JOURNAL OF LEARNING ANALYTICS

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

A Review of Psychometric Data Analysis and Applications in Modelling of Academic Achievement in Tertiary Education

Geraldine Gray, Colm McGuinness, Philip Owende, Aiden Carthy
Institute of Technology Blanchardstown, Ireland

geraldine.gray@itb.ie

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.

KEYWORDS: Learning analytics, educational data mining, psychometrics, classification, academic performance, ability, personality, Big-5, motivation, learning style, self-regulated learning, learning dispositions

1. INTRODUCTION

It is increasingly evident that significant numbers of college students do not complete the courses in which they enrol, particularly courses with lower entry requirements (ACT, 2012; Mooney et al., 2010). Enrolment numbers to tertiary education are increasing, as is the academic and social diversity in the student population (HEA, 2013; OECD, 2013). This adds to the challenge of both identifying students at risk of failing and provisioning the appropriate supports and learning environment to enable all students to perform optimally (Mooney et al., 2010). Tertiary education providers collect an ever-increasing volume of data on their students, particularly activity data from virtual learning environments and other online resources (Drachsler & Greller, 2012). As a result, the application of data analytics to educational settings is emerging as an evolving and growing research discipline (Sachin & Vijay, 2012; Siemens & Baker, 2012), with the primary aim of exploring the value of such data in providing learning

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

professionals, and students, with actionable information that could be used to enhance the learning environment (Siemens, 2012; Chatti et al., 2012). A key challenge for learning analytics is the need to develop capability to explore and identify data that will contribute to improving learning models, including data not currently gathered systematically by tertiary education providers (Buckingham Shum & Deakin Crick, 2012; Tempelaar et al., 2013).

Learning is a latent variable, typically measured as academic performance in assessment work and examinations (Mislevy, Behrens, & Dicerbo, 2012). Factors affecting academic performance have been the focus of research for many years (Farsides & Woodfield, 2003; Lent, Brown, & Hacket, 1994; Moran & Crowley, 1979). It remains an active research topic (Buckingham Shum & Deakin Crick, 2012; Cassidy, 2011; Komarraju, Ramsey & Rinella, 2013), indicating the inherent difficulty in both measurement of learning (Knight, Buckinham Shum, & Littleton, 2013; Tempelaar et al., 2013), and modelling the 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. Demographic data, such as age and gender, have 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, other factors 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, particularly at tertiary level (reviewed in Boekaerts, 2001; Robbins et al., 2004). Research into personality traits, specifically the BIG 5 factors of openness, conscientiousness, extroversion, agreeableness, and neuroticism, and their impact on academic achievement in tertiary education, suggests some personality factors are indicative of potential academic achievement (Chamorro-Premuzic & Furnham, 2004, 2008; De Feyter et al., 2012). For example conscientiousness, which is associated with persistence and self-discipline (Chamorro-Premuzic & Furnham, 2004), is correlated with academic performance, but not with IQ, suggesting conscientiousness may compensate for lower ability (Chamorro-Premuzic & Furnham, 2008). Openness, which is associated with curiosity, can be indicative of a deep learning style (Swanberg & Martinsen, 2010). Learning style (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 & Martinsen, 2010).

This paper reviews a range of psychometric factors that could be used to predict academic performance in tertiary education (section 2). It lays emphasis on factors that can be measured prior to, or during learner enrolment in tertiary education programmes. The unique focus is to facilitate, and inform, early engagement with students potentially at risk of failing (e.g., Arnold & Pistilli, 2012; Lauría et al., 2013). Furthermore, results from learner profiling during student induction can provide useful feedback to the learner on preferred approaches to learning tasks, and development of a personalized learning environment. A review of pertinent data analysis techniques is presented in section 3, with an emphasis

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

on empirical modelling approaches prevalent in educational data mining. Section 4 outlines the benefits of greater collaboration between educational psychology and learning analytics.

2. PSYCHOMETRIC VARIABLES RELEVANT TO ACADEMIC PERFORMANCE

The following discussion of student-centred factors focuses on four key areas: aptitude, temperament, motivation, and learning strategies. These were chosen based on being directly or indirectly related to academic performance and measurable in the early stages after student enrolment. The following sections outline the available evidence on correlations between individual attributes and academic achievement. All studies cited were based on tertiary education.

2.1. Cognitive Ability: How It Is Measured and Its 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, Hezlett, & Ones, 2004). In addition to general cognitive intelligence, there is widespread evidence for a multi-dimensional construct of intelligence comprising of a range of sub-factors (Flanagan & 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, summarized in Table 1.

Cognitive ability tests have been criticized based on 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 developing a causal relationship. In an analysis of a range of IQ studies measuring 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, 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 compare children in Western and non-Western standards of education. These have shown that children tended to score well on tests that measured skills valued by their parents (Sternberg (1999 p. 8). It is notable that correlations between general intelligence and academic performance are stronger at secondary school level than in tertiary level education (Bartels et al., 2002; Cassidy, 2011; Colom & Flores-Mendoza, 2007; Eysenck, 1994; Matarazzo & Goldstein, 1972). Therefore, prior academic

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

performance, such as High School Grade Point Average (HSGPA),[1] and/or standardized tests, such as American College Testing (ACT)[2] scores or Scholastic Aptitude Test (SAT)[3] scores, are frequently used as measures of ability when modelling academic performance in tertiary education.

Table 2 illustrates 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 a moderate correlation between academic performance and SAT scores (r=0.388, 90% CI [0.353, 0.424]) and a marginally higher correlation between academic performance and HSGPA (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 insignificant correlation (r=0.16, 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) 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, n=581) could be attributed to their measure of prior academic performance, which was based on A level[4] scores in just two subjects. The relatively high level of correlation reported by Cassidy (2011) could also 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 in order to predict students' final GPA aggregate.

