From Learning to e-Learning: Mining educational data.

A novel, data-driven approach to evaluate individual differences in students' interaction with learning technology

LORENZO VIGENTINI

Thesis submitted for the degree of PhD in Psychology.
The University of Edinburgh
September 2010.

"Every day you may make progress. Every step may be fruitful. Yet there will stretch out before you an ever-lengthening, ever-ascending, ever-improving path. You know you will never get to the end of the journey. But this, so far from discouraging, only adds to the joy and glory of the climb."

Sir Winston Churchill

Declaration

I hereby declare that this thesis is my own work and has not been submitted for any other degree or professional qualification.

Lorenzo Vigentini   Date. 25/09/2010

Acknowledgements

It took some time, but I finally made it! This thesis is dedicated to my grandparents, who, despite all my misadventures, always believed in me, and to the memory of McG (Dr. Brendan McGonigle) without whom I would have not been able to even start this project.
I am sure that if he was still around this thesis would have gone in a very different direction.

I would like to thank Dr. Hamish Macleod and Dr. Peter Caryl who bravely took over my supervision half way through and provided very valuable guidance and support over the years: without your direction and pragmatism I would be still working on the thesis!

Special thanks go to Sneh Jaswal for useful comments on an early draft and particularly to ‘Roberto’ Higgins who patiently helped with editing and correcting the final draft fixing my *scotalian*.

Abstract

In recent years, learning technology has become a very important addition to the *toolkit* of instructors at any level of education and training. Not only offered as a substitute in distance education, but often complementing traditional delivery methods, e-learning is considered an important component of modern pedagogy. Particularly in the last decade, learning technology has seen a very rapid growth following the large-scale development and deployment of e-learning financed by both Governments and commercial enterprises. These turned e-learning into one of the most profitable sectors of the new century, especially in recession times when education and re-training have become even more important and a need to maximise resources is forced by the need for savings.

Interestingly, however, evaluation of e-learning has been primarily based on the consideration of users' satisfaction and usability metrics (i.e. system engineering perspective) or on the *outcomes* of learning (i.e. gains in grades/task performance). Both of these are too narrow to provide a reliable effect of the *real* impact of learning technology on the learning processes and lead to inconsistent findings.

The key purpose of this thesis is to propose a novel, data-driven framework and methodology to understand the effect of e-learning by evaluating the *utility* and *effectiveness* of e-learning systems in the context of higher education, and specifically, in the teaching of psychology courses. The concept of learning is limited to its relevance for students' learning in courses taught using a mixture of traditional methods and online tools tailored to enhance teaching. The scope of e-learning is intended in a *blended* method of delivery of teaching.

A large sample of over 2000 students taking psychology courses in year 1 and year 2 was considered over a span of 5 five years, also providing the scope for the analysis of some longitudinal sub-samples.

The analysis is accomplished using a psychologically grounded approach to evaluation, partially informed by a cognitive/ behavioural perspective (online usage) and a differential perspective (measures of cognitive and learning styles). Relations between behaviours, *styles* and academic performance are also considered, giving an insight and a direct comparison with existing literature.

The methodology adopted draws heavily from data mining techniques to provide a rich characterisation of students/users in this particular context from the combination of three types of metrics: cognitive and learning styles, online usage and academic performance.
Four different instruments are used to characterise styles: ASSIST (Approaches to learning, Entwistle), CSI (Cognitive Styles Inventory, Allinson & Hayes), TSI (Thinking Styles Inventory and the mental self-government theory, Sternberg) and VICS-WA (Verbal/Imager and Wholistc/Analytic Cognitive style, Riding, Peterson) which were intentionally selected to provide a varied set of tools.
Online usage, spanning over the entire academic year for each student, is analysed applying web usage mining (WUM) techniques and is observed through different layers of interpretation accounting for behaviours from the single clicks to a student's intentions in a single session.
Academic performance was collated from the students' records giving an insight in the end-of-year grades, but also into specific coursework submissions during the whole academic year allowing for a temporal matching of online use and assessment.

The varied metrics used and data mining techniques applied provide a novel evaluation framework based on a rich profile of the learner, which in turn offers a valuable alternative to regression methods as a mean to interpret relations between metrics. Patterns emerging from styles and the way online material is used over time, proved to be valuable in discriminating differences in academic performance and useful in this context to identify significant group differences in both usage and academic performance.
As a result, the understanding of the relations between e-learning usage, styles and academic performance has important practical implications to enhance students' learning experience, in the automation of learning systems and to inform policymakers of the effects of learning technology has from a user and learner-centred approach to learning and studying.

The success of the application of data mining methods offers an excellent starting point to explore further a data-driven approach to evaluation, support informed design processes of e-learning and to deliver suitable interventions to ensure better learning outcomes and provide an efficient system for institutions and organization to maximise the impact of learning technology for teaching and training.

Table of Contents

List of Figures

List of Tables

Chapter 1. A context for this research

The principal aim of this thesis is to explore the utility and effectiveness of *e-learning* and *learning technology* in the context of higher education, specifically in the teaching of psychology courses. Even though these two terms are often used interchangeably, it should be noted from the start that the latter is intended with a focus on the medium, with enabling characteristics, whilst the former should be intended with a focus on the learner stressing the processes involved in interacting with learning technology.

The analysis and evaluation will be accomplished using a psychologically grounded approach to evaluation, partially informed by a cognitive, behavioural and differential perspective. Elements of human-computer interaction and web usage mining will also be drawn into the picture; these provide a contextual ground to criticise the surface approach to evaluation often taken in the current literature when discussing the benefits of IT and learning technology in education.

Furthermore, we will attempt to construct an evaluation framework based on a rich profile of the learner: projected with a wider reach, this could be useful to enhance the learning experience, in the automation of learning systems and to inform policymakers of the effects of learning technology from a user and learner-centred approach to learning and studying.

Three core themes justify further research in this area:

- The pervasiveness of technology and its place in education
- The *massification* of education
- The emerging issues surrounding the "*digital natives* debate".

These themes contributed to pose fundamental questions which we will introduce in this chapter and attempt to argue throughout this thesis. In particular questions such as 'does learning technology work?' (If it does) 'what type of e-learning works, in what contexts, how can it be improved and why may it fail?' Given the controversy generated by the massive investments in e-learning, a crucial question one ought to ask is whether the investments justify the outcomes. Such questions are proposed as a broad sketch of the themes touched hereafter.

In this chapter we will examine the three core aspects mentioned above as precursors of the research conducted. In particular, we will argue that e-learning has became one of the cornerstones of educational strategies because the landscape of higher education (HE) has dramatically changed, partially driven by governments' policies, and it is struggling to keep pace with the infiltration of technology which is a pervasive presence in society. We will argue that the reasons why e-learning is offered as a panacea to solve HE problems are often too general and superficial. The discussion of these issues and a brief contextualization of the case of the University of Edinburgh will lead to a more detailed examination of learning technology and pedagogy in relation to students' learning in the next chapter.

1.1. Pervasiveness of technology and its downsides

Technology is so woven into the fabric of modern life and society that it has become all but invisible. People look at it without seeing it, but we often notice when this is not available: the availability of communication tools and accessibility to media is taken for granted and we most certainly developed an over dependence on technology, information and gadgetry. This is particularly the case for richer countries, all of which started to include internet access as a key point of their inclusion policies and often extend it as a *right* of each citizen (recent examples are the revised charter of rights in the UK or Spain).

Since the introduction of microcomputers in the late 70s, the pace of development and technological innovation has increased exponentially. Looking at the timeline since the introduction of written word in print and the introduction of new communication tools in the last century with radio, telephone, TV and the internet, generational changes became very dense in the last 100 years, with multiple major changes often occurring within one's lifetime.

Advances in technology and miniaturization dramatically helped to increase the pace, and IT and computers have become a center piece in everyone's life driving our communications and interaction in business and private life alike for conducting transactions or exchanging information with colleagues, family and friends.

As early as 1984, Susan Hellweg (Hellweg, 1984) conducted a survey of ninety-four Fortune 500 corporations to identify the extent of the pervasiveness and perceived impact of five electronic communication technologies (electronic mail, videotex, interactive computers, video teleconferencing, and word processing). She was able to determine that word processing was the most pervasive technology used (with secretaries as primary users), followed, in order, by interactive computers, electronic mail, videotex, and video teleconferencing (mostly used by managers and technical staff). Slight to moderate changes were identified in employee productivity, workload capacity and job satisfaction. However, it was found that whilst word processing has made information processing faster in offices, it has not made those offices less reliant upon written communication. Furthermore the findings revealed that there was some apprehension among employees in the offices surveyed about the rapid emergence of the technologies, specifically in terms of having to learn how to use them and of coping with the changes involved. This is a fundamental aspect as dramatic changes in working practices depend on the ability of the workforce to be flexible and ready to embrace changes without too much additional training.
Hence, *computer literacy*, intended as the knowledge and ability to use computers and technology efficiently has become an essential asset. As companies become ever more dependent on technology, the value a potential employee has may be measured in terms of his or her technological competency. The highest goal of a computer-literate person is to be able to learn and use new computer programs without large amounts of help. Computer literacy gives a person of any age an edge in both their career and education.

Certainly since this 1984 report, governments around the world secured massive investments attempting to promote computer literacy both in education and supporting further training in the industry. The scientific establishment also addressed the questions of what and how should be taught when teaching computers. A policy report commissioned and published by the National Research Council in 1999 and authored by a mix of computer scientists and education scholars discussed aspects of literacy and computer literacy.
The authors used the term *fluency* rather than *literacy*, giving the following brief explanation:

> "Generally, computer literacy has acquired a skills connotation, implying competency with a few of today's computer applications, such as word processing and email. Literacy is too modest a goal in the presence of rapid change, because it lacks the necessary staying power. As the technology changes by leaps and bounds, existing

skills become antiquated and there is no migration path to new skills […] To adapt to changes in the technology […] involves learning sufficient foundational material to enable one to acquire new skills independently after one's formal education is complete. This requirement of a deeper understanding than is implied by the rudimentary term computer literacy motivated the committee to adopt fluency as a term connoting a higher level of competency." (NRC, 1999, p. 2)

The authors also decided not to argue that information technology fluency was required of everyone, but to focus specifically on college graduates, defined as "[…] individuals who want to be able to use information technology effectively." (NRC, 1999 p. viii). The most interesting aspect of this report -with no bibliography- was the very detailed specification of the knowledge and abilities required to achieve *fluency with information technology*. Table 1.1, reproduced from the report, enlists all the components of contemporary skills, foundational concepts, and intellectual capabilities as proposed (NRC, 1999p 4).

In her review of 20 years of literacy research, Kate Williams (Williams, 2003) argued that the position of the NRC document was quite limited and suggested a broader set of these *competencies* (11) of which only 3 were addressed by the NRC report (Literacy as a technical skill, literacy as a conceptual, historical and social entity). The discussion of all aspects of literacy (and computer literacy) is well beyond the aim of this thesis, but it is necessary to mention the lively debate in this area, in which the research conducted implicitly relies on some basic assumptions about computer literacy and fluency, some of which will be addressed in the next section.

Intellectual capabilities	**Information technology concepts**	**Information technology skills**
1. Engage in sustained reasoning 2. Manage complexity 3. Test a solution 4. Manage problems in faulty solutions 5. Organize and navigate information structures and evaluate information 6. Collaborate 7. Communicate to other audiences 8. Expect the unexpected 9. Anticipate changing technologies 10. Think about information technology abstractly	1. Computers 2. Information systems 3. Networks 4. Digital representations of information 5. Information organization 6. Modeling and abstraction 7. Algorithmic thinking and programming 8. Universality 9. Limitations of information technology 10. Societal impact of information and information technology	1. Setting up a personal computer 2. Using basic operating system features 3. Using a word processor to create a text document 4. Using a graphics and/or artwork package to create illustrations, slides, or other image-based expressions of ideas 5. Connecting a computer to a network 6. Using the Internet to find information and resources 7. Using a computer to communicate with others 8. Using a spreadsheet to model simple processes or financial tables 9. Using a database system to set up and access useful information 10. Using instructional materials to learn how to use new applications or features

Table 1.1. Components of fluency with information technology (NRC 1999 p 4).

If we narrow the focus to the UK, the Government invested considerable resources to improve access and encourage all citizens to achieve basic IT literacy. The outocme is evident in a recent Omnibus Survey (2006) which characterized the way people use IT, particularly the Internet. On average 6 in 10 adults in the UK use the internet daily and more than half use the internet as the main information resource.

