.73]	0.089 [-0.19, 0.39]
Intrinsic goal	0.235 [-0.02, 0.48]	0.477 [0.21, 0.65]	0.186 [-0.28, 0.57]	0.599 [0.03, 0.82]	0.121 [-0.17, 0.34]
Self-regulation	0.221 [-0.03, 0.45]	0.307 [0.00, 0.54]	0.150 [-0.24, 0.55]	0.502 [0.09, 0.77]	0.000 -0.26, 0.26]
Study effort	0.153 [-0.10, 0.39]	0.347 [0.08, 0.56]	0.303 [-0.11, 0.63]	0.547 [0.22, 0.76]	-0.026 [-0.30, 0.21]
Study time	0.077 [-0.15, 0.32]	0.231 [-0.05, 0.49]	0.269 [-0.21, 0.65]	0.340 [-0.04, 0.59]	-0.154 [-0.38, 0.12]
Deep learner	0.358 [0.13, 0.54]	0.483 [0.25, 0.65]	0.203 [-0.24, 0.59]	0.649 [0.32, 0.84]	0.053 [-0.25, 0.39]
Strategic learner	-0.284 [-0.48, -0.01]	-0.468 [-0.66, -0.17]	-0.034 [-0.42, 0.35]	-0.333 [-0.66, 0.15]	-0.151 [-0.47, 0.23]
Shallow learner	-0.176 [-0.34, -0.01]	-0.243 [-0.50, 0.02]	-0.266 [-0.51, 0.03]	-0.580 [-0.85, -0.09]	0.131 [-0.14, 0.34]
Group work	0.061 [-0.20, 0.26]	-0.059 [-0.32, 0.21]	-0.157 [-0.50, 0.20]	-0.168 [-0.52, 0.30]	-0.136 [-0.40, 0.16]
Age	0.406 [0.14, 0.64]	0.388 [0.22, 0.55]	0.265 [-0.21, 0.55]	0.675 [0.36, 0.80]	0.189 [-0.22, 0.46]
Gender	0.111 [-0.04, 0.28]	0.091 [-0.02, 0.26]	0.079 [-0.34, 0.50]	-0.231 [-0.43, -0.06]	0.179 [-0.22, 0.39]
Visual	0.244 [0.00, 0.46]	0.032 [-0.25, 0.31]	-0.213 [-0.63, 0.27]	0.063 [-0.40, 0.49]	0.054 [-0.27, 0.38]
Auditory	0.103 [-0.14, 0.31]	0.010 [-0.29, 0.28]	0.057 [-0.37, 0.46]	0.053 [-0.35, 0.50]	-0.391 [-0.63, -0.13]
Kinaesthetic	-0.256 [-0.48, 0.00]	-0.035 [-0.29, 0.25]	0.138 [-0.31, 0.50]	-0.105 [-0.52, 0.33]	0.304 [0.05, 0.51]

Figure C.5: Heat map of correlations with confidence intervals between non-cognitive factors and GPA, by course of study