2.2. Temperament: Definition and Relevance to Academic Performance

Theories of temperament focus on aspects of personality discernible at birth (Boeree, 2006; John et al., 2008). Historically, research that links temperament with academic achievement has lacked a 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 & Schouwenburg, 1996). While many factors are 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 & Heggestad, 1997; Boeree, 2006; John et al., 2008). These are commonly referred to as the Big Five (Cattell & Mead, 2008; Goldberg, 1992, 1993; Tupes & Cristal, 1961) or the related Five-Factor Model (Costa & McCrae, 1992). The five factors — openness, agreeableness, extroversion, conscientiousness, and neuroticism — are described in Table 3.

[1] High School Grade Point Average (HSGPA) is a secondary school, end-of-year, aggregate measure of academic performance, which can be a combination of continuous assessment results and end of term exams.
[2] ACT tests are based on high school curriculum in English, Mathematics, Reading, and Science (www.act.org).
[3] SAT measures general intelligence in addition to maths and verbal subscales (Frey & Detterman, 2003). Frey and Detterman (2003) found SAT scores to be highly correlated with IQ (r=0.82, p<0.001).
[4] Hong Kong's secondary school termination exam. Students can select from a range of subjects.

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

Table 1: CHC's ten broad factors of cognitive ability (McGrew, 2009)

Factor	Symbol	Description
Fluid Intelligence	Gf	Ability to solve problems independently of knowledge learned.
Crystallized Intelligence	Gc	Acquiring and organizing knowledge and skills, and ability to use such knowledge in solving problems.
Visual processing	Gv	Ability to process and analyze visual information.
Auditory Processing	Ga	Ability to process and analyze auditory information.
Processing Speed	Gs	Ability to perform automatic cognitive tasks quickly (measured in minutes).
Reaction Time/ Decision Speed	Gt	Speed at which an individual can react to a stimulus, or make decisions (measured in seconds).
Short-Term Memory	Gsm	Ability to hold information with immediate awareness and reuse within a few seconds.
Long-Term Retrieval	Glr	The ability to store and retrieve information over a longer period.
Quantitative Knowledge	Gq	The 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).

Table 2: Correlations between cognitive ability and academic performance

Study	n	Age	Academic Performance	g	SAT/ACT	Prior ability
Brady-Amoon & Fuertes (2011)	271	m=21.26	GPA			0.16
Cassidy (2011)	97	m=23.5	GPA			0.519**
Chamorro-Premuzic & Furnham (2008)	158	m=19.2	GPA	0.24*		
Conrad (2006)	300	m=19.48	GPA-self-reported		0.28*	
Duff et al. (2004)	146	17-52	GPA			0.274*
Eppler & Harju (1997)	212	m=19.2	GPA		0.37***	
Eppler & Harju (1997)	25	m=29.8	GPA		0.09	
Furnham & Zhang (2006)	64	[20-55]	Mean exam results	0.22		
Kaufman et al. (2008)	315	m=23.5	GPA			0.28
Kobrin et al. (2008)	151,316	18+	GPA		0.35	0.36
Ning & Downing (2010)	581	m=20.24	GPA			0.1*
Robbins et al. (2004)	Meta-analysis		GPA		0.39	0.448

*p<.05, **p<.01, ***p<0.001, m=mean, n=number of participants

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

While the Big Five concept is empirical rather than a theory of personality (Srivastava, 2010), good reliability and consistency has been reported (de Raad & Schouwenburg, 1996; John et al., 2008).

Table 3: Big Five personality factors (McCrae & Costa, 1991; Goldberg, 1992)

Big Five Factor	Traits of high scorers	Traits of low scorers
Extroversion	Enjoys human interaction, cheerful, outgoing.	Cautious, likes to be alone, can lack enthusiasm.
Neuroticism	Temperamental, moody, nervous, finds stress difficult to cope with.	Calm, even-tempered, unafraid.
Openness	Openness to new ideas and imagination, broad minded, tolerant, intellectual curiosity.	Likes routine and familiarity, factually orientated, practical.
Agreeableness	Kind, trusting, warm, unselfish.	Stubborn, rude, uncooperative.
Conscientiousness	Organized, thorough, reliable.	Relaxed, lazy, careless.

2.2.1. Relating Personality to Academic Performance

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

Some significant correlations between openness and academic performance have been reported, but correlations with academic performance are not as high (see Table 4). 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 & Martinsen, 2010). Sub-factors of openness, namely intellectual curiosity, creativity, and open-mindedness, have been associated with effective thinking and learning dispositions (Buckingham Shum & Deakin Crick, 2012; Tishman, Jay & Perkins, 1993). Knight et al. (2013) argues 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 limitations (Hembree, 1988). Where academic performance is measured under less stressful conditions, such as continuous assessment work, the relationship between neuroticism and academic performance is less well-defined

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

(Chamorro-Premuzic & Furnham, 2009, p. 75). Kappe and van der Flier (2010) found neuroticism to be positively correlated with academic performance (r=0.18, p<0.05, n=133) when assessment is free from time constraints and supervision.

Table 4: Correlations between temperament and academic performance

Study	n	Age	Academic Performance	Conscien-tious	Open	Extrovert	Neurotic	Agreeable
Chamorro-Premuzic & Furnham (2008)	158	18-21	GPA	0.37**	0.21**	0.16	-0.05	0.02
Chamorro-Premuzic & Furnham (2003)+	70	17-21	grades	0.33**	-0.06	0.05	-0.28**	0.34**
Conrad (2006)	300	m=19.48	GPA self-reported	0.35*	-0.02	0	-0.6	0.11
Dollinger et al. (2008)	338	m=21.9	GPA	0.26*	0.03	0.02	0.05	0.16*
Duff et al. (2004)	146	17-52	GPA	0.21	0.06	0.06	-0.13	0.115
Gray & Watson (2002)	300	18-21	GPA	0.36*	0.18*	-0.09	0	0.15*
Kappe & van der Flier (2010)++	133	18-22	GPA	0.46**	-0.08	0.05	-0.06	0
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	0.22**
O'Connor & Paunonen (2007)	meta-analysis		various	0.24	0.05	-0.05	-0.03	0.06
Trapmann et al. (2007)	Meta-analysis		GPA	0.216	0.083	0.011	-0.044	0.041

*p<.05, **p<.01, m=mean, n=number of participants

+ Figures based on 1st year exam results. Correlations for agreeableness were lower for 2nd and 3rd year results (0.06 and 0.03 respectively).