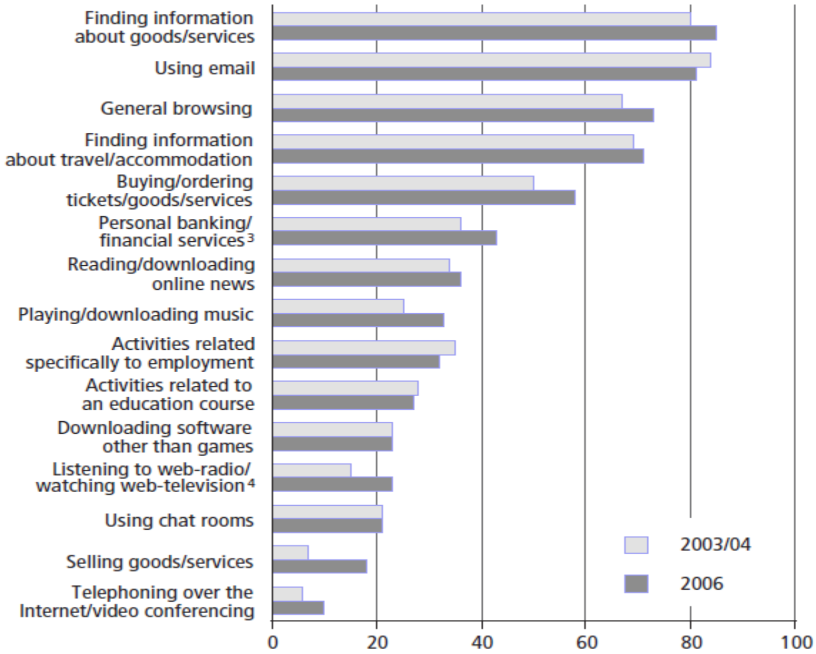
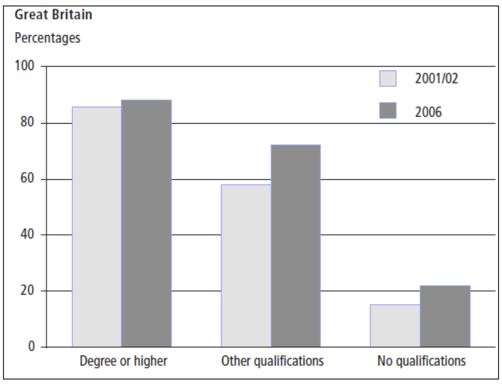


Figure 1.1. Percentage of adults aged 16 and over who had used the Internet and breakdown of their activities online in the three months before the interview.

Source: Omnibus Survey, Office for National Statistics data were collected between January and April 2006



United Kingdom	Percentages
Self-study in the sense of learning-by-doing	48
Informal assistance from colleagues, relatives, friends	45
Formal educational institution (school, college, university, etc)	30
Vocational training courses (on the demand of employer)	24
Self-study using books, cd-roms, etc	17
Training courses in adult education centre (but not on employer's initiative)	15
Some other way	1

Figure 1.1. Distribution of users of ICT in UK (2006) and where and how adults obtained ICT skills.

Sources: Omnibus Survey, Office for National Statistics; Skills for Life Survey 2003, Department for Education and Skills; Survey data for 2006 were collected in January, February and April.

ICT in Schools Survey, British Educational Communications and Technology Agency (Becta), Department for Education and Skills

Data also showed that there are demographic differences with heavier internet users between the ages of 31-49, males, graduates, with managerial or professional roles. The level of education of these users (Figure 1.2) partly justifies targeting university students for ICT training as mentioned earlier in the NRC report. One of the most interesting facts emerging from this source is that half of the people surveyed reported that they developed their IT skills in a self-directed way or assisted by peers rather than from formal teaching/training. This poses a very interesting premise for the discussion in the next section. In fact, in 2006 only a couple of computer applications (i.e. word processors and spreadsheets) featured in most adults' learning programmes and, in general, the aim of ICT seemed to be to provide support tools for learning (Figure 1.3), which are rather limited in scope.

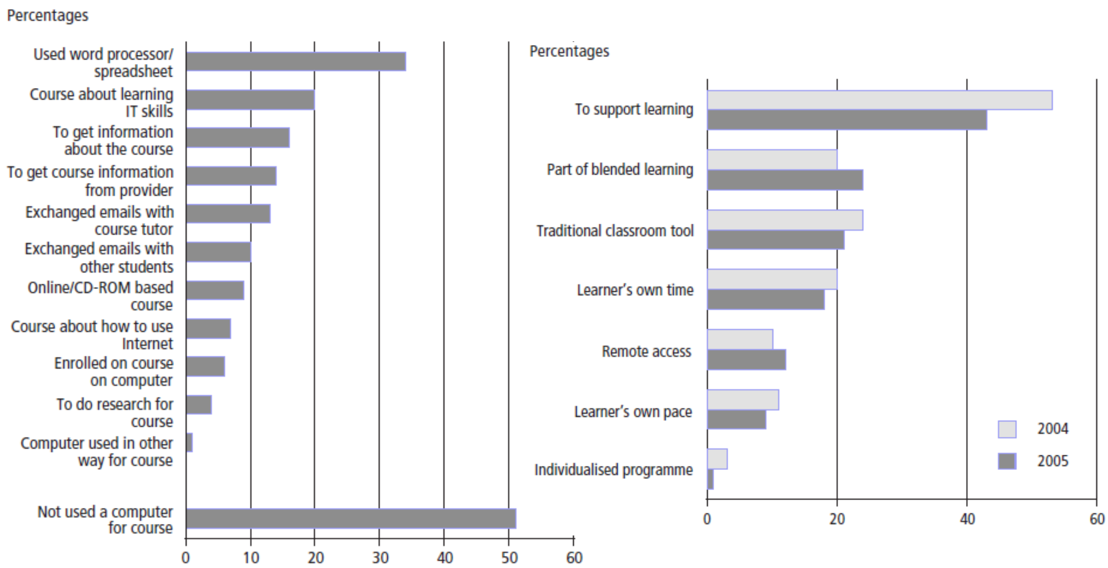


Figure 1.2. Use of ICT in adult learners for taught programmes in 2006 (left) and colleges mainstream programmes (right). Sources National Adult Learner survey and BECTA.

1.2. An overview of modern education

For those born in the first half of the 20th century it is very hard to think about higher education without making an association with traditional Institutions. In the UK venerable universities such as Oxford (founded about 1096), Cambridge (about 1208) in England or St Andrews (1409), Glasgow (1451) and Edinburgh (1582) in Scotland, all have a rich history and traditions, instinctively projecting an aura of accomplishment which is demonstrated by the achievements of many of their graduates who became important figures in many disciplines. Even though older Universities have the tendency of protecting traditional or longstanding practices, in the past three decades even these seasoned institutions have had to face a set of completely new challenges which are shaking traditions to their foundations.

Using figures from the OECD (Organisation for Economic Co-operation and Development between 1975 and 2000), the proportion of adults with higher educational qualifications almost doubled from 22% to 41%. Most rich countries are still struggling to cope with this growth in education demands, which forces newer generations to seek higher level qualifications to be able to compete in the job market and these figures do not include developing countries such China. A fairly recent article in the Economist, (Wooldridge, 2005) suggested that one of the reasons pushing the change is "the rise of the knowledge economy". In the *information society* knowledge is replacing physical resources as the main driver of economic growth and universities are not only the places to prepare workers, but also the backbones supporting change thanks to their labs, libraries and IT support systems (Eisenberg, 2008; Garnham, 2000).

Globalization is another factor which is affecting changes: the era of the *elitarian* approach to talent selection is certainly coming to an end. In fact, Universities are competing for talent as much as for fee-paying students to ensure their existence. From the same source, people from the OECD countries studying abroad have doubled to 1.9m in the last 20 years, which means big business. The World Bank estimated that there are more than 80m students worldwide engaging in higher education training, supported by 3.5m people employed by higher education institutions. Global spending in higher education amounts to $300 billion per year, equivalent to 1% of the global economic output. In principle this heralds a potential golden age for universities, but in practice, the role of governments is actually making things worse, failing to promote multilingualism and stifling mobility with ever tighter immigration rules. Furthermore, even though in most cases they embrace this *massification* of education, they do very little to fully support their enthusiasm. Academic jobs attract lower salaries than similar jobs in the industry and tighter management cannot make up for a lack of resources where an increase in demands from the student population cannot be accommodated in traditional campuses.

Another effect of massification is a reduced attention given to individual students, which is bound to become an issue: the number of students has increased, but often this is not matched by an increase in teaching staff, with departments often relying on platoons of postgraduates to deal with the day-to-day teaching. Old universities were born in a world where only a tiny minority of the population went into higher education; nowadays an undergraduate degree is almost a prerequisite to enter the job market. Yet, many academics have been reluctant to make allowances for this massification: traditional lectures are still one of the foundations for the *transfer* of knowledge. To mention an extreme case, Italian universities still insist that all students undergo a *viva voce* examination by a full professor for every module studied. The average length of such examination, when there are hundreds (if not thousands) of students to be examined is about 5 minutes: hardly a fair or satisfactory system to evaluate *proficiency*. In the UK, a shift toward fewer items of formative assessment in the curriculum seems to have become the norm: the increase of multiple choice-type (MCQ) assessments, especially with larger classes, is an effective way to conduct summative assessment reducing the workload of teaching staff.

Cultural conservatives believe that the best way forward is a return to traditions. The two ruling principles of modern higher-education policy (democracy and utility) are "degradations of the academic dogma" (cited from sociologist Robert Nisbet).

> "They (cultural conservatives) think it is foolish to waste higher education on people who would rather study "Seinfeld"[1] than Socrates, and disingenuous to confuse the pursuit of truth with the pursuit of profit." (Wooldridge, 2005)

Techno-utopians[2] jumped on the opportunity to argue against cultural conservatives and preached the benefit of technological innovation to counter the limitations of traditional

---

[1] Wikipedia entry for the term *Seinfeld*: "Seinfeld is an American situation comedy that originally aired on NBC from July 5, 1989 to May 14, 1998, lasting nine seasons. Many of its catchphrases have entered into the popular culture lexicon. The show led the Arthur Nielsen Media Research Ratings in its sixth and ninth seasons and finished among the top two (along with NBC's ER) every year from 1994 to 1998. In 2002, TV Guide named Seinfeld as the greatest television program of all time. A 2006 sitcom industry poll conducted by the United Kingdom's Channel 4 voted Seinfeld as the third best sitcom ever, ranking behind Frasier and Fawlty Towers." This quote is used here specifically to demonstrate the level of details that users of the Wikipedia are willing to share. (accessed on the 9/9/08, please note that content might change: http://en.wikipedia.org/wiki/Seinfeld)

[2] Ibid Wikipedia: "Techno-utopianism or technoutopianism refers to any ideology based on the belief that advanced science and technology will eventually bring about an utopia or, more precisely, a techno-utopia, a future society with ideal living conditions for all its citizens." (accessed on the 9/9/08, please note that content might change: http://en.wikipedia.org/wiki/Techno-utopianism)

education. In this respect learning technology, automated systems and online learning are all seen as the panacea to solve education's biggest problems. However, education should not be just about transmitting a body of facts: the Internet seems to do this quite well already, and the use of Wikipedia quotes in this context is intentionally critical, raising some interesting issues about the *authority* of the information available. Instead, the core of higher education *should be* learning to reason and to argue as well as equipping learners to become independent and self-directed. A community of scholars is then essential to support the process of learning, to become a member of a community of professionals, or, to use an expression coined by an educationalist, a *community of practice* (Wenger, 1999; Wenger, McDermott, & Snyder, 2002). Such a take might be considered idealistic, given the current trends, especially because education always seems to fall for an *individualistic* and *mentalistic* approach to teaching. Students are expected to do individual work, are assessed individually (in most circumstances) and low achievers are usually realising the self-fulfilling prophecy that *they are not good enough*. Obviously there are exceptions to this representation, however the idea of individual acquisition of knowledge is pervasive and it is the main driver for the need to earn degrees and qualifications.

This is an oxymoron: building knowledge and the progress of science are very much a social effort: lectures are by definition a way to *share* knowledge, books are written to divulge to the community, researchers publish in peer reviewed journals and research grants are reviewed by committees. Cultures are shaped by shared norms and beliefs. In the digital era, the Internet has become the biggest and most accessible way to access this *shared culture*, gather information on any topic and certainly one of the best places to dwell in communities of practice in which knowledge is shared by experts with novices. The expression 'did you Google X' has become a synonym for searching, and according to some critics (Keen, 2007) the "cult of the amateur" which is driving the information society is "killing our culture" and "assaulting our economy". We will return to this sharp criticism at a later stage, however, even the opposite techno-utopian view according to which technology-driven learning and teaching can fill the gap where the current education system fails (and this seems to be the case for the 4 out of 10 people who pursue self-directed learning using the Internet as well as traditional forms of teaching), might also fail as the human presence is essential to support the process of learning. This might be one of the most important reasons why e-learning fails in industry-based training or might explain why about 40% of students enrolled in distance learning do not complete their qualifications.