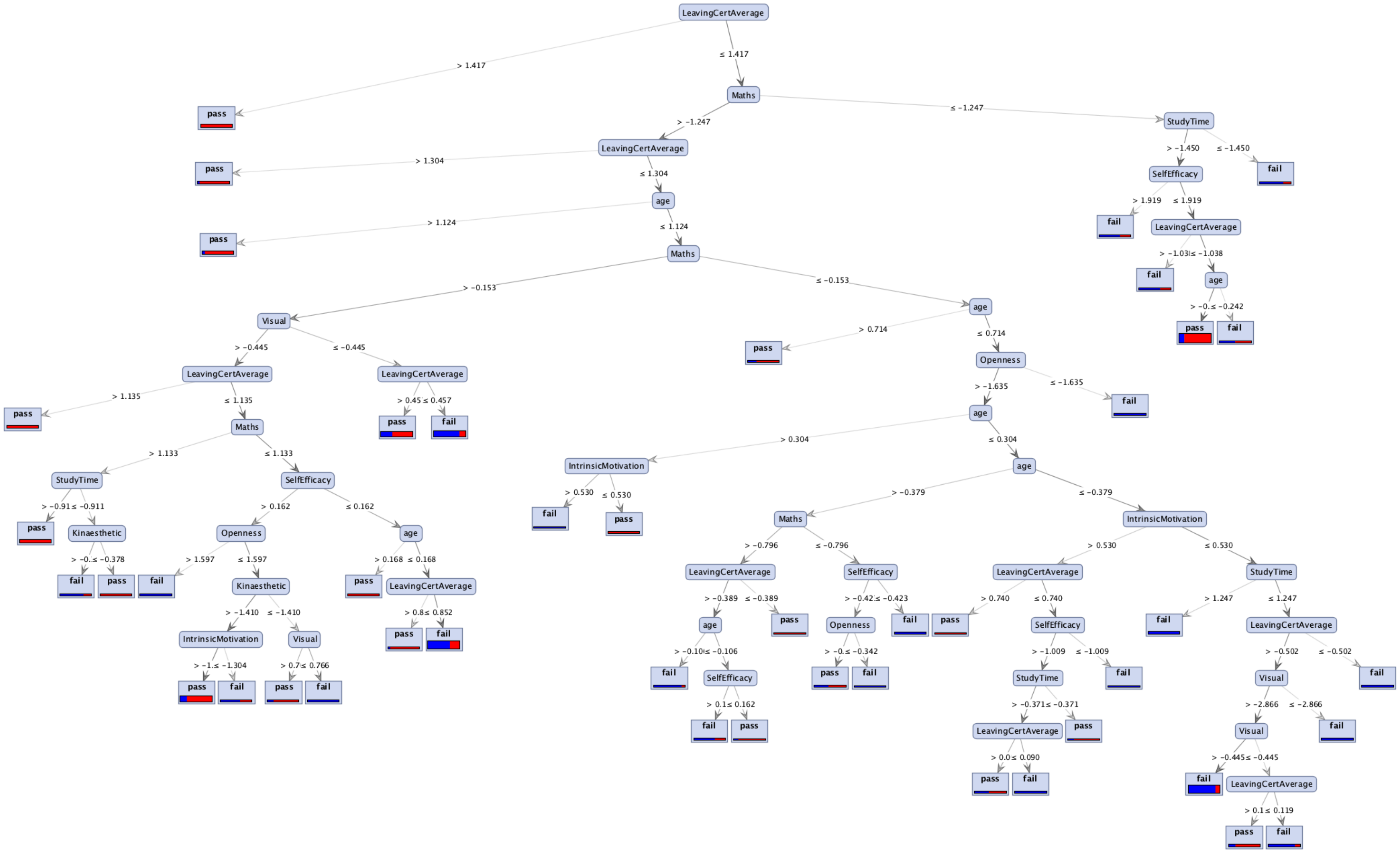

Figure C.6: Decision Tree model ($Model_{2012}$)

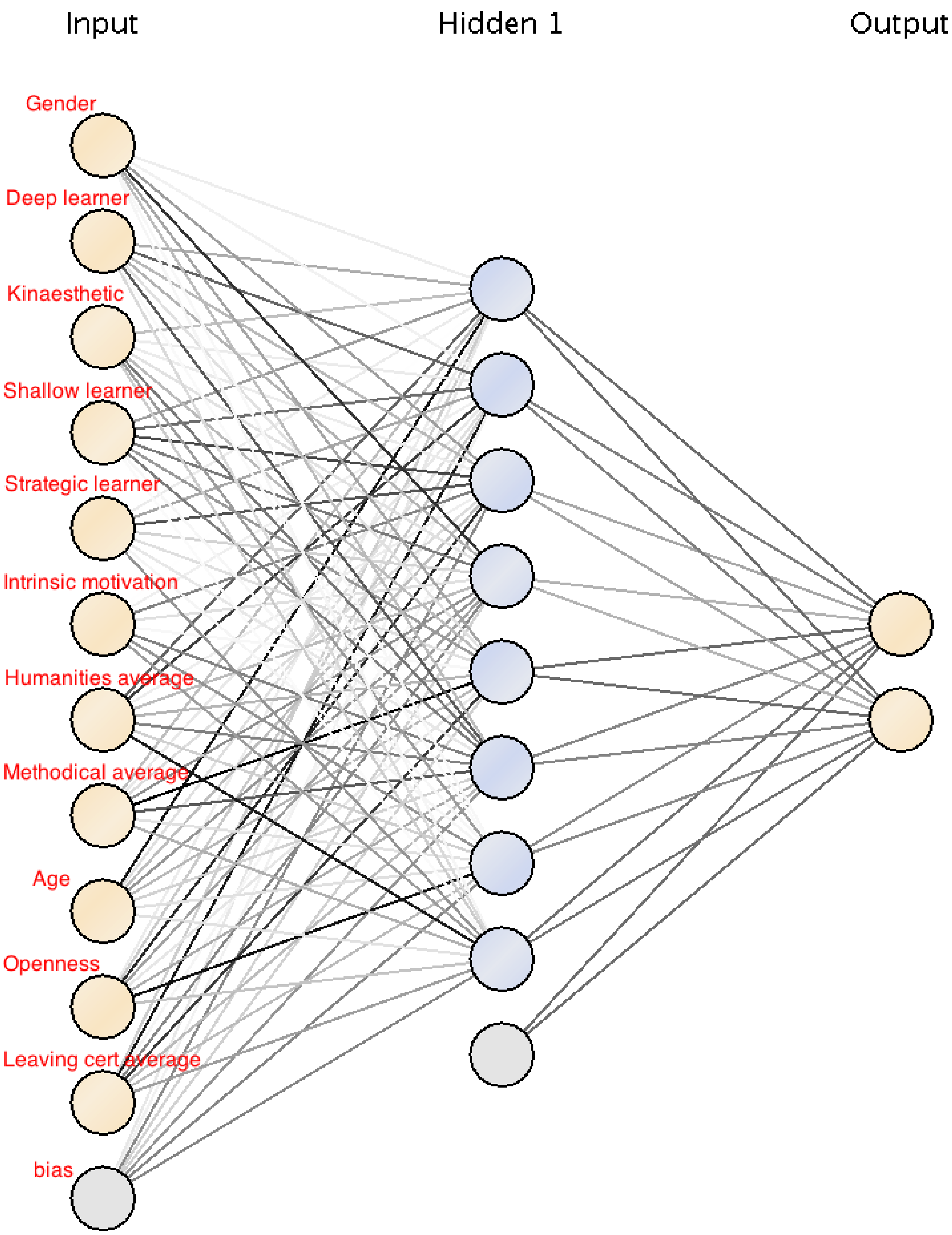

Figure C.7: Back-propagation Neural Network model ($Model_{2012}$). Line weightings reflect the magnitude of input weights.