++ Matched exam technique to personality type.

+++ Measured Emotional Stability, the reverse of Neuroticism.

Research is inconsistent regarding the remaining two personality dimensions of extroversion and agreeableness and their relationship with academic performance. Introverts tend to have better study habits and are less easily distracted (Entwistle & Entwistle, 1970 as cited in Chamorro-Premuzic &

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

Furnham, 2009, p. 78), while extroverts tend to perform better in class participation, oral exams, seminar presentations, and multiple-choice style questions (Furnham & Medhurst, 1995; Kappe & 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 correlated with absenteeism in first year of study.

2.3. Motivation and Correlations with 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 characterized by a range of complementary theories that aim to explain both the level of individual motivation and the nature of the motivation (Steel & 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 categorization 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.

2.3.1. Expectancy Theory of Motivation

Expectancy models of motivation explore the extent to which a person regards outcome as being a consequence of behaviour. Levels of expectancy motivation are therefore influenced by the extent to which a person believes he or she is in control of the outcome (Cassidy, 2011). There are two strands of expectancy motivation (Eccles & Wigfield, 2002; Pintrich & DeGroot, 1990):

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 5 gives a summary of correlations found between expectancy motivation and academic performance. A meta-analysis of a range of studies recorded correlations varying between 0.38 and 0.5 (Brown et al. 2008). A number of studies identified self-efficacy as a useful predictor of academic performance (Brady-Amoon & Fuertes, 2011; Cassidy, 2011; Yusuf, 2011). Indirect relationships between self-efficacy and academic performance mediated either by other motivational factors or learning strategies are also cited (Breiman, 2001; Yusuf, 2011). On the other hand, Pintrich & 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

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

considered, thereby concluding that self-efficacy facilitates cognitive engagement, but cognitive engagement itself is more directly linked to academic performance.

2.3.2. Goal Theory of Motivation

High self-efficacy is associated with a student setting challenging goals in terms of academic achievement. Such achievement goals fall into two categories: performance goals, where an individual is looking for favourable feedback, and learning goals, where an individual desires to increase competency (Covington, 2000; Dweck, 1986; Dweck & Leggett, 1988; Eccles & Wigfield, 2002; Eppler & 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 & Deakin Crick, 2012). Interestingly, Dweck (1986) 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).

Table 5: Correlations between expectancy motivation and academic performance

Study	n	Age	Academic performance	Self-efficacy	Outcome Expectancy
Brady-Amoon & Fuertes (2011)	271	m=21.26	GPA	0.22*	
Bruinsma (2004)	117	18	Y1 credits	0.26**	
Cassidy (2011)	97	m=23.5	GPA	0.397***	0.195
DiBenedetto & Bembenutty (2013)	113	18+	Module grade	0.37**	0.08
Diseth (2011)	177	m=21.21	Specific exam	0.44**	
Klassen et al. (2008)	261	m=23.3	Self-reported GPA	0.36**	
Komarraju & Nadler (2013)	257	18+	GPA	0.3**	
Robbins et al. (2004)	Meta-analysis		GPA	0.496	

*p<.05, **p<.01, ***p<0.001, m=mean, n=number of participants

Studies have found learning goals to be more strongly correlated with academic performance than performance goals (see Table 6). 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, Diseth

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

(2011) 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.

2.3.3. Self-determination theory (needs-based motivation)

Self-Determination Theory (SDT) focuses on our innate psychological need for competency (Deci & Ryan, 2000) and aims to explore the difference in the types of goals learners adopted, and the justification. SDT distinguishes between intrinsic motivation, where motivation arises from enjoyment of activity, and extrinsic motivation, where the outcome is attractive (Deci & Ryan, 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; Atherton, 2009; Entwhistle, 2005). Alternatively, SDT has been viewed as two separate factors that can both be present (Dweck & Leggett, 1988; Eppler & 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).

Table 6: Correlations between goal orientation and academic performance

Study	n	Age	Academic Performance	Learning goals	Performance goals
Diseth (2011)	177	m=21.2	Specific exam	0.21**	0.39**
Dollinger et al. (2008)+	338	m=21.9	Exam performance	0.21**	
Eppler & Harju (1997)	212	m=19.2	GPA	0.3***	0.13
Eppler & Harju (1997)	50	m=29.8	GPA	0.28*	0.08
Robbins et al. (2004)+	meta-analysis		GPA	0.179	
Wolters (1998)	115	m=19.1	Average grade	0.36***	-0.21*

*p<.05, **p<.01, ***p<0.001, m=mean, n=number of participants
+These studies cited correlations for achievement goals in general, rather than learning or performance goals specifically.

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 7). 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, Karau, and Schmeck (2009) found significant correlation between intrinsic motivation and academic

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

performance in a study of participants from a variety of disciplines. The study included three sub-factors of intrinsic motivation from the Academic Motivations Scale (AMS): motivation to know ($r=0.17$, $p<0.01$), motivation to accomplish ($r=0.22$, $p<0.01$), and motivation to experience stimulation ($r=0.13$, $p<0.05$). 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 and using a 60-item motivation scale, did not find correlations to be significant, suggesting that factors impacting on academic performance can vary for different student groups.