Policymakers certainly face the difficult problem of how to create a system of higher education that balances the twin demands of excellence and mass access, which makes room for global elite universities while also catering for large numbers of average students, which exploits the opportunities provided by new technology while also recognizing that education requires humans to run the system. These are all interesting aspects which provide an essential understanding for the interaction between IT and education from a socio-cultural perspective.

However, to keep the focus on the core area of investigation of this thesis, it is necessary to contextualise the psychology courses in the University of Edinburgh and to a wider extent, in the UK and Scottish higher education landscape.

1.3. Psychology: its popularity and misconceptions

In agreement with the boom in HE mentioned earlier, figures from the Higher Education Statistics Agency (HESA) indicate that, to date, the University of Edinburgh is the third largest university in Scotland closely behind Strathclyde University and the University of Glasgow with an intake of over 24000 new students every year. Like many other Universities it has experienced a 9% increase of students' intake in the past five years alone. From '95, in which 17576 students were admitted to their first year (undergraduate or postgraduate) the number of students admitted in 2007 was up to 24225, a 27.6% increase. This figure accounts for about 10% of students' intake in Scottish universities and 1% of the UK students admissions.

In the UK, the British Psychological Society (BPS) accredits degrees and courses at universities throughout the UK. In 2008, the BPS included nearly 700 undergraduate degrees, and more than 70 postgraduate professional training psychology courses.
Within the UK, Psychology degrees are a very popular subject with an average 2.9% of the new enrolments. Its popularity has also increased of about ¾ of a percentage point only in the past five years. There are many reasons explaining these trends: for this research it is useful to understand why so many take psychology courses in their degrees.

							United Kingdom			Other European Union			Non-European-Union			Total Gender	
Academic year	Total HE students	Total Psychology	FT UGs	FT PGs	PT UGs	PT PGs	Total	Female	Male	Total	Female	Male	Total	Female	Male	Female	Male
2002-03	2175115	50780	35795	4635	3735	6615	46860	37195	9665	2335	1825	505	1585	1215	370	40235	10540
%		2.33	1.65	0.21	0.17	0.30	2.15	79.37	20.63	0.11	78.16	21.63	0.07	76.66	23.34	79.23	20.76
2003-04	2247440	64480	38580	5120	13495	7290	60110	47555	12555	2360	1870	490	2010	1570	440	50995	13485
%		2.87	1.72	0.23	0.60	0.32	2.67	79.11	20.89	0.11	79.24	20.76	0.09	78.11	21.89	79.09	20.91
2004-05	2287540	68265	41175	5470	14125	7500	63580	50485	13095	2560	2035	525	2130	1625	505	54145	14125
%		2.98	1.80	0.24	0.62	0.33	2.78	79.40	20.60	0.11	79.49	20.51	0.09	76.29	23.71	79.32	20.69
2005-06	2336110	71185	43200	5800	14180	8005	66085	52445	13635	2845	2265	585	2255	1735	520	56445	14740
%		3.05	1.85	0.25	0.61	0.34	2.83	79.36	20.63	0.12	79.61	20.56	0.10	76.94	23.06	79.29	20.71
2006-07	2362815	72475	43725	6415	14290	8045	66890	53250	13640	3070	2445	625	2520	1915	605	57610	14870
%		3.07	1.85	0.27	0.60	0.34	2.83	79.61	20.39	0.13	79.64	20.36	0.11	75.99	24.01	79.49	20.52
Total period	11409020	327185	202475	27440	59825	37455	303525	240930	62590	13170	10440	2730	10500	8060	2440	259430	67760
%		2.87	1.77	0.24	0.52	0.33	2.66	79.38	20.62	0.12	79.27	20.73	0.09	76.76	23.24	79.29	20.71

Table 1.2. Overall Admission of new students in Psychology compared with the total HE student population over the period 2002-07. Source HESA reports (http://www.hesa.ac.uk/, accessed August 2008)

							United Kingdom			Other European Union			Non
Subjects	Single Subject Rank	Total subject intake	FT UGs	FT PGs	PT UGs	PT PGs	Total	Female	Male	Total	Female	Male	Total
Nursing	1	932505	441985	6430	440735	43355	899285	798360	100920	9705	8585	1120	23510
Creative arts & design		737965	579555	45185	79540	33675	656830	394870	261955	34625	21655	12965	46510
Engineering & technology		681480	380785	107425	105645	87630	483650	70750	412900	59915	9350	50570	137915
Business studies	2	656765	330125	76105	113290	137235	490285	230185	260100	47105	20750	26345	119375
Historical & philosophical studies		493510	250760	36495	151220	55045	453025	252840	200185	15365	8680	6670	25115
Training teachers	3	441835	122700	147405	69420	102325	425295	314190	111110	9050	7105	1950	7485
Law	4	428750	247445	66540	61710	53060	351410	213070	138345	21135	11930	9210	56200
Computer science	5	423690	260415	45645	86485	31135	348075	74520	273560	19500	3405	16095	56110
Physical sciences		390100	243535	63395	47560	35595	342620	140145	202475	19545	8635	10910	27920
Languages less English		388670	187530	29255	147240	24630	326570	212635	113950	29530	20725	8810	32575
Management studies	6	333515	150020	40610	48980	93890	255915	124050	131865	18060	8055	10005	59545
Psychology	**7**	**327185**	**202475**	**27440**	**59825**	**37455**	**303525**	**240930**	**62590**	**13170**	**10440**	**2730**	**10500**
English studies	8	290255	192990	17115	64035	16115	247855	180595	67250	10815	7295	3515	31580
Medicine & dentistry		280670	191995	31330	945	56400	242125	140225	101895	10170	6150	4015	28380
Social work	9	270580	79580	17080	141135	32790	265445	213690	51760	2375	1930	445	2765
Architecture, building & planning		260350	125375	27585	64090	43305	219585	63210	156375	14845	6305	8545	25905
Mass communications & documentation		230270	165255	26080	16515	22410	198275	115105	83170	12440	8280	4165	19555

							United Kingdom			Other European Union			Non
Subjects	Single Subject Rank	Total % intake	FT UGs	FT PGs	PT UGs	PT PGs	Total	Female	Male	Total	Female	Male	Total
Nursing %	1	8.17	3.87	0.06	3.86	0.38	7.88	88.78	11.22	0.09	88.46	11.54	0.21
Creative arts & design %		6.47	5.08	0.40	0.70	0.30	5.76	60.12	39.88	0.30	62.54	37.44	0.41
Engineering & technology %		5.97	3.34	0.94	0.93	0.77	4.24	14.63	85.37	0.53	15.61	84.40	1.21
Business studies %	2	5.76	2.89	0.67	0.99	1.20	4.30	46.95	53.05	0.41	44.05	55.93	1.05
Historical & philosophical studies %		4.33	2.20	0.32	1.33	0.48	3.97	55.81	44.19	0.13	56.49	43.41	0.22
Training teachers %	3	3.87	1.08	1.29	0.61	0.90	3.73	73.88	26.13	0.08	78.51	21.55	0.07
Law %	4	3.76	2.17	0.58	0.54	0.47	3.08	60.63	39.37	0.19	56.45	43.58	0.49
Computer science %	5	3.71	2.28	0.40	0.76	0.27	3.05	21.41	78.59	0.17	17.46	82.54	0.49
Physical sciences %		3.42	2.13	0.56	0.42	0.31	3.00	40.90	59.10	0.17	44.18	55.82	0.24
Languages less English %		3.41	1.64	0.26	1.29	0.22	2.86	65.11	34.89	0.26	70.18	29.83	0.29
Management studies %	6	2.92	1.31	0.36	0.43	0.82	2.24	48.47	51.53	0.16	44.60	55.40	0.52
Psychology %	**7**	**2.87**	**1.77**	**0.24**	**0.52**	**0.33**	**2.66**	**79.38**	**20.62**	**0.12**	**79.27**	**20.73**	**0.09**
English studies %	8	2.54	1.69	0.15	0.56	0.14	2.17	72.86	27.13	0.09	67.45	32.50	0.28
Medicine & dentistry %		2.46	1.68	0.27	0.01	0.49	2.12	57.91	42.08	0.09	60.47	39.48	0.25
Social work %	9	2.37	0.70	0.15	1.24	0.29	2.33	80.50	19.50	0.02	81.26	18.74	0.02
Architecture, building & planning %		2.28	1.10	0.24	0.56	0.38	1.92	28.79	71.21	0.13	42.47	57.56	0.23
Mass communications & documentation %		2.02	1.45	0.23	0.14	0.20	1.74	58.05	41.95	0.11	66.56	33.48	0.17
total students 2002-07:	**11409020**												

Table 1.3. Ranking of the intake of new students across different disciplines.

Please note that this table has comparative value with either single subjects (as in the HESA report) or discipline areas. Ranking was done on single subjects only using an average of the period under exam (2002-07).Source HESA reports (http://www.hesa.ac.uk/, accessed August 2008)

The admission figures provided by HESA (table 1.2 and 1.3) show the distribution of new enrolments in psychology during 2002-07 across UK universities with an average of 65000 new students every year. Table 1.2 provides a summary of new admissions taking into account other popular choices for comparison. Even though some entries refer to a discipline which comprises different subject areas (using the same denominations adopted in the HESA reports), ranking in this table is assigned considering the average entries over the period for single subjects only. The bottom panel of Table 1.3 also highlights some expected disparities in the subject choice with stereotypical higher rates of females in the social sciences and males in engineering and computer science. Students enrolling in the psychology course at the University of Edinburgh account for about 1% of new admissions every year.

1.3.1. Why do students enrol in psychology?

Despite its popularity, students do not seem to have a full understanding of what psychology entails. The issue is particularly important because in Edinburgh the courses can be taken as “outside subjects” for other degrees in year 1 and year 2 and the fact that a fair number of students change their degree path before getting into Honours.

From anecdotal evidence collected in the past 3 years as School Liaison Officer, the most common answers from prospective students to the question ‘Why did you decide to apply for a psychology degree?’ are ‘I don’t know’ or ‘I think it will be quite interesting’.

When meeting them, usually in the year before their A Levels or Highers exams at age 16 or 17, one of the common denominators seems to be that prospective students want to learn more about themselves, gain a more objective view on their relationships with family and friends and only a small number of them can grasp the *scientific* value of psychology as a discipline. In a more recent survey with 130 psychology students in their second year conducted on the cohort which started their degree programme in 2007-08, when asked to rank 6 reasons for taking a psychology course, the top three choices were Interest (i.e. find out more), Personal development (i.e. potentially useful) or intellectual (i.e. academic value). The fact that little is known about the reasons for choosing psychology is also supported by others in the field of psychology and education. For example, Professor Dewart, Head of Psychology at Westminster, in her inaugural lecture back in December 2005, contended that very little is known about psychology students themselves.

> “Despite the growing popularity of psychology degrees we know very little about what motivates students to choose psychology, and what their expectations and experiences are when they do (…) Taking a degree in psychology can be seen as a process of 'studying minds', but little study has been conducted into the minds of psychology

> students (...) As psychologists we can begin to investigate these 'studying minds'. We can attempt to identify the pressures they experience and develop ways in which they can be supported. The findings of this research may have implications for the student experience in general and for future developments in psychology education”. (Dewart, 2005)

A similar message was beaconed in the latest International Conference on the Teaching of Psychology in July 2008 by the likes of Charles Brewer, head of the American Psychological Association (APA) Teaching Psychology Division, Annie Trapp (Higher Education Academy) and at the PLAT Conference by Jane Halonen and David Boud. In fact, most people teaching in psychology, in their way of teaching, do not explicitly take into account the psychological theory which affects directly how students learn. To exemplify this statement, it is useful to consider two extreme cases presented in the literature: Winer and colleagues (Winer, Cottrell, Gregg, Fournier, & Bica, 2002) summarized a series of studies that found the belief that extramission theory of vision (i.e. X-ray vision like Superman) is fairly common for the naïve layperson and hard to correct. They showed that, depending on how one tests for it, more than half the population held some version of this belief. After reviewing a number of studies that tried to correct extramission beliefs, Winer et al. "found no evidence that traditional readings presented immediately before the test, formal classroom experiences, or the combination of both improved performance" (p. 421). They did find, however, that when university students were shown a highly simplified lecture on vision containing explicit statements and evidence refuting extramission beliefs, there was a reduction in those beliefs. The improvement, though, was temporary and disappeared after 5 months. As cynical at it might appear, it may be that students leave psychology courses with their misconceptions intact. Furthermore, they may feel more confident in their mistaken beliefs because they have taken a psychology course (Landau & Bavaria, 2003), which support their mis-beliefs. Recently, Richard Miller (2008) showed that dissonance can be used effectively in overcoming students’ beliefs in paranormal and pseudoscience. However, he also indicated that the effects are stronger when students feel in control of their thinking processes and demonstrated that overexposure to counter-attitudinal messages can produce the opposite effect.