Appendix D

Java and R Code Used

This Appendix gives code listings for Java and R code used in the study, specifically:

1. Section D.1 gives the Java code used to calculate factors of prior academic performance as discussed in Section 3.3.1.

2. Correlation significance was verified using 1,999 Bootstrap Confidence Intervals with BCa as explained in Section 3.4.1. Section D.2 gives the R code used.

3. The R code used for analysis of group differences detailed in Section 3.4.2 is given in Section D.3.

4. As explained in Section 3.4.3, regression models predicting GPA were based on optimal attribute subsets identified using an exhaustive search as implemented in the *regsubsets* function in R. The code used is given in Section D.4.

5. The R code to test for significant differences in model accuracies using Fishers exact test, and McNemar's test, is given in D.5 and described in Section 3.5.10.

6. Section D.6 lists the R code used to generate SMOTE (synthetic minority oversampling) examples as explained in Section 4.3.4.

D.1 Calculations for CAO points

```
/*
 * To calculate the total number of CAO points and average LC for each student.
 * @author Geraldine Gray
 */
package leavingcertdata;

import java.io.BufferedReader;
import java.io.FileReader;
import java.io.FileWriter;
import java.util.StringTokenizer;
import java.io.IOException;
import java.io.PrintWriter;
import java.text.DecimalFormat;

public class Main {
    //Files
    static PrintWriter pwResults;
    static FileWriter fwResults;
    static PrintWriter pwInput;
    static FileWriter fwInput;
    //array to hold top six results
    static Integer Points[] = new Integer[6];
    //Record if student was in a special category
    static int special = 0;
    static int numOfSubjects = 0;
    static int numberOfStudents = 0;
    static int totalPoints=0;
    static Boolean firstStudent = true;
    //variables for current and previous student numbers
    static String oldStudent = null;
    static String currentStudent = null;

    public static void main(String[] args) {
        //path and INPUT FILE name.  Change path and filename as appropriate
        String pathName = "/Volumes/PhDdata/leavingCert/";
        String fileName = "SubjectsWithPoints2013.csv";
        String strFile = pathName + fileName;
        // initialise points array
        initialisePointsArray();
        // open output file and output column headings
        openOutputFiles(pathName);
       // process inputfile
        readCsv(strFile);
        closeFiles();
    }

    public static void initialisePointsArray() {
        for (int i = 0; i < 6; i++) {
            Points[i] = 0;
        }
    }

    public static void openOutputFiles(String pathName) {
        //open the five output files and write column headings
        try {
            // Write to OUTPUT FILE, change name as appropriate
            fwResults = new FileWriter(pathName + "/3ResultsWithCAOPoints2013.csv");
            pwResults = new PrintWriter(fwResults);
            // output studentID and CAOPoints
            pwResults.print("Id,");
            pwResults.println("CAOPoints,NumberSubjects,AverageMark");
        } catch (Exception e) {
            System.out.println("Exception opening output file: " + e);
        }
    }

    public static void readCsv(String strFile) {
```

```
//open the output files and write column headings
    try {
        BufferedReader br = new BufferedReader(new FileReader(strFile));
        String strLine = "";
        StringTokenizer st = null;
        String currentPoints = null;
        String notNeeded = null;
        // Ignore first line (column headings)
        strLine = br.readLine();
        //read comma separated file line by line
        while ((strLine = br.readLine()) != null) {
            st = new StringTokenizer(strLine, ";");
            currentPoints = st.nextToken();
            currentStudent = st.nextToken();
            //Is it a new student? if so write out values for previous student and reset variables
            if (!(currentStudent.equals(oldStudent))) {
                calcCaoPoints();
                special = 0;
                numOfSubjects = 0;
                totalPoints=0;
                initialisePointsArray();
                oldStudent = currentStudent;
            }
            //Process current row
            updateCaoPoints(currentPoints);
        }
        //Process last student
        oldStudent = currentStudent;
        calcCaoPoints();
    } catch (Exception e) {
        System.out.println("Exception while reading csv file: " + e);
    }
}

static void updateCaoPoints(String points) {
    // if this is a special (no points) catch it here
    if (points.equals("special")) {
        special++;
    } else {
        numOfSubjects++;
        //find current lowest number of points & its position
        int smallest = Points[0];
        int smallPosition = 0;
        for (int i = 1; i < 6; i++) {
            if (Points[i] < smallest) {
                smallest = Points[i];
                smallPosition = i;
            }
        }
        // Is latest value greater than current lowest? if so replace current lowest
        int pointValue = Integer.parseInt(points);
        if (pointValue > smallest) {
            Points[smallPosition] = pointValue;
        //Calculate overall total for average LC
        totalPoints=totalPoints+pointValue;
        }
    }
}

static void calcCaoPoints() {
    int averageMark;
    if (!(oldStudent == null)) {
        numberOfStudents++;
   //calculate CAO points by adding top 6 results
        int CAOPoints = 0;
        for (int i = 0; i < 6; i++) {
            CAOPoints += Points[i];
    //calculate average points of across all subject results
            if (numOfSubjects==0)
```

```
              averageMark=0;
               else
                        averageMark=totalPoints/numOfSubjects;
            }
            pwResults.println(oldStudent + "," + CAOPoints + "," + numOfSubjects + "," + averageMark);
        }
    }

    static void closeFiles() {
        try {
            //Flush the output to the file
            pwResults.flush();
            pwResults.close();
            fwResults.close();
        } catch (Exception e) {
            System.out.println("Exception while writing csv file: " + e);
        }
    }
}
```