Table 7: Correlations between self-determination and academic performance

Study	n	Age	Academic Performance	Intrinsic motivation	Extrinsic motivation
Bruinsma (2004)	117	m=18	Y1 credits	0.09	
Goodman et al. (2011)	254	[17-29]	GPA	0.281**	0.205**
Kaufman et al. (2008)	315	m=23.5	GPA	0.08	-0.05
Komarraju et al. (2009)	308	18-24	self-reported GPA	0.2**	0.11
Komarraju & Nadler (2013)	257	m=20.48	self-reported GPA	0.11	0.05
Wolters (1998)	115	m=19.1	Average grade	0.14	0.05

*p<.05, **p<.01, ***p<0.001, m=mean, n=number of participants

2.3.1. Impact of motivation on 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 are less consistent, and are influenced to some extent by the selection of factors included in any specific study. For example, Chamorro-Premuzic and Furnham (2003) and Breiman (2001) 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 noted 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 that self-efficacy had a causal relationship with academic performance. In a meta-analysis covering a range of psychosocial and study skills impacting on academic performance at the tertiary level, excluding personality factors, 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

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

changed than ability or personality. This would suggest that while correlations exist between factors of personality and motivation, factors of personality, particularly conscientiousness, and factors of motivation, particularly self-efficacy and achievement goals, each have value, and are worth further consideration in models of student learning.

2.4. Defining learning strategies and their relationship with academic performance

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

2.4.1. Learning style constructs

Many constructs and frameworks exist for learning styles: instructional preference, information processing style, and cognitive personality style (see Coffield et al. (2004) for a detailed review). Approaches to learning have their foundation 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 memorize content regardless of their level of understanding. Later studies added strategic learners as a third category (Entwhistle; 2005, p. 19) whose priority is to do well, and will adopt either a shallow or a 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 & Säljö, 2005).

Table 8: Correlations between learning orientation and academic performance

Study	N	Age	Academic Performance	Deep	Shallow	Strategic
Cassidy (2011)	97	m=23.5	GPA	0.308**	-0.013	0.316**
Chamorro-Premuzic & Furnham (2008)	158	m=19.2	GPA	0.33**	-0.15	0.18*
Duff et al. (2004)	46	m=24.3	GPA	0.097	-0.054	0.153
Snelgrove (2004)	289	18+	GPA	0.20*	-0.13	0.17*
Swanberg & Martinsen (2010)	687	m=24.5	Single exam	0.16	-0.25	

*p<.05, **p<.01, ***p<0.001, m=mean, n=number of participants

Many studies concur with a negative correlation between a shallow learning approach and academic performance (see summary in Table 8). Some studies show higher correlations between academic performance and a deep learning approach (e.g., Chamorro-Premuzic & Furnham, 2008; Snelgrove,

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

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 & Deakin Crick, 2012; Knight et al., 2013), and hence, may elicit secondary, follow-up actions.

2.4.2. Self-regulated learning

Self-regulated learning is recognized as a complex concept that overlaps with a number of other concepts including temperament, learning approach, and motivation, specifically self-efficacy and goal setting (Bidjerano & Dai, 2007; Boekaerts, 1996). While many students may set goals, the ability to self-regulate learning can be the difference between achieving, or not achieving, the goals set (Covington, 2000). Self-regulated learners take responsibility for setting and achieving their own learning goals by planning their learning, having effective time management, using appropriate learning strategies, continually monitoring and evaluating the quality of their own learning, and altering their learning strategies when required (Schunk, 2005; Zimmerman, 1990). Such learners regard learning as a process they can control, but their motivation factors can vary (Pintrich & DeGroot, 1990). To be motivated to self-regulate, a learner must be confident in setting goals and organizing study, and also be confident that study efforts 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 effort (Bembenutty, 2009; Komarraju & Nadler, 2013; Zimmerman, 1990; Zimmerman & Kitsantas, 2005). Volet (1996) argues that self-regulated learning is more significant in the tertiary level than earlier levels of education because of the shift from a teacher-controlled environment to one of self-regulated study.

A number of studies cite significant correlations between academic performance and factors of self-regulation (see Table 9 for a summary). For example, a longitudinal study of first year students ($n=581$) found academic performance to be more strongly correlated with self-testing strategies ($r=0.48$, $p<0.001$) and monitoring levels of understanding ($r=0.42$, $p<0.001$) than effort management ($r=0.24$, $p<0.01$) (Ning & Downing, 2010). Conversely, in a study of undergraduates across all years of study, Komarraju and Nadler (2013) found effort management to have a higher correlation with academic performance ($r=0.39$, $p<0.01$) than other factors of self-regulation. They also found that monitoring and evaluating learning aspects of self-regulation 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 resulted in a deep learning approach, which in turn resulted 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).

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

2.5. Regression Models of Academic Performance Based on Psychometric Variables

Table 10 presents examples of hierarchical regression models that have attempted to explain variance in academic performance. Relatively high levels of model accuracy related to studies that include 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) accounted for 53% of the variance in a regression model including prior academic performance, self-efficacy, and age (n=97). However, the high model accuracy may be due to the measure of prior academic performance used (first year GPA). Chamorro-Premuzic and Furnham (2008) accounted for 40% of the variance in a regression model that included prior academic ability, personality factors, and a deep learning strategy. A similar proportion of variance (44%) was reported by Dollinger et al. (2008) in a regression model including prior academic ability, personality factors, academic goals, and study time. Not all studies concur with these results.

Table 9: Correlations between academic performance and self-regulation

Study	N	Age	Academic Performance	Effort regulation	Time management	Self-regulation
Bidjerano & Dai (2007)	217	m=22	GPA, self-reported	0.23**	0.33**	
Dollinger et al. (2008)	338	m=21.9	exam performance		0.21**	
Goodman et al. (2011)	254	[17-29]	GPA	0.276**		
Komarraju & Nadler (2013)	257	18+	GPA	0.39**	0.31**	0.14*
Ning & Downing (2010)	581	m=20.24	GPA	0.24**		0.42***
Snelgrove (2004)	289	18+	GPA			0.26*
Sundre & Kitsantas (2004)	62	18-24	Single MCQ			0.35**

*p<.05, **p<.01, ***p<0.001, m=mean, n=number of participants

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 accounting for 14% of the variance 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) accounted for 21% of variance 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 ability. Komarraju et al. (2011) accounted for 15% of the variance in a model including personality and learning approach. Eppler and Harju (1997) accounted for 11% of the variance in a model including factors of motivation and work commitments, while Bidjerano and Dai (2007) also accounted for an 11% variance in a model including factors of personality and self-regulation. These results suggest that ability is an important determinant of academic performance, particularly in models of standard students. Authors also found that psychometric variables accounted for additional variance beyond that accounted for by prior academic performance (Cassidy, 2011;

Chamorro-Premuzic & Furnham, 2008; Dollinger et al., 2008; Kaufman et al., 2008; Swanberg & Martinsen, 2010).