Many teachers fail to address misconceptions because they believe the primary focus of teaching is *presenting information accurately and clearly* and this is sufficient for good teaching. Gardner (Gardner, 1993a) showed that a similar issue is found in physics and biology, subjects which are traditionally associated with fact-based instruction.

The fact that students have misconceptions would be irrelevant if such beliefs had no impact on further learning. However a large body of literature indicates that one's schema, or belief system, can have a major impact on what is noticed, what is learned, what is forgotten, and how memories may become distorted (Bower, Black, & Turner, 1979; Bransford & Johnson, 1972).

The extreme cases presented for supernatural powers or pseudoscience are used here as simple tools to highlight the importance of students' *expectations* and *motivations*. In the past couple of years a number of research papers were published focusing on students' expectations, not only of psychology, but for their entire university experience. For example, using a Canadian sample of undergraduate students Jackson and colleagues (Jackson, Pancer, Pratt, & Hunsberger, 2006) determined that four cluster could characterize sufficiently well the expectation during the transition to University: Optimistic (35%), Prepared (25%), Fearful (14%) and Complacent (27%). Furthermore they were able to determine that expectancies predicted adjustment. These expectancies are not just a reflection of chronic differences in coping strategy, but they possibly arise as much from what students are told by teachers, parents, friends as from past experience. Furthermore, while students may be helped by providing realistic information about university, to help students (the authors report a dropout rate of up to 40% in many cases) is to provide the 'fearful' students adequate support to develop coping strategies.

Even if this argument seems to drift away from the core questions of the thesis, expectations and motivation are essential in our context as we will see that only a third of students registering in our first year psychology course do so as part of a single-subject degree. The majority take the course as an outside subject and this has obvious implication for the way in which the course can achieve its set outcomes.

It is also important to mention expectations and motivations as these will not be explored directly in this thesis. Motivation leading to action, will be considered when students make the explicit decision to use the additional, non-compulsory, resources available to them in the form of e-learning tools. Therefore we will be able to consider this as a proxy measure for motivation. Expectations will be only considered for group data when evaluating what career paths students take and through the National student survey (in the next sections). This indirect approach could be seen as a methodological weakness, but we try to counter this by showing the mismatch between students' *perception* and *utility* of some of the tools used, which makes it even more important to properly address expectation and motivation in further research.

1.3.2. Learning for what? Issues of educational vs learners' needs

Another aspect we would like to address before moving on the discussion of learning technology is that of professional recognition and career prospects for psychology graduates. Unlike other disciplines, psychology is inherently multidisciplinary, with strong links with many other areas and offers the possibility to pursue further training in both sciences and social sciences, but the flexibility of the transferable skills gained during the undergraduate degree are often considered satisfactory to work in a variety of placements as hinted by the HESA first destination Survey 2001 (table 1.4).

The BPS is the body responsible for accreditation and professional recognition in the UK and accredits undergraduate degrees on two bases: they are either accredited to give a holder of the degree the society's Graduate basis of Registration (GBR, now GBC), or simply to qualify them for Graduate Membership of the Society.
There are several areas of psychology in which people can qualify as a Chartered Psychologist and it is these that the Society recognises as the main types of psychologists. The GBR is an essential step to become a Chartered Psychologist in one of the recognised Divisions.

At the time of writing the BPS recognises 9 different types of psychologist:

- Clinical psychologists
- Counselling psychologists
- Educational psychologists
- Forensic psychologists
- Health psychologists
- Neuropsychologists
- Occupational psychologists
- Sport and exercise psychologists
- Teachers and researchers in psychology

			First degree						Higher degree			Other postgraduate			Other undergraduate				
Academic year	Total HE students	Total Psychology	**Total first degree**	First class	Upper second	Lower second	Third class/ Pass	no class	**Total higher degree**	Doctorate	Other higher degree	**Total other post-graduate**	PGCE	Other post-graduate quals	**Total other under-graduate**	Foundation degree	HND DipHE	Other under-graduate quals	ratio admisions v graduations
2002-03	557790	12845	8900	775	4955	2695	285	190	2770	780	1990	575	0	575	605	10	130	465	3.95
%		0.23	**69.29**	6.03	38.58	20.98	2.22	1.48	**21.56**	6.07	15.49	**4.48**	0.00	4.48	**4.71**	0.08	1.01	3.62	
2003-04	595640	15050	10405	1010	5660	3040	400	295	3065	730	2335	785	0	785	795	5	150	640	4.28
%		0.25	**69.14**	6.71	37.61	20.20	2.66	1.96	**20.37**	4.85	15.51	**5.22**	0.00	5.22	**5.28**	0.03	1.00	4.25	
2004-05	633045	16665	11435	1085	6335	3315	455	240	3305	805	2505	705	0	705	1215	5	180	1025	4.10
%		0.26	**68.62**	6.51	38.01	19.89	2.73	1.44	**19.83**	4.83	15.03	**4.23**	0.00	4.23	**7.29**	0.03	1.08	6.15	
2005-06	640850	18125	12200	1330	6755	3440	470	205	3765	830	2935	750	0	750	1405	10	210	1190	3.93
%		0.28	**67.31**	7.34	37.27	18.98	2.59	1.13	**20.77**	4.58	16.19	**4.14**	0.00	4.14	**7.75**	0.06	1.16	6.57	
2006-07	651060	18575	12455	1405	6895	3400	455	300	3770	900	2870	750	0	750	1595	15	230	1355	3.90
%		0.29	**67.05**	7.56	37.12	18.30	2.45	1.62	**20.30**	4.85	15.45	**4.04**	0.00	4.04	**8.59**	0.08	1.24	7.29	
Total period	3078385	81260	55395	5605	30600	15890	2065	1230	16675	4045	12635	3565	0	3565	5615	45	900	4675	4.03
%		0.26	**68.17**	6.90	37.66	19.55	2.54	1.51	**20.52**	4.98	15.55	**4.39**	0.00	4.39	**6.91**	0.06	1.11	5.75	

Table 1.4. Distribution of Psychology graduates by level and year of graduation with a ratio of admissions per graduation.

.Source HESA reports (http://www.hesa.ac.uk/, accessed August 2008)

Related and cognate professions, especially with a clinical involvement are acknowledged and monitored, but these are regulated in the UK by different bodies. The most notable are Counsellors and Psychiatrists who have separate bodies (respectively the Royal College of Psychiatrists and the British Association for Counselling and Psychotherapy) to regulate their profession.

The reason for considering the official structure of the professional options in psychology, is that it is important to offer a framework to understand students' expectations and misconceptions beyond the single courses.

In the table above it is quite interesting to observe that (as expected) the number of graduates in psychology has been increasing steadily, 7% in the period under consideration. This is just above the overall increase of new admissions (6.6% for the same period), without over-saturating the market (like in other sectors), but providing a steady supply of graduates ready for further training or employment.

	Psychology (science)	%	Psychology (social sciences)	%	Psychology (All)	%
Total Undergraduate in Psychology	**2740**		**790**		3530	
Health & social work	640	23.4%	170	21.5%	810	22.9%
Property development, renting, business & research activities	435	15.9%	115	14.6%	550	15.6%
Education	355	13.0%	85	10.8%	440	12.5%
Wholesale & retail trade/Repair of motor vehicles, motorcycles & personal & household goods	295	10.8%	95	12.0%	390	11.0%
Public administration & defence/Social security	295	10.8%	95	12.0%	390	11.0%
Financial activities	240	8.8%	85	10.8%	325	9.2%
Manufacturing	115	4.2%	30	3.8%	145	4.1%
Hotels & restaurants	105	3.8%	40	5.1%	145	4.1%
Other community, social & personal service activities	105	3.8%	25	3.2%	130	3.7%
Transport, storage & communication	80	2.9%	25	3.2%	105	3.0%
Electricity, gas & water supply	30	1.1%	5	0.6%	35	1.0%
Construction	15	0.5%	10	1.3%	25	0.7%
Not known	15	0.5%	5	0.6%	20	0.6%
Private households with employed persons	10	0.4%	0	0.0%	10	0.3%
Mining & quarrying	5	0.2%	0	0.0%	5	0.1%
Agriculture & forestry	0	0.0%	0	0.0%	0	0.0%
Fishing	0	0.0%	0	0.0%	0	0.0%
International organisations & bodies	0	0.0%	0	0.0%	0	0.0%
Total	26	0.9%	88	11.1%	114	3.2%

Table 1.5. Distribution of placements for psychology graduates in the UK between 2001-02.

Table 1.5 –with its limitations[3]- is testifying that the stereotypical preconception of a psychologist as a clinical psychologist, who engages in some form of psychotherapy with

[3] It should be noted that data about destinations is much less complete than students' data because it is very difficult to track students when they leave University. The small percentage of known destination compare to the number of graduates is a clear indicator, therefore these data should be considered with caution.

paying customers, is reflected in the typology of employments which seems to attract psychology graduates most. Experience in the public sector, mainly as supervised assistant psychologist or social worker, is considered a requirement before pursuing a further qualification in clinical psychology. An alternative route is the one of researcher working under the supervision of a qualified psychologist and this also seems to be quite popular. As indicated in the table, however, other options are also feasible spanning in virtually all sectors.

1.3.3. What do students say about their psychology degrees

Another interesting resource which students refer to when they apply for a psychology course is the National Student Survey which provides an overall (although possibly biased) picture of what students think of their courses. As much as the RAE is used to assess the scientific value of a department and profile the research strengths and potential (also financial), the student survey is widely used by the media to compile league tables of students' satisfaction which are inevitably taken into account by prospective students.

For example, in the 2007-09 academic years, the psychology department in Edinburgh scored below average in the aspects of assessment and feedback. Even though we need to keep in mind the limitations of the tool, the NAS provides an insight into students' mismatched expectations, and could be a useful indicator to re-think about current practices within a Department or School. A further criticism which was put forward by the likes of Kulik and McKeachie (1975) or Cohen (Cohen, 1980, 1981) is regarding the *true* value of students' feedback. The core aspect of such criticism is based on the fact that students, as novices in the area of study, cannot provide useful criticism because they don't have the necessary knowledge nor the expertise to judge the quality of what is presented to them. A counter argument, however, is that academics are so used to theoretical skirmishes that they are blinded by the fact that students at University have been *learners* for nearly two decades before they reach university, therefore they can use this expertise to help instructors to provide a better teaching experience.

There will be scope in this thesis to evaluate these two positions as we will make a case that students don't necessarily know what is best for them, however, in agreement with the meta-analysis conducted by Cohen (Cohen, 1980), we strongly feel that students' feedback is an essential aspect to consider in informing better teaching quality and this is certainly facilitated by new mediums, which could help to make feedback more frequent, more contextualized and more efficient.

Understanding the 'digital divide'

According to Kate Williams (Williams, 2001) and the exchanges from many contributors of the U.S.-based listserv *digitaldividenetwork*, the term *digital divide* came into regular use in the mid-1990s after Bill Clinton applied it to identify the gaps in ownership of computers between groups in a speech in Knoxville.

The term specifically identifies the gap between those people with effective access to digital and information technology and those without access to it. The concept might include the imbalances in physical access to technology as well as the imbalances in resources and skills needed to effectively use it. Groups often discussed in the context of a digital divide include socioeconomic (rich/poor), racial (majority/minority), generational (young/old) or geographical (urban/rural).

More recently Mark Prensky (2006) emphasised more the generational difference between those who were born and have always been immersed in the technological world of computer and gadgetry and those who had to embrace technology at some point in their lives, defined as *digital immigrants*. His argument is simple and appealing to many: the world has evolved, technological and cultural changes marvel those who lived in the last century; schools and the educational system has changed little and adapted slowly. My grandparents are still trying to come to terms with mobile phones: they happily carry around a mobile and use its more traditional functionalities, but they can still recollect that in their childhood there was a single telephone in the local village and their opportunities to use it were very rare. Their usage patterns remained anchored to the idea that access to the phone should be only for important communications.

Computers access in schools is a very recent development, but in many primary schools, especially in the countryside it is still a commodity which is not available to all.
The wide availability of technology would make one think that "more is better"; the reality, however is quite different. Pallfrey and Gasser in "Born digital" (2008) noticed the fact that one is born with technology doesn't make him/her an expert and often immigrants become better than the natives. As obvious as this statement might sound, the fact is that the whole area is still very open, mainly based on case studies and simply put, doesn't provide the empirical evidence which is demanded in most area of research to support hypotheses and statements. In particular at a recent talk given at Harvard, Gasser (2008) advocated a need to

explore the issues surrounding the digital natives' debate by dissecting the problems into layers. The figure here is a close representation.