D.2 Bootstrap Confidence Intervals using the Bias corrected and accelerated method

```
library(boot)

#Read dataset
data<-read.csv("FinalDataset.csv",header=TRUE)

#optionally, calculate the bootstrap statistics on a subgroup within the dataset. Below defines subgroups for male and female:
#1 Males:
data<-data[data$gender==0,]
#2: Females
data<-data[data$gender==1,]

################################################################
#Select prior academic performance factors for students with leaving cert results
CAO<-data[data$CAOPoints>0,c("GPA","CAOPoints","MethodicalAverage","HumanitiesAverage","Maths","English","AppliedAverage","age")]

#Verify the correct number of rows are selected:
nrow(CAO)

#Define a function for each correlation required
fCAO <- function(CAO, i) cor(CAO[i, 1], CAO[i, 2])
fMaths <- function(CAO, i) cor(CAO[i, 1], CAO[i, 5])
fEng <- function(CAO, i) cor(CAO[i, 1], CAO[i, 6])
fNumerate <- function(CAO, i) cor(CAO[i, 1], CAO[i, 3])
fHum <- function(CAO, i) cor(CAO[i, 1], CAO[i, 4])
CAOapplied<-CAO[CAO$AppliedAverage>0,]
fApp <- function(CAOapplied, i) cor(CAOapplied[i, 1], CAOapplied[i, 7])

#Generate 1999 bootstrap statistics using  correlation functions defined above
resultCAO<-boot(data = CAO, statistic = fCAO, R = 1999)
resultMaths<-boot(data = CAO, statistic = fMaths, R = 1999)
resultEng<-boot(data = CAO, statistic = fEng, R = 1999)
resultNumerate<-boot(data = CAO, statistic = fNumerate, R = 1999)
resultHum<-boot(data = CAO, statistic = fHum, R = 1999)
resultApp<-boot(data = CAOapplied, statistic = fApp, R = 1999)

#Calculate BCa confidence intervals, default is 95% confidence interval (conf=0.95)
boot.ci(resultCAO, index=1, type="bca")
boot.ci(resultMaths, index=1, type="bca")
boot.ci(resultEng, index=1, type="bca")
boot.ci(resultNumerate, index=1, type="bca")
boot.ci(resultHum, index=1, type="bca")
```

```
boot.ci(resultApp, index=1, type="bca")

#check other confidence intervals, for example
boot.ci(resultApp, index=1, type="bca", conf=0.999)
boot.ci(resultApp, index=1, type="bca", conf=0.99)

##########################################################################################
#Select non cognitive factors (psychometric)
#Note: Before running these commands, update 'data' to hold either the entire data, or the subgroup of interest)

PsyData<-data[,c("GPA","Conscientiousness","Openness","SelfEfficacy","ExtrinsicMotivation","IntrinsicMotivation","SelfRegulation",
"StudyEffort", "StudyTime", "DeepLearner", "StrategicLearner", "ShallowLearner", "GroupWork", "age", "gender", "Visual",
"Auditory", "Kinaesthetic")]

#verify the correct number of rows are selected:
nrow(PsyData)

#Define a function for correlations with GPA
fCon <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 2])
fOpen <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 3])
fSE <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 4])
fEM <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 5])
fIM <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 6])
fSR <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 7])
fStE <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 8])
fStT <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 9])
fDeep <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 10])
fStr <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 11])
fSh <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 12])
fGr <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 13])
fAge <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 14])
fGen <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 15])
fV <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 16])
fA <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 17])
fK <- function(PsyData, i) cor(PsyData[i, 1], PsyData[i, 18])

#Generate 1999 bootstrap statistics using correlation functions defined above
resultCon<-boot(data = PsyData, statistic = fCon, R = 1999)
resultOpen<-boot(data = PsyData, statistic = fOpen, R = 1999)
resultSE<-boot(data = PsyData, statistic = fSE, R = 1999)
resultEM<-boot(data = PsyData, statistic = fEM, R = 1999)
resultIM<-boot(data = PsyData, statistic = fIM, R = 1999)
resultSR<-boot(data = PsyData, statistic = fSR, R = 1999)
resultStE<-boot(data = PsyData, statistic = fStE, R = 1999)
resultStT<-boot(data = PsyData, statistic = fStT, R = 1999)
resultDeep<-boot(data = PsyData, statistic = fDeep, R = 1999)
resultStr<-boot(data = PsyData, statistic = fStr, R = 1999)
resultSh<-boot(data = PsyData, statistic = fSh, R = 1999)
resultGr<-boot(data = PsyData, statistic = fGr, R = 1999)
resultAge<-boot(data = PsyData, statistic = fAge, R = 1999)
resultGen<-boot(data = PsyData, statistic = fGen, R = 1999)
resultV<-boot(data = PsyData, statistic = fV, R = 1999)
resultA<-boot(data = PsyData, statistic = fA, R = 1999)
resultK<-boot(data = PsyData, statistic = fK, R = 1999)