3. 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, the studies cited in section 2 primarily used correlation (78% of the studies) and regression (54% of the studies), with some papers citing path analysis results (14%) and structural equation models (11%). 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 additively, 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, overlaps occur in the constructs being measured. In addition, there is evidence to suggest variance is not constant for all attributes. For example, De Feyter et al. (2012) found that low levels of self-efficacy had a positive, direct effect on academic performance for neurotic students only, and for stable students, average or higher levels of self-efficacy had a direct effect on academic performance. In addition, 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 recognizing the continuing importance of statistical models, Freedman (1987) and Breiman (2001) argued that alternative-modelling 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 approach is better (Breiman, 2001, p. 18). It is therefore pertinent to ask if data mining's empirical modelling approach can add value to psychometric 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, visualization, 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 & 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).

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

Table 10: Regression models with beta values for significant attributes.

				Ability		Personality				Motivation				Learning Style			Self-regulation		Other factors		
Study	**n**	**Age+**	**Variance**	**g**	**prior**	**C**	**O**	**N**	**A**	**SE**	**AG**	**IM**	**AM**	**De**	**Sh**	**St**	**Effort**	**Time**	**Age**	**Gender**	**Job**
Bidjerano & Dai (2007)	217	22	18%	0.28													0.27				
Bidjerano & Dai (2007)	217	22	11%			0.14											0.31				
Chamorro-Premuzic & Furnham (2008)	158	19.2	40%	0.29		0.49								0.21							
Cassidy (2011)	97	23.50	53%		0.54					0.26									0.36		
Dollinger et al. (2008)	338	21.9	43%	0.44	0.32													0.21			-0.1
Duff et al. (2004)	146	[17-52]	34%		0.39	0.37													0.3		
Eppler & Harju (1997)	262	21.2	21.5%	0.3								0.34									-0.14
Eppler & Harju (1997)	262	21.2	11%									0.32									-0.16
Kaufman et al. (2008)	315	25.9	14%		0.24	0.12						0.15	0.16								
Komarraju et al. (2011)	308	[18-24]	15%			0.33	0.14	0.19	0.15												
Swanberg & Martinsen (2010)	687	24.5	21%		0.3										0.15	-0.17				-0.14	

g=general cognitive intelligence; C=Conscientiousness; O=Openness; N=Neurotic; A=Agreeableness; SE=self-efficacy; AG=academic goals; IM=Intrinsic Motivation; AM=Achievement Motivation; De=Deep; Sh=Shallow; St=Strategic.

+ Mean age, except where a range of ages is given.

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

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

3.1. Classification Algorithms Used on Educational Data

A Decision Tree (DT) 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 "grade=good": "*if Conscientiousness > 5.6 and Self-Efficacy > 6.3 then Grade = Good*." 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., 2006). 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 & Kamber, 2006).

Rule-based classifiers define class membership based on a set of *if…then…* rules. Basic implementations generate models similar to a decision tree model (Tan et al., 2006) despite the difference in search strategies used. Rule-based classifiers implement a depth first search; decision trees implement a breath first search (Gupta & 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 "*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 & Sánchez, 2005; Tang et al., 2012).

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 & 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., 2006), and can converge to its optimal accuracy quickly, making it suitable for relatively small datasets (Ng & 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 & 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

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

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 Neural Network (NN) 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 & 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., 2006). Nonetheless, NNs performance has been found to be comparable with other statistical approaches, particularly when approximating complex patterns based on numeric input values (Sargent, 2001; Groth, 2000).

A Support Vector Machine (SVM) models class membership by approximating a hyperplane that defines a linear boundary between two classes (Cortes & 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 optimization problem to which a globally optimal solution can be found (Tan et al., 2006). While SVMs are limited to numeric attributes and binary classification tasks, Dixon and Brereton (2009) found SVMs outperformed other learners when modelling datasets that are not normally distributed.

k-Nearest Neighbour (*k*-NN) uses instances from the original dataset to classify a new row of data, and so works with the full dataset rather than a generalized model (Tan et al., 2006; Cover & Hart, 1967). For example, a student would be classified according to the class membership of the *k* rows in a dataset most similar to the characteristics of that student, where *k* is a configurable parameter determining neighbourhood size. Decisions made are local, and decision boundaries can be irregular in shape, making k-NN suitable to datasets not easily generalizable because of pattern complexity (Tan et al., 2006).

Ensembles aggregate the predictions of a collection of classification models (Breiman, 1996; Banfield et al., 2004). 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 is 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., 2006, 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., 2006).

3.1.1. Review of Model Performance

Table 11 summarizes a selection of educational data mining studies, the algorithms used, and the

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

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 psychometric factors, 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 interaction, or Intelligent Tutoring Systems (ITS) aimed towards curriculum adaptation to each learner by monitoring progress and measuring skill levels (Baker & 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 co-linearity 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 DT, logistic regression and SVM, Lauría et al. (2013) reports comparable performance when modelling prior academic performance, demographic data, and ITS usage data. Gray et al. (2013) agreed that model performance was comparable when modelling students as a single group, but found models capable of representing complex patterns (SVM, NN and k-NN) outperformed other models (DT, logistic regression, Naïve Bayes) when modelling subgroups split by age. Bergin (2006) achieved good accuracy with Naïve Bayes when modelling a small dataset of prior academic performance and psychometric data, 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.