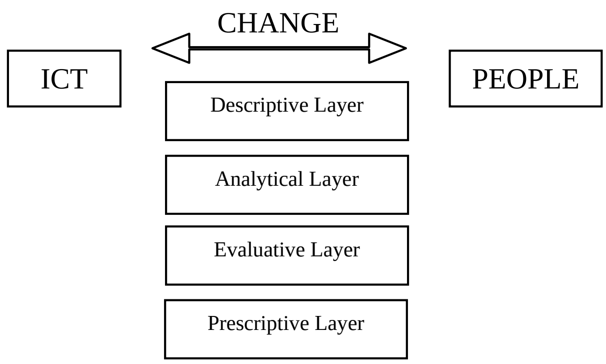


Figure 1.1.3. A framework to study the relations between ICT and people focusing on the changes occurring or emerging from the interaction. Reproduced from Gasser (2008).

This framework maps very well the purposes of this thesis and the evaluation framework which we will use to characterise the research done. We will come back to this framework in the next section when addressing the current state of evaluation in e-learning and later in the thesis when discussing the dynamics of interaction between students and e-learning.

1.4. E-learning: a solution to bridge the gap?

One of the most interesting pieces of information emerging from the surveys mentioned above and then touched upon in the discussion about digital natives, however, is the fact that half of the people using technology are *self-taught* and *learn-by-doing* (fig 1.3).
This observation is a crucial one: what is driving half the people surveyed to independently engage with technology and learn how to use it? Is this more effective (or efficient) than formal training, courses or curriculum-based activities?
It is clear that finding out *what* people do on the internet and *why* in an ecologically valid setting, has become an essential aspect to understand both the technology and the interactivity between users and the internet as well as focusing the attention of the education community to embed technology in the curriculum from an early age.

In Edinburgh we have interesting pre-conditions to support ICT. The University has in place a specific ICT strategy which strongly promotes a wide adoption of IT and learning technology. In the UK context, the University of Edinburgh maintains control of its communication and information systems infrastructure, which allows access to one of the fastest broadband connection lines in the world and the provision of *Eddie*, one of the world Super computers.

The use of learning technology, especially in the College of Medicine and Veterinary medicine, is considered cutting edge at international level; the School of Education is hosting the first fully online MSc in e-learning and an internal e-learning fund has allowed considerable developments across all colleges.

However, beyond the institutional level, the pedagogical conviction that e-learning *always* makes a positive change stands on shaky foundations. Research evidence presented by Russell (2001) in which hundreds of studies are reviewed with mixed results just adds to the debate. The voices of those arguing against the effectiveness of e-learning advocate reconsidering more carefully the issues of "augmentation vs. disruption" of e-learning (Grainne Conole, Laat, Dillon, & Darby, 2008; Heilesen & Josephsen, 2008). In other words, they question whether technology is really helping to improve teaching and learning and if the expectations and investments are actually delivering the expected results: all evidence points to the necessity of clarifying the methodological aspects of the evaluation.

Despite such criticisms, national entities like JISC persevere in promoting the use of technology and continue to showcase examples of good practice in applying learning technology to pedagogy. Even though these case studies are presented with a fairly rigorous framework, the evidence is not always convincing, especially when good practice fits only the specific context.

As in the descriptive layer mentioned by Gasser, finding out the answer to simple questions such as *what* people do, *when* and *how* they do it when interacting with learning technology is the first step to identify the bases to formulate questions about the contextual and general validity of these observations.

An analytical evaluation of the data is essential to provide evidence for the evaluative layer and we will advocate that one of the ways to approach the problem is to shift the focus on the learners and take into consideration the usage and level of success in e-learning to provide clear metrics which allows both designers and instructors to maximise the benefits of the tools and learn from examples of bad design to avoid wastage of funding.

Providing firm, shared methodological grounds to evaluate e-learning efforts is therefore paramount. This can be achieved by going back to evaluate the *tools* as well as the *evaluation methods* and we will address this aspect in more details in the next chapter.

Perhaps, puzzling to the most critical, is that since Skinner's teaching machines, the idea of employing technology and delivery systems in learning environments has never lost a core optimistic slant, despite empirical evidence maintaining a different stance. Why is it the case that the belief in the *goodness* of technology overcame rational analysis?

1.5. Posing the 'right' questions

The dynamic and vibrant landscape offered in the HE sector, with respect of e-learning and pedagogy are quite obvious, as well as the potential implications which successful learning technology could offer. What is most interesting, however, is that the current *hype* for e-learning is weakly supported even if heavily funded. This makes it necessary to take a step back from traditional methods of evaluation which will be reviewed in the next chapter. In fact, we will argue that the frameworks available to evaluate e-learning are limited and often offer merely a description of what a new tool has or has not done for a specific course or context.

Instead, we propose that e-learning should be explored using a strong psychological framework which takes into account the behavioural aspects of the interaction with e-learning and also the psychological precursors which might determine the behaviours. These can provide a rich profile of a student/learner, before we can characterise the specific interaction with the tools and partially pre-determine the type of interaction which will occur.

Just as a doctor would not prescribe medication without knowing a fairly detailed history of the patient and a comprehensive report of their symptoms, as instructors we cannot provide suitable pedagogical interventions nor, as technologists, can we provide effective tools to aid learning without a rich knowledge of both the context and the learners.

The thesis is intentionally placed at the boundaries between psychology, education and IT, however the methods used will often be atypical of these disciplines.

Describing the interaction between the users and the system and characterising the learners in details are therefore the core methods of this research which is very much exploratory. The core problem addressed, is how to identify *usable relations* from behavioural data emerging from the interaction with learning technology, and uncover prescriptive patterns which will help to shape better instruction and inform system designers to make new tools more effective.

Looking at the empirical evidence, we will propose a new evaluation framework grounded on data which is far richer than usability tests or records of students' satisfaction, typically presented in the literature. We expect to generate a number of specific, actionable criteria to shape how learning technology could augment instruction rather than replace it (as many might be deceived to believe).

In the next chapter we will review some core concepts in students' learning and consider the impact of ICT and its evaluation in HE. In chapter three and four we will discuss how psychological theory could provide an interesting way of characterising students. A differential psychology perspective, using metrics of intelligence and personality, has been a typical approach in the literature in which there is evidence of the relations between both IQ and personality traits with academic performance (AP) at university. After reviewing the evidence we will take a different direction and use measures of cognitive and learning styles to take some distance from abilities and focus on individual preferences. In chapter five we will present the specific case for psychology students in Edinburgh which offered the opportunity to explore a sample of over 2000 students with some longitudinal sub-samples. In chapters six to 8 we will then characterise the samples and historical trends in performance, empirical data on usage (behavioural data) and psychological profiles. In chapter nine we will attempt to integrate the data with an hypotheses-driven approach to provide a new model for evaluation which can be used in both HE and business, to improve the pedagogical value of the use of learning technology.

Chapter 2. From learning to e-learning: education, instructional technology and personalisation

"If you can look into the seeds of time,
And say which grain will grow and which will not,
Speak then to me, who neither beg nor fear
Your favours nor your hate."

(W. Shakespeare, Macbeth 1.3.58-61)

The title of this chapter, which is inferred from the title of the thesis, is hinting at its essential role to set out the scope of analysis of this thesis, to define and conceptualise the language and terms used.
Although *learning* is probably one of the most prolific and debated area of research in psychology and education, it would be impossible to systematically and coherently summarise such a vast body of literature in a single chapter.

Instead, this chapter has four key purposes. First we identify and select some key issues about learning theory in psychology and education that will be relevant for the overarching argument of the thesis and to lay out the foundations of the following discussion. In this context, *learning* is specifically intended as *student's learning*. Even by narrowing the scope, however, it is necessary to draw some boundaries as the entire field of educational psychology is devoted to discover how learning and instruction interact and how student learning is achieved and can be improved. For this reason we will focus on what qualifies *intended learning* and briefly mention a number of theories from psychological research, carefully relating theory to instructional design.

Secondly we will specifically address what is meant by *learning technology* and *e-learning*. Some, like Jay Cross at the WebCT/Blackboard conference back in 2002, contended that the "e" should be dropped altogether. Especially in the context of higher education, e-learning should be thought more in terms of *blended* learning (Bonk, Graham, Cross, & Moore, 2005; Heinze & Procter, 2004; Lenhart, Simon, & Graziano, 2001). Even though we agree that the core focus should be *learning*, it will be argued throughout this section that the "e" cannot be ignored or taken for granted. The area of research is very young, but the "e" is offering a completely new approach to support teaching and learning which relies on the understanding of instructional design and psychology of learning accrued in the past 50 years. It must be put into perspective to provide a new and more coherent approach to evaluate implementations of learning technology.

Thirdly, we will consider the issue of evaluation of e-learning in more detail: in the first chapter, we talked about aspects of access and accessibility, political issues and change, which make the thread of three of the core themes in e-learning research proposed by Conole and colleagues (Conole & Oliver, 2007). In a similar vein, we consider the interdisciplinary nature of learning and instruction and will focus on the *interactivity* (and engagement) as well as *social interaction* as essential aspects of the research on learning. These are essential to support the implementation, use and evaluation of e-learning in the curriculum.

Finally, hinting at the discussion in chapters three and four, and based on the discussion about the evaluation of e-learning, we will question whether "one size fits all" with learning technology and suggest that individual characteristics might actually be more important for a successful implementation than a comprehensive theory of e-learning effectiveness.

2.1. Learning and learning from instruction.

What is learning? Shuell postulated that either explicitly or implicitly, all conceptions of learning require three criteria:

> "(a) a change in an individual's behaviour or ability to do something, (b) a stipulation that this change must result from some sort of practice or experience, and (c) a stipulation that this change is an enduring one. The primary purpose of the latter two qualifications is to exclude certain types of behavioural changes that do not seem to represent what we mean by learning (maturation, temporary changes due to drugs etc.)" (Shuell, 1986, p. 412).

The latter specification is quite important because it allows differentiation between intended learning from other types (i.e. vicarious learning). An alternative and more specific view is given by Curry:

> "Intended learning is both a process and a product. The process is adaptive, future focused, and holistic, affecting an individual's cognitive, affective, social, and moral volitional skills. The product is observable as a relatively permanent change in behaviour, or potential behaviour. The process is observable in the improved ability of the individual to adapt to environmental stimuli." (Curry, 1983a, p. 2)

Either one of these two definitions is a suitable starting point for this discussion because neither is specifying the *content* nor the *source* of learning. These definitions reflect a general trend of the research in the psychology of learning which, for over half a century, has been heavily influenced by behaviourism, and which puts emphasis on outcomes rather than knowledge or knowledge structures (content).

2.1.1. Behaviourists' views of learning: outcomes vs. content

In the last century a systematic investigation of learning started with Thorndike (1911); in his seminal work he posed the foundation of most of the following research based on two basic laws: the Law of Effect and the Law of Exercise. A heated debate and an overwhelming amount of research on animals (from rats to mammals) was carried out well into the 1940s and 1950s with a theoretical standoff between Tolman, Watson (1924), Hull (1945), Spence (1951) and their opponents: even if the Law of Exercise waned under the pressure of overwhelming contradicting evidence, the Law of Effect remained a prominent feature and Skinner one of its strongest supporters. In fact Skinner himself, in a classic article (1954 reproduced in Ely & Plomp, 1996) suggested that:

> "Some promising advances have been made in the field of learning. The Law of Effect has been taken seriously; we have made sure that effects do occur and that they occur

> under conditions which are optimal for producing the changes called learning. Once we have arranged the particular type of consequence called reinforcement, our techniques permit us to shape up the behaviour of an organism almost at will." (Ely & Pomp 1996, p. 200)

Skinner was convinced that efficient learning is dependent on four basic principles and even suggested that teaching machines would do a better job than teachers if these principles are followed: 1) practice of the correct responses (Law of Exercise); 2) knowledge of results and reinforcement of the right answer (Law of Effect); 3) minimum delay of reinforcement (immediacy of feedback); 4) successive small steps should be hinted (what Thorndike called Law of associative shifting). The impact of Skinner's theory of learning in the article 'The science of learning and the art of teaching' influenced psychology profoundly. It was only 20 years later that McKeachie (1974) challenged each of the main principles and refuted them. McKeachie used a number of examples to dispute Skinner's view. The gist of his comments, is one example illustrated by Guthrie (1935) to undermine the immediacy of reinforcement: "a child spanked an hour after writing on the living room wall will usually learn not to write on the wall if he is told why he is being punished" (McKeachie 1974, p. 8). Then with a long list of studies he came to the conclusion that:

> "The research evidence, I believe, demonstrates that each point enunciated by Skinner is untrue – at least in as general a sense as he believed. This does not mean that Skinner's attempts to influence education have been bad or that the principles are completely false; rather his attempt to make a systematic effort at application has revealed that what we psychologists once took to be the verities hold only under limited conditions." (McKeachie 1974, p. 10)

There are two fundamental problems with Skinner's systematic attempt to apply learning theory to teaching: on one hand, Skinner expected a simple mapping between animals and humans without taking into account individual differences. On the other, the stimulus-response approach proved to be too simplistic to be useful in a realistic class environment.