#Generate 1999 bootstrap statistics using correlation functions defined above
#Default is 95% confidence interval (conf=0.95)
boot.ci(resultCon, index=1, type="bca")
boot.ci(resultOpen, index=1, type="bca")
boot.ci(resultSE, index=1, type="bca")
boot.ci(resultEM, index=1, type="bca")
boot.ci(resultIM, index=1, type="bca")
boot.ci(resultSR, index=1, type="bca")
boot.ci(resultStE, index=1, type="bca")
boot.ci(resultStT, index=1, type="bca")
boot.ci(resultDeep, index=1, type="bca")
boot.ci(resultStr, index=1, type="bca")
boot.ci(resultSh, index=1, type="bca")
boot.ci(resultGr, index=1, type="bca")
```

```
boot.ci(resultAge, index=1, type="bca")
boot.ci(resultGen, index=1, type="bca")
boot.ci(resultV, index=1, type="bca")
boot.ci(resultA, index=1, type="bca")
boot.ci(resultK, index=1, type="bca")

#Check other confidence intervals, for example:
boot.ci(resultStT, index=1, type="bca", conf=0.999)
boot.ci(resultOpen, index=1, type="bca", conf=0.99)

#Alternatively, overwrite the functions above to check for
#correlations between psychometric factors themselves:
#Conscientiousness
fOpen <- function(PsyData, i) cor(PsyData[i, 2], PsyData[i, 3])
fSE <- function(PsyData, i) cor(PsyData[i, 2], PsyData[i, 4])
fEM <- function(PsyData, i) cor(PsyData[i, 2], PsyData[i, 5])
fIM <- function(PsyData, i) cor(PsyData[i, 2], PsyData[i, 6])
fSR <- function(PsyData, i) cor(PsyData[i, 2], PsyData[i, 7])
fStE <- function(PsyData, i) cor(PsyData[i, 2], PsyData[i, 8])
fStT <- function(PsyData, i) cor(PsyData[i, 2], PsyData[i, 9])
fDeep <- function(PsyData, i) cor(PsyData[i, 2], PsyData[i, 10])
fStr <- function(PsyData, i) cor(PsyData[i, 2], PsyData[i, 11])
fSh <- function(PsyData, i) cor(PsyData[i, 2], PsyData[i, 12])
fGr <- function(PsyData, i) cor(PsyData[i, 2], PsyData[i, 13])
fAge <- function(PsyData, i) cor(PsyData[i, 2], PsyData[i, 14])
fGen <- function(PsyData, i) cor(PsyData[i, 2], PsyData[i, 15])
fV <- function(PsyData, i) cor(PsyData[i, 2], PsyData[i, 16])
fA <- function(PsyData, i) cor(PsyData[i, 2], PsyData[i, 17])
fK <- function(PsyData, i) cor(PsyData[i, 2], PsyData[i, 18])

#similarly for Openness which was the 3rd attribute, e.g.
fSE <- function(PsyData, i) cor(PsyData[i, 3], PsyData[i, 4])
fEM <- function(PsyData, i) cor(PsyData[i, 3], PsyData[i, 5])
 . . .
#similarly for Self-efficacy which was the 4th attribute, e.g.
fEM <- function(PsyData, i) cor(PsyData[i, 4], PsyData[i, 5])
fIM <- function(PsyData, i) cor(PsyData[i, 4], PsyData[i, 6])
 . . .

#similarly for Extrinsic goal, the 5th attribute, e.g.
fIM <- function(PsyData, i) cor(PsyData[i, 5], PsyData[i, 6])
fSR <- function(PsyData, i) cor(PsyData[i, 5], PsyData[i, 7])
 . . .
#similarly for Intrinsic goal, the 6th attribute, e.g.
fSR <- function(PsyData, i) cor(PsyData[i, 6], PsyData[i, 7])
fStE <- function(PsyData, i) cor(PsyData[i, 6], PsyData[i, 8])
 . . . .
#similarly for metacognitive self-regulation which was the 7th attribute, e.g.
fStE <- function(PsyData, i) cor(PsyData[i, 7], PsyData[i, 8])
fStT <- function(PsyData, i) cor(PsyData[i, 7], PsyData[i, 9])
 . . . .
#similarly for Study effort,  the 8th attribute, e.g.
fStT <- function(PsyData, i) cor(PsyData[i, 8], PsyData[i, 9])
fDeep <- function(PsyData, i) cor(PsyData[i, 8], PsyData[i, 10])
 . . . .
#similarly for Study time,  the 9th attribute, e.g.
fDeep <- function(PsyData, i) cor(PsyData[i, 9], PsyData[i, 10])
fStr <- function(PsyData, i) cor(PsyData[i, 9], PsyData[i, 11])
 . . .
#similarly for Deep learner ,  the 10th attribute, e.g.
fStr <- function(PsyData, i) cor(PsyData[i, 10], PsyData[i, 11])
fSh <- function(PsyData, i) cor(PsyData[i, 10], PsyData[i, 12])
 . . .
#similarly for Strategic learner, the 11th attribute, e.g.
fSh <- function(PsyData, i) cor(PsyData[i, 11], PsyData[i, 12])
fGr <- function(PsyData, i) cor(PsyData[i, 11], PsyData[i, 13])
 . . . .
#similarly for Shallow learner, the 12th attribute, e.g.
fGr <- function(PsyData, i) cor(PsyData[i, 12], PsyData[i, 13])
```

```
fAge <- function(PsyData, i) cor(PsyData[i, 12], PsyData[i, 14])
 . . . .
#similarly for Group, the 13th attribute, e.g.
fAge <- function(PsyData, i) cor(PsyData[i, 13], PsyData[i, 14])
fGen <- function(PsyData, i) cor(PsyData[i, 13], PsyData[i, 15])
 . . . .
 #similarly for Age, the 14th attribute, e.g.
fGen <- function(PsyData, i) cor(PsyData[i, 14], PsyData[i, 15])
fV <- function(PsyData, i) cor(PsyData[i, 14], PsyData[i, 16])
 . . .
 #similarly for Gender, the 15th attribute, e.g.
fV <- function(PsyData, i) cor(PsyData[i, 15], PsyData[i, 16])
fA <- function(PsyData, i) cor(PsyData[i, 15], PsyData[i, 17])
 . . .
 #similarly for Visual, the 16th attribute, e.g.
 fA <- function(PsyData, i) cor(PsyData[i, 16], PsyData[i, 17])
fK <- function(PsyData, i) cor(PsyData[i, 16], PsyData[i, 18])
#and Auditory with Kinesthetic
fK <- function(PsyData, i) cor(PsyData[i, 17], PsyData[i, 18])
```