Table 11: Data Mining models for predicting academic performance in tertiary education

Study	Algorithm	Accuracy	n	Class label	Demographic Data	Prior Education	Psychometric data	ITS
Bergin (2006)	Ensemble (stackingC)	82%	102	weak/strong		x	x	
Gray et al., (2013)	SVM	82%	636	weak/strong		x	x	
Herzog (2006)	Decision Tree (C5.0)	83%	4564	degree completion time	x	x		
Dekker et al. (2009)	Decision Tree (J48)	79%	1002	drop out		x		
Lauría et al. (2013)	Decision Tree	87%	6445	weak/strong	x	x		x

Study	Algorithm	Accuracy	n	Class label	VLE	ITS
Baker et al. (2011)	Bayesian Network (BKT)	AUC: 0.7029	76	next question correct		x
Merceron & Yacef (2005)	Decision Tree (C4.5)	87%	224	pass/fail		x
Minaei-Bidgoli et al. (2003)	Ensemble	94%	227	pass/fail		x
Pardos et al. (2011)	Ensemble (Neural Networks)	AUC:0.77	5,422	Performance on ITS		x
Romero et al. (2008)	Fuzzy Rule (MaxLogit-Boost)	62%	438	module performance 4 bins	x	

n=number of instances; AUC=Area under the Curve

4. BENEFITS OF GREATER COLLABORATION BETWEEN EDUCATIONAL PSYCHOLOGY AND LEARNING ANALYTICS

Notably for this literary review, a limited number of educational data mining studies have investigated the role of psychometric factors in models of learning (Buckingham Shum & Deakin Crick, 2012; Shute & Ventura, 2013). Bergin (2006) found that adding self-efficacy and study hours improved model accuracy, but due to the small sample size (n=58) could not draw reliable conclusions from the findings. Lauría et al. (2012) also achieved good model accuracy when modelling psychometric data with prior academic performance and other demographic attributes. Gray et al. (2013) suggested that while good accuracies can be achieved without the addition of psychometric data, the inclusion of psychometric data could offer better insights into factors influencing academic performance. In addition, including psychometric data in models of learning can provide useful feedback on the learning dispositions that assessment design rewards (Nelson et al., 2012). Buckingham Shum and Deakin Crick (2012) argues 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) argues 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 psychometric data to capture learner disposition and approaches to learning task. Interestingly, evidence from Shute and Ventura (2013) suggests some learner dispositions can be inferred from their online behaviour (e.g., persistence and creativity).

Learning analytics can offer benefits over and above traditional data analysis methods prevalent in the social sciences, including a greater range of modelling approaches, scalability, analysis of relevant trace data and a quick feedback cycle. Studies cited above suggest data mining algorithms can offer additional insights over and above standard statistical modelling (e.g., Herzog, 2006). In addition, increased use of technology has resulted in a wealth of digital trails generated by learners, providing large volumes of trace data collected during the learning process (Knight et al., 2013). Many data mining algorithms have implementations adapted for this big-data environment, for example, Decision Tree (Ben-Haim & Tom-Tov, 2010), k-NN (Liang et al., 2009), Neural Networks (Gu et al., 2013), SVM and regression (Luo et al., 2012), and supporting tools are available (Prekopcsák et al., 2011), facilitating quick analysis and feedback (Siemens & Long, 2011). Recent developments in learning analytics frameworks (e.g., the learning warehouse, Buckingham Shum & Deakin Crick, 2012) illustrate the potential for learning analytics to support automation of the full life cycle from data gathering through to deployment of recommendations and interventions based on analysis results.

5. CONCLUSION

This review has collated evidence on the importance of psychometric factors in the modelling of academic achievement in tertiary level education. While not accounting for all of the variance in the noted academic performance, learner ability, personality, motivation, and self-regulation have significant relationships with academic performance, and overlap with noteworthy learning dispositions. Since such attributes can be measured prior to student engagement in course work, they facilitate early recognition of learners at risk of failing, inform appropriate interventions, and provide early input to personalized learning environments. Prior academic performance is a good predictor of academic performance for standard students, but it does not perform as well for mature learners or learner groups with ethnic diversity. Conscientiousness is also a strong personality-based predictor of academic

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

performance, while self-efficacy is the best motivation-based predictor of academic performance. Self-regulation, particularly study time and study effort, are also significant. On these bases, there has been extensive work done by educational psychologists on the evaluation of psychometric predictors of academic performance using parametric models. However, there is evidence that datasets that include psychometric variables are complex in terms of redundancy and non-linearity of relationships, and therefore could be suited to the empirical modelling approaches used in data mining.

To date, the complementary disciplines of learning analytics and educational data mining have focused predominantly on analyzing data systematically gathered in educational settings, which at the tertiary level includes factors of prior academic performance, demographic data, such as age and gender, and data gathered by logs recording student behaviour in online learning environments. Though both are relatively new disciplines, initial results are encouraging across a variety of analysis techniques. However, there is scope for more research investigating the contribution of additional data that could be gathered by tertiary education providers, as well as how this data should be modelled to enhance current student models, and offer actionable feedback on the learning process. Further work is needed to determine if greater inclusion of psychometric data in algorithmic models of student learning can add value to the knowledge learned from these models.

REFERENCES

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

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

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

Apter, M.J. (1989). *Reversal theory: Motivation, emotion and personality*. London: Routledge.

Arnold, K.E., & Pistilli, M.D. (2012). Course signals at Purdue: Using learning analytics to increase student success. *Second International Conference on Learning Analytics and Knowledge*, ACM, Vancouver, British Columbia, Canada.

Atherton, J.S. (2009). Learning and teaching: Motivation. Retrieved from http://www.learningandteaching.info/learning/motivation.htm#Levels%20of%20Motivation

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

Baker, R.S.J.d., Pardos, Z.A., Gowda, S.M., Nooraei, B.B., & Heffernan, N.T. (2011). Ensembling predictions of student knowledge within intelligent tutoring systems. *Proceedings of Nineteenth International Conference on User Modeling, Adaptation, and Personalization,* Girona, Spain, 13–24.

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

Banfield, R.E., Hall, L.O., Bowyer, K.W., & Bhadoria, D. (2004). A comparison of ensemble creation techniques. *Multiple Classifier Systems, Lecture Notes in Computer Science,* Springer, Berlin, Heidelberg, 223–232.