Nevertheless, the idea behind operant conditioning, in principle, explained well how desired patterns of behaviour could be positively reinforced and is still, to large extent, driving pedagogy as exemplified by the use of assessment as the tool to provide positive or negative reinforcement for students (Biggs, 1996). It is worth noting that prompt reward might work in particular contexts and with certain types of individuals: for example token economies and systematic tangible reward have been used successfully with clinical cases, in penitentiaries and also to address problem behaviour in schools (Cook, 1999; Higgins, Williams, & McLaughlin, 2001; McGinnis, Friman, & Carltyon, 1999).

Similar trends and techniques can be found in e-learning and learning technology: a predominantly behaviourist approach has driven the work of people like Gagné (1965) and Merrill (2002), who meticulously explored systems of instruction in which a hierarchical model of diagnosis and intervention can drive instruction in a similar way to what Skinner called *teaching machines*.

Greeno pointed out that it was not until the 60's and 70's that the shift away from behaviourism offered by cognitive psychologists gained momentum. The shift in the interpretation was characterised by a "(…) discrete change between states of knowledge rather than change in probability of response (…)" (Greeno, 1980, p.716). This is particularly significant as a cognitive explanation of learning implies that behaviour must be the *result* of learning, rather than that, which is itself learned. In this sense, changing the environment central to the behaviourist perspective (i.e. providing appropriate reinforcement when the appropriate response is made), is insufficient for cognitive approaches, which focus on changing the learner (i.e. encourage the learner to use appropriate strategies).

2.1.2. Cognitive approaches

Cognitive approaches to learning can be traced back to the tradition of developmental psychology and Jean Piaget in particular, which gives us an opportunity to address the issues about the *source* of learning. For example, researchers embracing Piagetian views endorse a developmental process largely dictated by the individual's abilities and relative stages of development. The child is discovering and learning from experience as a little scientist. This fundamentally constructivist approach holds that a person constructs his/her comprehension of the surrounding world through learning and knowledge; this implicitly minimises, if not excludes, the impact of a teaching model based on the transfer of knowledge from the instructor to the pupils. On the other hand, researchers embracing the Vygotskian interpretation of development, see the interactive and social aspects as central to facilitate the learning process.

Even though the stage theory and the methods for investigating the acquisition of cognitive schemes and skill have been widely criticised, the principles of adaptation and assimilation which Piaget used to explain how the individual strives to maintain an equilibration with the surrounding environment, are still considered very influential. Ausubel (Ausubel, Novak, &

Hanesian, 1968), for example, built on the additive nature of the process of assimilation and stressed that the most influential factor affecting learning is ‘what the learner already knows’.

One of the biggest contributions of cognitive approaches to learning and which permeates Bruner’s view of education is the realization that learning is an active, constructive and goal oriented process. This involves three almost simultaneous processes of acquisition of new information, transformation of knowledge and evaluation (Bruner, 1985). Therefore, rather than observing a behaviour occurring, mostly focusing on *simple* learning, cognitive psychologists have an interest in analyzing performance and abilities in terms of the processes involved to achieve and improve learning. This involves exploring performance via tests of mental abilities, inductive and deductive reasoning, study of memory capacity and components of memory, and perceptual processing at various levels.

The attempt to make sense of *complex* learning and behaviours by studying the internal mental processes was mainly driven by the idea that learning is not a unitary process. Hierarchical models, with the strong support of a very fast escalation of computing power afforded by technological advancements, allowed researchers to explore the formalisms, heuristics and even modelling of some learning processes. In fact, from the 1960s there has been a great overlap in the work carried out in cognitive sciences, artificial intelligence and human-computer interaction, which were strongly grounded on the metaphor of the brain as a very powerful information processing system which could be emulated by computers and artificial systems. As we know, over 50 years later, such beliefs have lost their original appeal, but the research conducted in these three very different disciplines strongly contributed to refine our understanding of a variety of cognitive processes, including learning.

Miller, Galanter and Pribram’s (Miller, Galanter, & Pribram, 1960) seminal book was the first, which popularised the hierarchical organization of behaviours and has directed the study of processes of increasing complexity (Resnick, 1981). Ausubel subsumption theory, (Ausubel & Youssef, 1963) and Rumelhart & Norman (Norman, 1980, 1981; Rumelhart, 1979) produced the first formal models of cognitive processes and the latter is the first comprehensive theory of human cognition which was based on three core types of learning: accretion, restructuring or schema creation and tuning or schema evolution. In 1981 a further step was taken with the publication of the ACT[4] theory (Anderson, 1982). Most cognitive psychologists distinguish between declarative and procedural knowledge. In contrast to Rumelhart & Norman, Anderson believes that there is only a single set of learning processes that explains learning from acquisition to abstraction and problem solving. At the centre of the ACT theory, a network of propositions (declarative knowledge) and a set of IF-THEN rules (procedural knowledge) can explain both the acquisition and transition between the two domains.

A completely different perspective from Piagetian developmental theories is the view that knowledge and learning is interactive and meaning construed socially. Vygotsky is one of the researchers emerging from the Russian cultural-historical tradition who disputed that rather than self-discovery, children's potential could only be achieved thanks to the interaction with others. One of the most well known concepts advocated by Vygotsky is the concept of Zone of Proximal Development (ZPD) and the related idea of scaffolding. The latter attracted a great deal of attention and has become an umbrella term with a certain amount of controversy which was addressed in an entire issue of the Journal of the Learning Sciences (2004, vol 13, 3). Vygotsky conceptualised the ZPD as:

> "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (Vygotski, Vygotsky, & Cole, 1978, p. 86)

We will go into more details about scaffolding in the second part of this chapter, however this has great importance in learning design and learning technology as the role of the teacher (or automated system) is different from the rules-bound solution emerging from the behaviourist perspective both in terms of content and type of intervention.

Within this framework it is possible to identify three strands of research in psychology relevant to the argument of this chapter: one is specifically exploring learning per se, the other, which we can call instructional psychology (see Resnick 1981 for a review) can be considered as an area of psychology applied to education. Instructional psychology is also very similar, as an applied branch of research, to human-computer interaction (see Carroll,

---

[4] ACT (Adaptive Control of Thought) and the revised ACT-R are cognitive architectures developed by J. Anderson and heavily influenced by cognitive psychology and neuroscience. The core aim is to explain human behaviours in context, by specifying the way cognitive processes works, the *modules* involved and the processes using a symbolic programming-like system.

1997; Olson & Olson, 2003 for reviews) and both are relevant for this thesis. Furthermore, there are two fundamental reasons why the *source* of learning is essential in this argument. Generally speaking, and certainly in the past century, with the introduction of schooling, most of the learning is achieved via formal instruction in a class setting. The content is standardised by the curriculum and the scope for personalised instruction in such setting is limited as the teacher/instructor must ensure an equal distribution of resources and that attention is devoted to each student.

The second aspect removed from the equation in class instruction are the concepts of *motivation* and *readiness*: in the domain of learning technology, Malone (1980) and Prensky (2006) are some of those who strongly criticise the failure of ‘edutainment’ compared to entertainment, with the former failing to achieve similar levels of interactivity and engagement compared with video games, even though the technology available is the same (something should be said about budgetary differences in the two).

2.1.3. Situated cognition & the social dimension of learning

In the 1980s and 1990s, a widespread disillusionment with the formal methods of cognitive science and the relatively slower progress in artificial intelligence, compared with the pace of miniaturization and technological innovation as well as robotics, encouraged a new paradigmatic shift. The first generation of cognitive scientists had to recognize that the models (cognitive schemes or knowledge) which computational models of cognition afforded, contributed to provide a new interpretational window, but were severely limited.

> “It was widely thought that good problem solving and other intellectual performances reflected general strategies (supported by g) operating on whatever database of knowledge happened to be needed. True ability resided in the general strategies, with the database an incidental necessity.” (Perkins & Salomon, 1989, p. 17)

Research evidence, however, showed that there was no such thing as a general heuristic able to operate outwith the domain of implementation. Cognitive approaches pushed a shift from the observed association between stimulus and response to a knowledge-based understanding of processes. However, there is much more in common between behaviourism and cognitive approaches than might first appear:

> “Cognitivism, like behaviourism, understands knowledge and learning as resulting from experience within a stable, objective world. Community and culture can enter into cognitivist theory only insofar as they are decomposable into discrete elements that can participate in the stable, objective realm of experience. Thus, the opportunity to explore learning and knowledge as processes that occur in a local, subjective and socially

> construed world is very limited by both behaviourists and cognitivist paradigms. (…) What situated cognition theory promises as a next step is a model for dealing with knowledge and learning as fundamentally social and cultural, rather than an artefacts of an individual's journey through as impersonal and objective world." (Kirshner & Whitson, 1997, p. VII)

Clancey defined situated cognition as:

> "a philosophical perspective and engineering methodology that acknowledges the value of descriptive models of knowledge as abstractions but attempts to build robots in a different way. (…) the theory of situated cognition claims that when modelers equate human knowledge with a set of descriptions, such as a collection of facts and rules in an expert system, they are describing abstractly how the program should behave in particular situations, but they are not capturing the full flexibility of how perception, action and memory are related in the brain" (Clancey, 1997, p. 3)

As well as for artificial systems, the impact of situated cognition approaches in education and instructional design was twofold: on one side studying learning in context meant a more specific attention to metacognitive and self-monitoring processes. On the other it allowed identification of a social and cultural valence of learning which was previously ignored, but becomes essential in understanding specific processes and outcomes. Laurillard put forward a core criticism of situated cognition in education questioning the differences between academic knowledge and everyday knowledge and how the two are represented in human experience. In her view there is:

> "a distinction between natural environments which afford the learning of 'percepts' in everyday life, and unnatural environments which are constructed for learning 'precepts' in education. Situated cognition makes the same distinction in arguing that the one type of environment should emulate the other, but does not elaborate on the nature of the difference." (Laurillard, 2005, p. 20)

Laurillard exemplified her reasoning by advocating that:

> "We cannot experience structuralism in the same way as we experience good table manners. (…) Because we have to rely on the artificial structuring of our experience of precepts, via academic texts, for example, it is unlikely that the mechanisms we use in the natural environment will transfer directly to this unnatural environment. Thus, our means of access to precepts becomes critical to our success in learning them" (Laurillard, 1987, p. 202)

Volet and Järvelä (Volet & Järvelä, 2001) proposed a multi-dimensional cognitive-situative model of understanding for students' motivation and learning. The basic framework considers the 'person-in-context' in which the student's cognitions, motivations and emotions are interacting with the 'learning affordances' which are the social and physical spaces in which learning occur. Between these two factors there is one's 'experiential interface': according to this model, the most productive learning is realised if there is a congruence, or in other terms, if the individual is ‘tuned’ to the affordances of the learning context and if the community of practice supports the individual engagement in learning (Volet and Järvelä 2001). Such representation is an attempt to pull together the advantages of the two perspectives and rather than presenting them as antithetic views, it leads toward a unified approach.

2.1.4. Integrative views: experiential and interactionist approaches

The differences in the approaches presented thus far, emerging particularly in the various disciplines, were already observed in the 1980s. In his review of the contrasting explanation of learning Entwistle noted:

> “(...) it is important to recognise that all the explanation provided, be they behaviourist, psychometric, or humanistic, (…) are all partial in both senses of the word. Each theorist is focusing on a limited range of learning situations, and has collected a restricted range of evidence. (…) it is inevitable that each theory retains an element of subjectivity derived from the person’s own individual experiences and beliefs.” (N. Entwistle, 1987, p. 27)

Such diversity and contrasting evidence, however is useful to provide a complex characterization of learning. In a fairly recent integrative effort, Illeris (Illeris, 2002; Jarvis, 2006) talked about two fundamental processes of learning, namely *acquisition* and *interaction*. These are organised into three dimensions of learning: cognitive, emotional and social dimension, which can be thought as the vertices of triangle representing the space in which learning occurs. Figure 2.1 demonstrates how the processes and dimensions are integrated.