D.3 Analysis of group differences

```
#The following codes was used in Analysis of group differences. It includes testing for normality using two or three subgroups,
#testing for equality of variance across two and three groups, t-test for two groups and ANOVA for three groups with post hoc test.
library(splines)
library(stats4)
library(mvtnorm)
library(VGAM)
library(lawstat)
library(plyr)atim}

#Full dataset
data<-read.csv("FinalDataset.csv",header=TRUE)
#Participants with leaving certificate results
dataLC<-data[data$CAOPoints>0,]

#1. Testing attributes means are normally distributed by generating 50 bootstrap samples of each attribute.
#Also testing the list of means have equal variance.
#Code here is for gender (2 groups) and GPA. Code for age group (three groups) is below:

#Variable 'col' defines the attribute to test, change this to test other attributes.
col<-"GPA"
#The following lines define the subgroups. Change the conditions to define other subgroups
myColFemale<-data[data$gender>0.5,c(col)]
myColMales<-data[data$gender<0.5,c(col)]

#Generate 50 bootstrap samples for each subgroup:
resamplesFemale <- lapply(1:50, function(i)
sample(myColFemale, replace = T))
resamplesMale <- lapply(1:50, function(i)
sample(myColMales, replace = T))

#Calculate the mean for each sample
r.mean.female <- sapply(resamplesFemale, mean)
r.mean.male <- sapply(resamplesMale, mean)

#Test for normality
shapiro.test(r.mean.female)
shapiro.test(r.mean.male)

#Compare variances:
#Convert list of means to a dataframe, and rename the column to 'mean'
MeanFemale<-as.data.frame(r.mean.female)
MeanFemale<-rename(MeanFemale, c("r.mean.female"="mean"))
MeanMale<-as.data.frame(r.mean.male)
MeanMale<-rename(MeanMale, c("r.mean.male"="mean"))

#Add a gender column to each column of means
```

```
females<-cbind(gender="female", MeanFemale)
males<-cbind(gender="male", MeanMale)

#Combine both list of means into one data frame
both<-rbind(females,males)

#Test for equal variance
levene.test(both$mean, both$gender)

######################################################
# 2. As above, testing attributes means are normally distributed by generating 50 bootstrap samples of each attribute,
# and testing the list of means have equal variance.
# Code here is for age groups (3 groups) and CAO points. Dataset is stored in 'dataC'.

#Variable 'col' defines the attribute to test, change this to test other attributes.
col<-"CAOPoints"
#The following lines define the subgroups. Change the conditions to define other subgroups
myCol1<-dataC[dataC$age<=23,c(col)]
myCol2<-dataC[dataC$age<=28 & dataC$age>23,c(col)]
myCol3<-dataC[dataC$age>28,c(col)]

#Generate 50 bootstrap samples for each subgroup:
resamples1 <- lapply(1:999, function(i)
sample(myCol1, 852, replace = T))
resamples2 <- lapply(1:999, function(i)
sample(myCol2, 154, replace = T))
resamples3 <- lapply(1:999, function(i)
sample(myCol3, 201, replace = T))

#Calculate the mean for each sample
r.mean.1 <- sapply(resamples1, mean)
r.mean.2 <- sapply(resamples2, mean)
r.mean.3 <- sapply(resamples3, mean)

#Test for normality
shapiro.test(r.mean.1)
shapiro.test(r.mean.2)
shapiro.test(r.mean.3)

#Compare variances: Convert list of means to a dataframe, and rename the column the 'mean'
Mean1<-as.data.frame(r.mean.1)
Mean1 <-rename(Mean1, c("r.mean.1"="mean"))
Mean2<-as.data.frame(r.mean.2)
Mean2 <-rename(Mean2, c("r.mean.2"="mean"))
Mean3<-as.data.frame(r.mean.3)
Mean3 <-rename(Mean3, c("r.mean.3"="mean"))

#Add an age column to each column of means
group1<-cbind(age="young", Mean1)
group2<-cbind(age="mid", Mean2)
group3<-cbind(age="mature", Mean3)

#Combine both list of means into one data frame
both<-rbind(group1, group2, group3)

#Test for equal variance
levene.test(both$mean, both$age)

######################################################
#3. Running t.tests for two groups. The example here is running Welch's t-test for gender.

t.test(data$GPA ~ data$gender, var.equal=FALSE)
t.test(dataLC$CAOPoints ~ dataLC$gender, var.equal= FALSE)
t.test(dataLC$English ~ dataLC$gender, var.equal= FALSE)
t.test(dataLC$Maths ~ dataLC$gender, var.equal=FALSE)
t.test(dataLC$HumanitiesAverage ~ dataLC$gender, var.equal= FALSE)
t.test(dataLC$AppliedAverage ~ dataLC$gender, var.equal= FALSE)
t.test(dataLC$MethodicalAverage ~ dataLC$gender, var.equal= FALSE)
t.test(data$Conscientiousness ~ data$gender, var.equal= FALSE)
```

```
t.test(data$Openness ~ data$gender, var.equal= FALSE)
t.test(data$SelfEfficacy ~ data$gender, var.equal= FALSE)
t.test(data$IntrinsicMotivation ~ data$gender, var.equal= FALSE)