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

Bayer, J., Bydzovská, H., Géryk, J., Obsvac, T., & Popelnsky, L. (2012). Predicting drop-out from social behaviour of students. *Proceedings of the Fifth International Conference on Educational Data Mining,* Chania, Greece, 103–109.

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

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

Ben-Haim, Y., & Tom-Tov, E. (2010). A streaming parallel decision tree algorithm. *Journal of Machine Learning Research, 11*, 849–872.

Bergin, S. (2006). Statistical and machine learning models to predict programming performance, PhD thesis, Computer Science, National University of Ireland, Maynooth.

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

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

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

Boekaerts, M. (2001). Bringing about change in the classroom: Strengths and weaknesses of the self-regulated learning approach. *EARLI presidential address*, Centre for the Study of Education and Instruction, Leiden 2300 RB, The Netherlands.

Boeree, G. (2006). Personality theories. Retrieved from http://webspace.ship.edu/cgboer/perscontents.html

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

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

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

Breiman, L., Friedman, J., Stone, C.J., & Olshen, R.A. (1984). *Classification and regression trees*. Boca Raton, FL: Chapman and Hall/CRC.

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

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

Buckingham Shum, S., & Deakin Crick, R. (2012). Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics. *Second International Conference on Learning Analytics and Knowledge*, Vancouver, British Columbia, Canada.

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

Cassidy, S. (2011). Exploring individual differences as determining factors in student academic achievement in higher education. *Studies in Higher Education, 37*(7), 1–18.

Cattell, H.E.P., & Mead, A.D. (2008). The Sixteen Personality Factor Questionnaire (16PF). In *The SAGE handbook of personality theory and assessment*, Vol. 2, *Personality measurement and testing* (chapter 7). Thousand Oaks, CA: SAGE Publications Ltd.

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

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

Chamorro-Premuzic, T., & Furnham, A. (2008). Personality, intelligence and approaches to learning as predictors of academic performance. *Personality and Individual Differences, 44*, 1596–1603.

Chamorro-Premuzic, T., & Furnham, A. (2006). *Personality and intellectual competence*, Sussex UK: Psychology Press, 1st ed., 68–92.

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

Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). Should we be using learning styles? What research has to say to practice. Learning and Skills Research Centre, UK, www.LSRC.ac.uk

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

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

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

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

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

Cover, T.M., & Hart, P.E. (1967). Nearest neighbour pattern classification. *IEEE Transactions on Information Theory, 13*(1).

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

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

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

De Clercq, M., Galand, B., & Frenay, M. (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.

De Feyter, T., Caers, R., Vigna, C., & Berings, D. (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.

de Raad, B., & Schouwenburg, H.C. (1996). Personality in learning and education: A review. *European Journal of Personality, 10*, 303–336.

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

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

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

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.

Dixon, S., & Brereton, R. (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.

Dollinger, S.J., Matyja, A.M., & Huber, J.L. (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.

Drachsler, H., & Greller, W. (2012). The pulse of learning analytics: Understandings and expectations from the stakeholders. *Second International Conference on Learning Analytics and Knowledge*, ACM, Vancouver, British Columbia, Canada.

Duff, A., Boyle, E., Dunleavy, K., & Ferguson, J. (2004). The relationship between personality, approach to learning and academic performance. *Personality and Individual Differences, 36*, 1907–1920.

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

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

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

Entwhistle, N. (2005). Contrasting Perspectives in Learning. In *The Experience of Learning*, chapter 1. Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment. Retrieved from http://www.tla.ed.ac.uk/resources/EoL.html

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

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

Eppler, M.A., & Harju, B.L. (1997). Achievement motivation goals in relation to academic performance in traditional and non-traditional college students. *Research in Higher Education, 38*(5), 557–573.

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

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

Flanagan, D.P., & McGrew, K.S. (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, 36*(2), 151–182.

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

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

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

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

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

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

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

Goodman, S., Jaffer, T., Keresztesi, M., Mamdani, F., Mokgatle, D., Musariri, M., Pires, J., & Schlechter, A. (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.

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

Gray, G., McGuinness, C., & Owende, P. (2013). An investigation of psychometric measures for modelling academic performance in tertiary education. *Sixth International Conference on Educational Data Mining,* Memphis, Tennessee, 240–243.

Groth, R. (2000). *Data mining: Building competitive advantage.* Upper Saddle River, NJ: Prentice Hall

Gu, R., Shen, F., & Huang, Y. (2013). A parallel computing platform for training large scale neural networks. *IEEE International Conference on Big Data,* Santa Clara, California, USA, 376–384.

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

Han, J., & Kamber, M. (2006). *Data mining concepts and techniques*. Burlington, MA: Morgan Kaufmann.

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

HEA (2013). *Higher education key facts and figures 2011–12*. Retrieved from http://www.hea.ie/en/Publications

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

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

Jensen, A.R. (1981). *Straight talk about mental tests*. Glencoe, IL: The Free Press (MacMillan).

John, O.P., Naumann, L.P., & Soto, C.J. (2008). Paradigm shift to the integrative big five trait taxonomy: History, measurement and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114–158). New York, NY: Guildford Press.

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

Kappe, R., & van der Flier, H. (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.

Kaufman, J.C., Agars, M.D., & Lopez-Wagner, M.C. (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.

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

Knight, S., Buckingham Shum, S., & Littleton, K. (2013). Epistemology, pedagogy, assessment and learning analytics. *Third Conference on Learning Analytics and Knowledge* (LAK 2013). Leuven, Belgium.

Kobrin, J., Patterson, B.F., Shaw, E.J., Mattern, K.D., & Barbuti, S.M. (2008). Validity of the SAT for predicting first year college grade point average, Research Report. *College Board New York*, 2008–2005.

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

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

Komarraju, M., & Nadler, D. (2013). Self-efficacy and academic achievement: Why do implicit beliefs, goals, and effort regulation matter? *Learning and Individual Differences, 25,* 67–75.

Komarraju, M., Ramsey, A., & Rinella, V. (2013). Cognitive and non-cognitive predictors of college readiness and performance: Role academic discipline. *Learning and Individual Differences, 24,* 103–109.