In his definition the cognitive dimension has to do with learning the content and may be described as knowledge or skills which builds up the understanding and ability of the learner. In his words:

> “The endeavour of the learner is to construct meaning and ability to deal with the challenges of practical life and thereby develop an overall personal functionality.” (Illeris 2005, in Jarvis 2006, pg. 91)

The emotional dimension is the one that deals with mental energy, feelings and motivations; most of the inspiration comes from Freud’s and Rogers’ work as well as recent research in differential psychology. The social dimension encompasses the external interaction and specifically involves participation, communication and interaction. According to this view, psychology explored predominantly the cognitive dimension and only recently started to explore the emotional aspects, however the social influences have been largely derived by thinkers like Marx and Feuerbach. In his original work functionality, sensibility and sociality were expressed with the popular reference to the domains of Piaget, Freud and Marx.

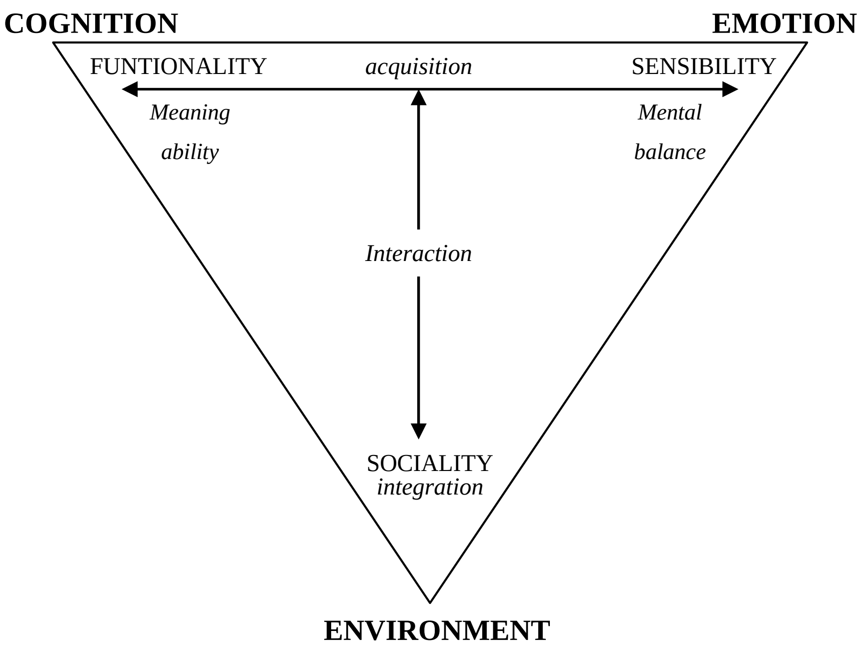


Figure 2.1. The processes and dimensions of learning.Adapted from Illeris 2005, in Jarvis 2006.

The recent work conducted by Wenger (1999) is a prototypical example of research right in the middle of the ideal triangle formed by the three vertices and his views on learning as a process of formation of meanings within a community of practice is clearly identifying the fundamental tension between these dimensions. In the British context, Jarvis explores learning from a philosophical perspective and highlights the tension between internal and external starting from a Cartesian perspective and embed the experience of learning in the wider socio-cultural context of western civilisations (Jarvis 2006).

2.1.5. Quick summary

In this section we explored some selected theories and perspectives on learning from both psychology and education. A greater emphasis was put on the implications of taking a stance from one of the different perspectives, particularly considering potential consequences for instructional design. The core message to be taken into the following discussion is that no single perspective is satisfactory enough to understand the interactive nature of learning from instruction. However, an integrative approach should be taken to evaluate how effective students are when interacting with instruction. In complete agreement with Ramsden:

> "(...)we can never assume that the impact of teaching on student learning is what we expect it to be. Students' thought and actions are profoundly affected by the educational context or environment in which they learn. (…) Good teaching involves striving continually to learn about students' understanding and the effect of teaching on it." (Ramsden, 1992, p. 6)

Hence, in the next section, we will shift the perspective and focus on students, in the attempt to identify useful approaches to understand students' learning and studying.

2.2. Understanding students' learning and studying

Understanding how students learn is very much dependent on what kind of interpretational approach is used and the focus can be broadly categorised based on three core dimensions: on the cognitive mechanisms which afford learning, on the emotional and motivational aspects (internal or external) or on the social and interactive aspects (with the context, teacher, peers or material). In this section we shift the focus on another aspect that will be expanded further in the next two chapters: what are the possible sources of individual variation in performance?

A very interesting phenomenon in the research mentioned thus far is that the two strands of psychology of learning and instructional psychology have developed in parallel and mostly independently from one another. Entwistle, with specific attention to the role of the teacher, asserted that:

> "one of the difficulties teachers have had with psychological explanations of learning is that most psychologists have concentrated exclusively on the characteristics of the individual. (…) Learning, for the teacher, is inseparably linked to the classroom context". (N. Entwistle, 1988)

It is however possible to draw a parallel between the type of learning theories and their applications and the table 2.1, is an attempt to make the observations more systematic.

Theories	Main characteristics	Potential e-learning applications	Literature
Behaviourism	Focuses on behaviour modification via stimulus-response pairs; Trial and error learning; Learning through association and reinforcement; Pedagogical focus is on control and adaptive response; Focus on observable outcomes	Much of current e-learning development represents little more than transfer of didactic approaches online, the 'web page turning mentality' linked directly to assessment and feedback	Skinner, Tennant, Bloom, Bennett, Gagne
Cognitive (information processing)	Focus on internal cognitive structures; views learning as transformations in these cognitive structures; Focus on human development Pedagogical focus is on the processing and transmission of information through communication, explanation, recombination, contrast, inference and problem solving; Useful for designing sequences of conceptual material which build on existing information structures	Salomon's notion of distributed cognition (Salomon, 1993) could lead to a more shared knowledge structure between individual and surrounding information rich environment of resources and contacts; Development of intelligent and learning systems, and the notion of developmental personalised agents	Piaget, Bruner, Norman, Ausubel, Resnick, Anderson, Hutchins, Wenger
Constructivist	Focus on the processes by which learners build their own mental structures when interacting with an environment; Pedagogical focus is task-orientated; Favour hands-on, self-directed activities orientated towards design and discovery; Useful for structured learning environments, such as simulated worlds; construction of conceptual structures through engagement in self-directed tasks	The concept of toolkits and other support systems which guide and inform users through a process of activities could be used to good effect to embed and enable constructivist principles; Access to resources and expertise offers the potential to develop more engaging and student-centred, active and authentic learning environments; Microworlds and simulations	Papert, Bruner, Duffy & Jonassen, Entwistle
Activity-based	Focus on the structures of activities as historically constituted entities; Action through mediating artefacts within a framework of activity within a wider socio-cultural context of rules and community; Pedagogical focus is on bridging the gap between historical state of an activity and the developmental stage of a person with respect to that activity e.g. current state of language use and child's ability to speak a language; The Zone of Proximal Development – the idea that assessing current ability gives limited insight into an individual's potential for development, which is better studied through examining their work alongside a more able peer	In the last decade there has been a shift from a focus on the information (and in particular content) aspects of ICT to an emphasis on communication, collaboration and understanding the factors which underpin the development of communities; In particular there has been a realisation that the development of content alone does not lead to more effective learning, and that there is a need to structure and foster learning environments to enable communities to develop Networking capabilities of the web enable more diverse access to different forms of expertise and the potential for the development of different types of communities	Vygosky (1934), Wertsh (1985), Engerstom (1987)
Socially situated learning (or interactionist)	Take social interactions into account and learning as social participation Emphasis on interpersonal relationships involving imitation and modelling Language as a tool for learning and the joint construction of knowledge Language has two functions: 1. As a communicative or cultural tool, used for sharing and jointly developing knowledge 2. As a psychological tool for organising our individual thoughts, for reasoning, planning, and reviewing our actions Dialogue between tutor and student can be articulated into 12 levels of engagement – both external and internal Knowledge is a matter of competences with respect to valued enterprise. Participating in the pursuit of this, i.e. active engagement, Meaning our ability to experience the world and our engagement with it as meaningful – is ultimately what learning is to produce	Multiple forms asynchronous and synchronous communication offer the potential for more diverse and richer forms of dialogue and interaction between students and tutors and amongst peers, as well as the use of archive materials and resource for vicarious forms of learning; Different online communication tools and learning environments and social for a offer the potential for new forms of communities of practice or facilities to support and enhance existing communities	Mercer, Vygostsky, Laurillard, Lave, Wenger
Experiential	Experience as foundation for learning Learning as the transformation of experience into knowledge, skill, attitudes, values emotions Reflection as a means of transforming experience Problem based learning a focus: Experience: Problem situation, identification and definition; Gather and reflecting on information; Theory formation and test in practice; Experience through Primary and Secondary; Reasoning and Reflection; Evaluation (Dewey, 1916)	Asynchronous communication offers new forms of discourse which is not time-bound and hence offers increased opportunity for reflection Archive and multiple forms of representation of different communications and experiences offer opportunities for reflection	Dewey, Kolb, Jarvis, Rogers, Marton, Covington
Systems theory	Focus on organisational learning, or on modelling the development of learners in response to feedback	New forms of distribution and storage, archiving and retrieval offer the potential for development of shared knowledge banks across organisations and forms of organisational distributed cognition Models of learning account adaptation in response to both discursive and active feedback	Senge, Laurillard

Table 2.1. Learning theories, concepts and applications: a synthesis from Entwistle 1987 and Conole et al. (2004).

It is clear from this summary that attention to the general theory rather than the individual has been the main focus of the research reported. Already in 1957 Cronbach, identified a clear divide between two strands of research in psychology:

> “While the experimenter is interested only in the variation he himself creates, the correlator finds his interest in the already existing variation between individuals, social groups, and species.” (Cronbach, 1957, p. 671)

In the next two chapters we will expand on the idea that in the context of instruction both approaches have a crucial role to enrich the evaluation of implementations of e-learning. However, here we would like to explore some of the possible sources for variation in students’ performance specifically in relation to instruction and learning technology. Focussing on specific models of learning will therefore be useful to evaluate both the learning process and the outcomes.

2.2.1. A matter of ‘alignment’

Cognitive process analysis of the like of Resnick (Greeno, Collins, & Resnick, 1996) has been applied to a variety of instructional tasks from simple addition and subtraction to learning geometry and computer programming providing a great insight into the way students learn. There are however different opinions about what makes it ‘optimal learning’. Gagné’s work, for example, was very specific in the attempt to identify an ‘optimal’ sequence which should characterise instruction. He identified eight types of learning, which are typically characterised by a four-phases process. The eight types of learning, namely signal learning, stimulus-response learning, chaining, verbal association, discrimination learning, concept learning, rule learning and problem solving, were a departure from the simple task analysis proposed in the previous two decades by classic behaviourists: for the first time Gagné included higher order processes such as problem solving (Gagné 1965).

The model he proposed was quite appealing in its simplicity: a stimulus situation was followed by an apprehending and acquisitions phase (constituting the core learning steps via simpler processes such as attending, perceiving, coding and acquiring), which were followed by a storage and retrieval phase (characterised by retention, recognition, recall, reinstatement and transfer) leading to an observed performance of some sort.

Instructional Event	Internal Mental Process
1. Gain attention	Stimuli activates receptors
2. Inform learners of objectives	Creates level of expectation for learning
3. Stimulate recall of prior learning	Retrieval and activation of short-term memory
4. Present the content	Selective perception of content
5. Provide "learning guidance"	Semantic encoding for storage long-term memory
6. Elicit performance (practice)	Responds to questions to enhance encoding and verification
7. Provide feedback	Reinforcement and assessment of correct performance
8. Assess performance	Retrieval and reinforcement of content as final evaluation
9. Enhance retention and transfer to the job	Retrieval and generalization of learned skill to new situation

Table 2.2. The relation between instructional events and internal mental processes.

The hierarchical and sequential nature of learning was the most important aspect of Gagné's interpretation. For example prior mastery of a prerequisite intellectual skill has to be achieved to readily accomplish the learning of new intellectual skills (positive transfer). Not having a particular skill does not completely prevent learning, but that the learning is not facilitated. Moving in a class context, such differential preparation is what causes the differences in performance between students who may, or may not, already have the prerequisite skills. We will come back to this idea of the learning hierarchy at a later stage.

Gagné created a nine-step process of instructional events, which addresses the 'conditions of learning'. Table 2.2 shows these instructional events in the left column and the associated mental processes in the right column. An explicit attempt is made to relate the instructional events to the specific mental process/state.