t.test(data$ExtrinsicMotivation ~ data$gender, var.equal= FALSE)
t.test(data$SelfRegulation ~ data$gender, var.equal= FALSE)
t.test(data$StudyEffort ~ data$gender, var.equal= FALSE)
t.test(data$StudyTime ~ data$gender, var.equal= FALSE)
t.test(data$DeepLearner ~ data$gender, var.equal= FALSE)
t.test(data$ShallowLearner ~ data$gender, var.equal= FALSE)
t.test(data$StrategicLearner ~ data$gender, var.equal= FALSE)
t.test(data$Visual ~ data$gender, var.equal= FALSE)
t.test(data$Auditory ~ data$gender, var.equal= FALSE)
t.test(data$Kinaesthetic ~ data$gender, var.equal= FALSE)
t.test(data$GroupWork ~ data$gender, var.equal= FALSE)
t.test(data$age ~ data$gender, var.equal= FALSE)
t.test(dataLC$LeavingCertAverage ~ data$gender, var.equal= FALSE)

######################################################
#4. Running ANOVA for three groups with Tukey post hoc test. The example here is for Age groups.

fit <- aov(GPA ~ ageCat, data=data)
summary(fit)
TukeyHSD(fit)
fit <- aov(Conscientiousness ~ ageCat, data=data)
summary(fit)
TukeyHSD(fit)
fit <- aov(Openness ~ ageCat, data=data)
summary(fit)
TukeyHSD(fit)
fit <- aov(SelfEfficacy ~ ageCat, data=data)
summary(fit)
TukeyHSD(fit)
fit <- aov(IntrinsicMotivation ~ ageCat, data=data)
summary(fit)
TukeyHSD(fit)
fit <- aov(ExtrinsicMotivation ~ ageCat, data=data)
summary(fit)
TukeyHSD(fit)
fit <- aov(SelfRegulation ~ ageCat, data=data)
summary(fit)
TukeyHSD(fit)
fit <- aov(StudyEffort ~ ageCat, data=data)
summary(fit)
TukeyHSD(fit)
fit <- aov(StudyTime ~ ageCat, data=data)
summary(fit)
TukeyHSD(fit)
fit <- aov(DeepLearner ~ ageCat, data=data)
summary(fit)
TukeyHSD(fit)
fit <- aov(ShallowLearner ~ ageCat, data=data)
summary(fit)
TukeyHSD(fit)
fit <- aov(StrategicLearner ~ ageCat, data=data)
summary(fit)
TukeyHSD(fit)
fit <- aov(Visual ~ ageCat, data=data)
summary(fit)
TukeyHSD(fit)
fit <- aov(Auditory ~ ageCat, data=data)
summary(fit)
TukeyHSD(fit)
fit <- aov(Kinaesthetic ~ ageCat, data=data)
summary(fit)
TukeyHSD(fit)
fit <- aov(CAOPoints ~ ageCat, data=dataLC)
summary(fit)
TukeyHSD(fit)
fit <- aov(AverageLC ~ ageCat, data=dataLC)
```

```
summary(fit)
TukeyHSD(fit)
fit <- aov(Maths ~ ageCat, data=dataLC)
summary(fit)
TukeyHSD(fit)
fit <- aov(English ~ ageCat, data=dataLC)
summary(fit)
TukeyHSD(fit)
fit <- aov(MethodicalAverage ~ ageCat, data=dataLC)
summary(fit)
TukeyHSD(fit)
fit <- aov(HumanitiesAverage ~ ageCat, data=dataLC)
summary(fit)
TukeyHSD(fit)
fit <- aov(AppliedAverage ~ ageCat, data=dataLC)
summary(fit)
TukeyHSD(fit)

########################################################
# T-test results were confirmed using Wilcoxon paired test.
#Code below illustrates this for gender

males<-data[data$gender==0,]
malesLC<-dataLC[dataLC$gender==0,]
females<-data[data$gender==1,]
femalesLC<-dataLC[dataLC$gender==1,]

wilcox.test(males$GPA,females$GPA)
wilcox.test(malesLC$CAOPoints,femalesLC$CAOPoints)
wilcox.test(malesLC$English,femalesLC$English)
wilcox.test(malesLC$Maths,femalesLC$Maths)
wilcox.test(malesLC$HumanitiesAverage,femalesLC$HumanitiesAverage)
wilcox.test(malesLC$MethodicalAverage,femalesLC$MethodicalAverage)
wilcox.test(malesLC$AppliedAverage,femalesLC$AppliedAverage)
wilcox.test(males$Conscientiousness,females$Conscientiousness)
wilcox.test(males$Openness,females$Openness)
# . . . and similarly for other study attributes

########################################################
# ANOVA results were verified using Kruskal-Wallis.
# Post hoc used Wilcoxon paired tests using Holm adjustment.
#Code below illustrates this for GPA bands.
data<-read.csv("FinalDataset.csv",header=TRUE)
dataLC<-data[data$CAOPoints>0,]

kruskal.test(CAOPoints~GPA,data=data)
pairwise.wilcox.test(data$CAOPoints, data$GPA, p.adjust.method = "holm")
kruskal.test(AverageLC~GPA,data=dataLC)
pairwise.wilcox.test(dataLC$AverageLC, dataLC$GPA, p.adjust.method = "holm")
kruskal.test(Maths~GPA,data=dataLC)
pairwise.wilcox.test(dataLC$Maths, dataLC$GPA, p.adjust.method = "holm")
kruskal.test(English~GPA,data=dataLC)
pairwise.wilcox.test(dataLC$English, dataLC$GPA, p.adjust.method = "holm")
kruskal.test(ScienceAverage~GPA,data=dataLC)
pairwise.wilcox.test(dataLC$MethodicalAverage, dataLC$GPA, p.adjust.method = "holm")
kruskal.test(HumanitiesAverage~GPA,data=dataLC)
pairwise.wilcox.test(dataLC$HumanitiesAverage, dataLC$GPA, p.adjust.method = "holm")
kruskal.test(Conscientiousness~GPA,data=data)
pairwise.wilcox.test(data$Conscientiousness, data$GPA, p.adjust.method = "holm")
kruskal.test(Openness~GPA,data=data)
pairwise.wilcox.test(data$Openness, data$GPA, p.adjust.method = "holm")
# . . . and similarly for other study attributes
```