Kuncel, N.R., Hezlett, S.A., & Ones, D.S. (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.

Larose, D.T. (2005). *Discovering knowledge in data: An introduction to data mining*. Hoboken, NJ: John Wiley and Sons, Inc.

Lauría, E.J.M., Baron, J.D., & Devireddy, M. (2012). Mining academic data to improve college student retention: An open source perspective. *Second International Conference on Learning Analytics and Knowledge* (LAK 2012). ACM, Vancouver, British Columbia, Canada.

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

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

Liang, S., Liu, Y., Wang, C., & Jian, L. (2009). A CUDA-based parallel implementation of k-nearest neighbor algorithm. *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery.* China, 53–60.

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

Luo, D., Ding, C., & Huang, H. (2012). Parallelization with multiplicative algorithms for big data mining. *IEEE 12th International Conference on Data Mining,* Brussels, Belgium, 489–498.

Marton, F., & Säljö, R. (2005). Approaches to Learning. In *The Experience of Learning*, chapter 3. Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment. Retrieved from http://www.tla.ed.ac.uk/resources/EoL.html

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

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

McCrae, R.R., & Costa, P.T.J. (1991). The NEO personality inventory: Using the five-factor model in counseling. *Journal of Counseling and Development, 69*(4), 367–372.

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.

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

Minaei-Bidgoli, B., Kashy, D.A., Kortemeyer, G., & Punch, W.F. (2003). Predicting student performance: An application of data mining methods with educational web-based system lon-capa. *Proceedings of the 33rd ASEE/IEEE Frontiers in Education Conference,* Boulder, Colorado, USA.

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

Mooney, O., Patterson, V., O'Connor, M., & Chantler, A. (2010). A study of progression in higher education: A report by the higher education authority, Technical report, Higher Education Authority, Ireland.

Moran, M.A., & Crowley, M.J. (1979). The leaving certificate and first year university performance. *Journal of Statistical and Social Enquiry in Ireland, 24*(1), 231–266.

Munzert, A.W. (1980). *Test Your IQ*. New York, NY: MacMillan.

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

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

Nelson B., Nugent R., & Rupp A. (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.

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

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

Nisbet, R., Elder, J., & Miner, G. (2009). *Statistical analysis and data mining applications*. Waltham, MA: Academic Press.

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

OECD (2013). Education at a glance 2013. Retrieved from http://www.oecd.org/edu/eag2013%20(eng)--FINAL%2020%20June%202013.pdf

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

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

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

Pintrich, P., Smith, D., Garcia, T., & McKeachie, W. (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.

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

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

(2014). A Review of Psychometric Data Analysis and Applications. *Journal of Learning Analytics*, *1*(1), 75–106.

Prekopcsák, Z., Makrai, G., Kenk, T., & Gáspár-Papanek, C. (2011). Radoop: Analysing big data with rapidminer and hadoop. *Second Rapidminer Community Meeting and Conference* (RCOMM), Dublin, Ireland.

Prinsloo, P., Slade, S., & Galpin, F. (2012). Learning analytics: Challenges, paradoxes and opportunities for mega open distance learning institutions. *Second International Conference on Learning Analytics and Knowledge*. ACM, Vancouver, British Columbia, Canada.

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

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

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

Romero, C., Ventura, S., Espejo, P.G., & Hervás, C. (2008). Data mining algorithms to classify students. *Proceedings of the First International Conference on Educational Data Mining,* Montreal, Canada, 8–17.

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

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

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

Schmitt, N., Oswald, F.L., Pleskac, T., Sinha, R., & Zorzie, M. (2009). 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), 1479–1497.

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

Shute, V., & Ventura, M. (2013). Stealth assessment: Measuring and supporting learning in video games. The John D. and Catherine T. MacArthur Foundation reports on digital media and learning.

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

Siemens, G., & Baker, R.S.J.d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the Second International Conference on Learning Analytics and Knowledge,* Vancouver, Canada*,* 252–254.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 46*(5), 30–23.

Sins, P.H.M., van Joolingen, W.R., Savelsbergh, E.R., & van Hout-Wolters, B. (2008). Motivation and performance within a collaborative computer-based modeling task: Relations between students'

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

achievement goal orientation, self-efficacy, cognitive processing, and achievement. *Contemporary Educational Psychology, 33*, 58–77.

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

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

Spearman, C. (1927). *The abilities of man*. London: MacMillan.

Srivastava, S. (2010). Measuring the big five personality factors. Retrieved from http://pages.uoregon.edu/sanjay/bigfive.html

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

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

Sundre, D.L., & Kitsantas, A. (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.

Swanberg, A.B., & Martinsen, Ø.L. (2010). Personality, approaches to learning and achievement. *Educational Psychology, 30*(1), 75–88.

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Boston, MA: Pearson Education.

Tang, M., Chen, X., Hu, W., & Yu, W. (2012). Generation of a probabilistic fuzzy rule base by learning from examples. *Information Sciences, 271,* 21–30.

Tempelaar, D.T., Cuypers, H., van de Vrie, E., Heck, A., & van der Kooij, H. (2013). Formative assessment and learning analytics. In E.D. Dan Suthers, Katrien Verbert, E.D. Xavier Ochoa Dan Suthers, Katrien Verbert, D. Xavier Ochoa Suthers, K. Verbert, E. Duval, & X. Ochoa (Eds.), *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (LAK ’13). ACM, New York, NY.

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

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

Tupes, E.C., & Cristal, R.E. (1961). Recurrent personality factors based on trait ratings (Report No AD 267 778), United States Air Force, Lackland, Texas.

Vancouver, J.B., & Kendall, L.N. (2006). When self-efficacy negatively relates to motivation and performance in a learning context. *Journal of Applied Psychology, 91*(5), 1146–1153.

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.

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

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

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

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

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

Zuffianó, A., Alessandri, G., Gerbino, M., Kanacri, B.P.L., Di Giunta, L., Milioni, M., & Caprara, G.V. (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.