D.4 Linear regression

```
library(MASS)
library(leaps)
```

```
#Choose a dataset:
#All participants
data<-read.csv("FinalDataset.csv",header=TRUE)
#Participants with leaving certificate results
data<-data[data$CAOPoints>0,]
#Participants with leaving certificate results and age <= 21
data<-data[data$CAOPoints>0 & data$age<=21,]
#Run regression model and plot results
models<-regsubsets(GPA~AverageLC+CAOPoints+ English+ Maths+ AppliedAverage+ HumanitiesAverage+ MethodicalAverage+Conscientiousness+
Openness+IntrinsicMotivation+ExtrinsicMotivation+ SelfEfficacy+SelfRegulation+StudyEffort+StudyTime+gender+GroupWork+age+DeepLearner+
ShallowLearner+StrategicLearner+Visual+Auditory,data=data,nbest=5)
plot(models, scale="r2")
```

D.5 Comparing model accuracies, Fisher and McNemar

Fisher's exact test, confirmed using Chi-squared

```
#Define matrix columns, rows and data entries.
rnames2 <- c("XVal", "2012")
cnames2 <- c("Correct", "Incorrect")
#Define the contingency table: the following line is adjusted based on the performance of each algorithm
cells2 <- c(1144,352,393,153)  #by row

#Create the contingency matrix
mymatrix <- matrix(cells2, nrow=2, ncol=2, byrow=TRUE, dimnames=list(rnames2, cnames2))

# Fishers exact test using matrix created above
fisher.test(mymatrix, y = NULL, hybrid = FALSE,
            alternative = "two.sided", conf.level = 0.95,
            simulate.p.value = FALSE, B = 2000)

#Compare results with chi-squared test
chisq.test(mymatrix)
```

McNemar's test with Holm correction

```
#McNemar's test to compare algorithms if samples are dependent.
#Define matrix columns, rows and data entries.
rnames2 <- c("Alg1Correct", "Alg1InCorrect")
cnames2 <- c("Alg2Correct", "Alg2InCorrect")
#Define the contingency table: the following line is adjusted based on the performance of each algorithm
cells2 <- c(578,146,244,239)  #by row

#Create the contingency matrix
mymatrix <- matrix(cells2, nrow=2, ncol=2, byrow=TRUE, dimnames=list(rnames2, cnames2))

#Run McNemar's test. Continuity correct should be set to TRUE for small error variances.
mcnemar.test(mymatrix, y = NULL, correct = FALSE)

#Post hoc test: create a vector of p-values, and adjust for family wise error, for example:
p <- c(0.7091, 0.8629, 0.5984, 0.6771,0.309,0.9601, 0.3934, 0.3901, 0,0.8305,0.6773,0.00017,0,0,0.00001)
p.adjust(p, method="holm")
```

D.6 Minority class balancing with SMOTE

```
library(grid)
library(lattice)
library(DMwR)
data<-read.csv("FinalDatasetAfterSetup.csv", header=TRUE, sep=";")
#data<-data[,c("GPA","AverageLC","CAOPoints","English","Maths","AppliedAverage","HumanitiesAverage","MethodicalAverage",
"Conscientiousness","Openness","IntrinsicMotivation","ExtrinsicMotivation","SelfEfficacy","SelfRegulation","StudyEffort","StudyTime",
"gender","GroupWork","age","DeepLearner","ShallowLearner","StrategicLearner","Visual","Auditory","Kinaesthetic")]
balData<-SMOTE(GPA ~ ., data, perc.under=200, k=5, perc.over=100)
write.csv(balData, file="FinalSmoteDataset.csv")
```