From a developmental psychology perspective, the strong Piagetian influence of a stage model for learning seems to be essential in setting the prerequisites which inform such a sequence. However, Bruner (1961) taking a more constructivist stance, defended the idea of a 'spiral curriculum' and was very convincing in presenting the case in which, with adequate transformation, or adaptation of the topic to the logic suitable for the child, any topic of the curriculum could be taught at any stage of development. The influence of what the teacher does (or must do) to accommodate the learner is quite clear.

Shuell, on the other hand, asserts that:

> "(...) if students are to learn desired outcomes in a reasonably effective manner, then the teacher's fundamental task is to get students to engage in learning activities that are likely to result in their achieving these outcomes, taking into account factors such as prior knowledge, the context in which the material is presented, and the realization that students' interpretation and understanding of new information depend on the availability of appropriate schemata. Without taking away from the important role played by the teacher, it is helpful to remember that what the student does is actually more important in determining what is learned than what the teacher does." (Shuell 1986, p 429).

Merrill, after reviewing theories of instructional design, went even further contending that there are five prescriptive principles common to most theories of instruction:

> "(a) Learning is promoted when learners are engaged in solving real-world problems. (b) Learning is promoted when existing knowledge is activated as a foundation for new knowledge. (c) Learning is promoted when new knowledge is demonstrated to the learner. (d) Learning is promoted when new knowledge is applied by the learner. (e) Learning is promoted when new knowledge is integrated into the learner's world." (Merrill 2002, p. 43)

These antithetic views about the role of instruction (in the figure of the teacher) and the role of the student emerge from the parallel development of theory about learning and theory of instruction from the research in psychology. In a very lucid assessment of the state of education (and higher education in particular), John Biggs (1999) summarised the interaction between teaching method, level of engagement and student orientation in the figure 2.2 below which represents two idealised students.

> "Susan, who is academically committed, bright, interested in her studies and motivated to do well. Robert, on the other hand is not as committed nor prepared, he is enrolled to obtain a degree, and possibly not even in the subject of choice." (Biggs & Guildford, 1999, p. 4)

The author acknowledges the fact that students like Robert are entering higher education in higher proportion than in the past, but he readily criticises a dismissive approach of the teachers blaming the student's lack of motivation and suggests that the challenge for the teacher is to close the gap. In his words, "Good teaching is getting the most students to use higher cognitive level processes that the more academic students use spontaneously" (Biggs 1999, pg 4)

Looking back at the table 2.1, the simplest way to categorise the way in which the theories differ is to look at the degree and type of control envisaged for the teacher/instructor and the role of students' engagement.

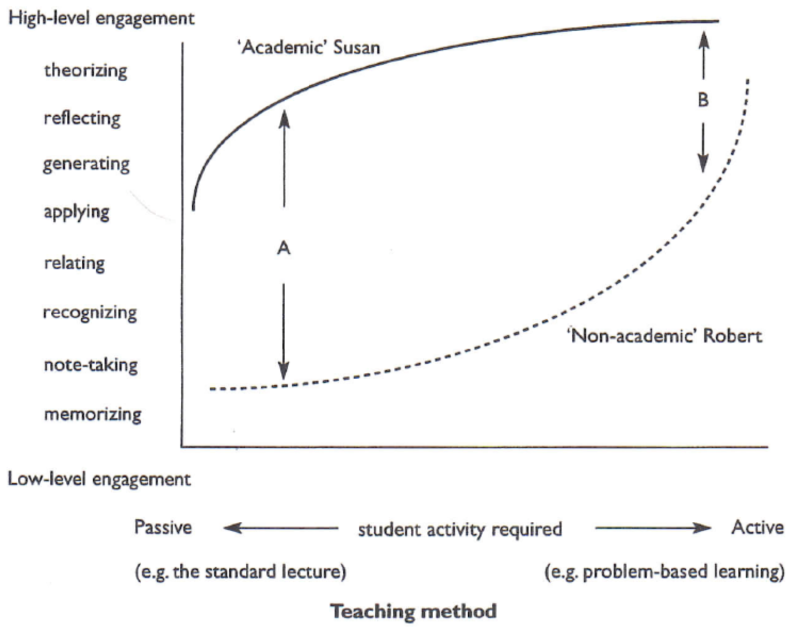


Figure 2.2. Student orientation, teaching methods and levels of engagement (from Biggs 1999)

At one extreme the teacher decides what is to be learned, writes instructional material in line with narrowly defined behavioural objectives and uses systematic (and objective) tests to ensure that learning is fulfilled. At the opposite end complete freedom of learning is advocated in which students take control of their learning, whilst the teacher acts as facilitator. The difficulty, as intimated also by Entwistle, is that:

> "These two extremes represent not just differing psychological theories deriving from contrasting sets of data, but also contrasting philosophical positions on the nature of both man and education." (Entwistle, 1987)

With a firm practical perspective Laurillard believes that "the best expression of an empirically based teaching strategy (…) is an iterative dialogue between teacher and students focused on a topic goal. (Laurillard 2005, p. 77)

Such dialogic teaching strategy is grounded in the progression through four distinct aspects: discursive, adaptive, interactive and reflective.

> "The strategy is undeniably prescriptive, but aspires to prescribe a form of interaction between teacher and student, rather than action on the student. In this way, it provides a structure capable of its own improvement" (Laurillard 2005, p. 78).

Assessment is the catalyst in the interaction between teaching strategies and learning, funnelling outcomes in the form of academic performance; assessment is also the externalisation of a teaching method and in higher education we are familiar with strategies which in combination are believed to satisfy most of the requirements. Lectures and readings are supposed to further acquisition of knowledge. Exercises and problems promoting practice, seminars and tutorials should stimulate discussion and practicals foster discovery. However, students respond differently and as Ramsden intimated, the core question really is "What do we want students to learn? What are the variations in the outcomes of their learning?"

2.2.2. A differential psychology perspective

Before we can explore the relations between academic performance, personality and intelligence in the following chapters, we need to provide a simple starting point for further discussion.

In 1963, Carroll postulated a model according to which five variables could account for the variability in academic performance: aptitude, the opportunity to learn and perseverance can be used as expressions of time investment. For example, aptitude is "the variable or variables that determine the amount of time a student needs to learn a given task" (Carroll, 1993a, p. 26). The other two are related to achievement and are the quality of instruction and the ability to understand instruction. Over the years the Carroll model was tested in a number of studies and the predictions found ample support; however, some like Shulman (1986), advocated that these variables oversimplified matters and other emotionally-related variables should be taken into account, but Bloom (1968), on the other hand, based most of his model for mastery learning on Carroll's work. This model is still providing source for inspiration in education and instructional design because at the core there was a statement that:

> "Most students (perhaps over 90 percent) can master what we have to teach them, and it is the task of instruction to find the means which will enable our students to master the subject under consideration" (Bloom 1968, p. 1).

This powerful promise is even more popular in today's higher education, in which, as identified earlier, it is more likely to find a wider range of abilities and attitudes. The message

coming from Bruner, Bloom and later from Block, was that appropriate teaching methods, particularly those which allow students to take their time to master a specific objective before progressing to a higher order level (achievement should be evaluated with criterion-referenced tests rather than norm-referenced tests), and based on a coherent taxonomy of educational objectives, are the best ways to account for individual differences.
However, a very different set of premises is given by Bruner:

> “There is not one kind of learning. It was the vanity of a preceding generation to think that the battle over learning theories would eventuate in one winning over all the others. Any learner has a host of learning strategies at their command. The salvation is in learning how to go about learning before getting irreversibly beyond the point of no return. We would do well to equip learners with a menu of their possibilities and, in the course of their education, to arm them with procedures and sensibilities that would make it possible for them to use the menu wisely.” (Bruner 1985, p 8)

This quote mentions two important aspects: we have already acknowledged that one of the reason why we have such diverse interpretations of learning is that different theories use different learning situations and rely on different data. However Bruner also suggests that because learners have different strategies to go about learning, instruction should provide a number of options to suit the students’ needs.

A similar view was also taken by Marton & Saljo (Marton & Saljo, 1976), Entwistle (Entwistle, 1988; Entwistle & Ramsden, 1983) and Biggs (Biggs, 1987; Biggs, 1996). Taking an experiential and phenomenographic approach to learning, Marton & Saljo investigated the ways in which students normally go about learning. Using a mix of questionnaire-based studies as well as interviews and qualitative data analysis they identified three modalities or approaches to learning: a deep, surface and strategic approach. These approaches, or learning styles (a clarification on the terminology will be provided in the chapter 4) were found to be quite important in determining the quality of students’ learning, to be related with academic performance, and also affected by the modes of teaching.

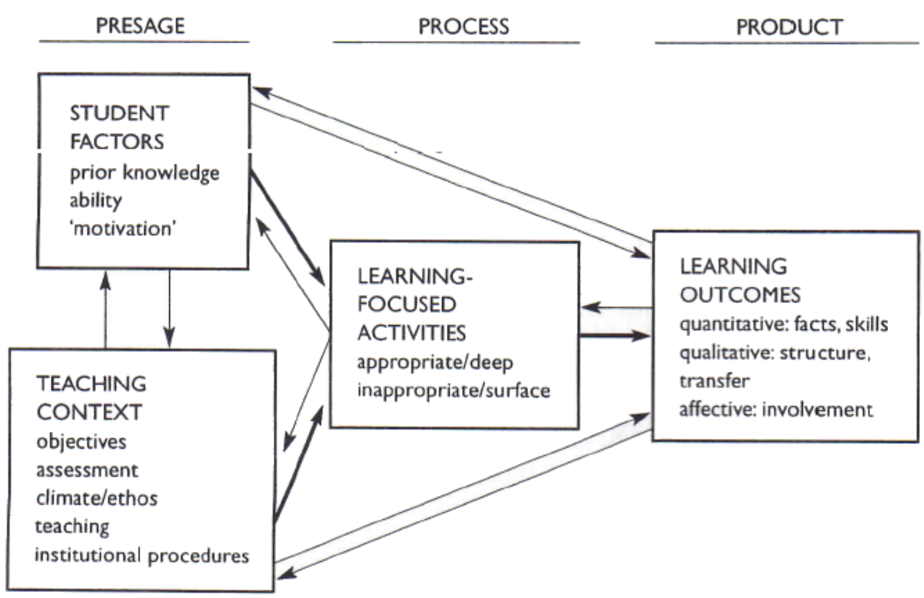


Figure 2.3. Bigg's 3Ps model (from Biggs 1999, p18)

The relevance of styles and differential aspects of learning has a significant impact on instructional design. For example, in his 3Ps model, Biggs indicated that there should be an alignment between the various stages of teaching and learning. In the 3Ps model (see figure 2.3 above), Biggs identifies three stages: presage (before learning occurs), process (during learning) and product (the outcome of learning).

Volet and Jarvela (2001) argued that a congruence between the student and the affordances of the context has to be maintained for optimal learning. What emerges from this discussion is that knowing more about students and about what are suitable strategies to match students individual differences are core aspects of good teaching and, as we will see in the next section of this chapter, such knowledge is also fundamental for the successful implementation of learning technologies.

2.3. Instruction and Learning Technology

2.3.1. Technology and education: a brief overview of the terms

After considering some ideas from the literature on learning in psychology, it is essential to understand what is intended and revisit the term *learning technology* as this is used in association with a plethora of other terms and concepts. For example, ICT (information and communication technologies) is also used to refer to the broad range of technologies used in education. This is often synonymous with learning technologies, which refers more specifically to the practical application of technology for learning and teaching. Conole & colleagues (2007) emphasized that the difference in terms is caused by a young field of research in continuous flux, but it is also an outcome of the very different perspectives taken in the fields of education, computer science, artificial intelligence, system engineering and ergonomics as well as human-computer interaction and psychology which partially contribute to the research in e-learning.

To date, there are at least a dozen different terms used commonly in the literature which are associated with learning technology: computer assisted/aided learning (CAL), computer assisted/aided instruction (CAI), computer-based training (CBT), computer-based learning (CBL), computer-supported learning resources (CSLR), technology mediated learning (TML), technology-based learning (TBL), technology-based training (TBT), Web-based training (WBT), Internet-based training (IBT) intelligent tutoring systems (ITS), learning management system (LMS), content management system (CMS), virtual learning environment (VLE) as well as e-learning, m-learning and blended learning, collaborative learning etc.

This *vocabulary soup* is continuously enriched with the addition of new tools and advancements, but the use of these terms is often unclear, and most terms can be considered as synonyms. In this thesis we will use the terms interchangeably, maintaining the terminology used in the primary sources rather than attempting an integration, which will facilitate the interpretation of the concepts, but we will consider them all under the umbrella of *learning technologies*. However, as indicated in the first chapter, wherever possible learning technology will be used specifically focusing on the medium and e-learning as the process of learning with technology.

One of the most common mis-conceptions about e-learning arises from the idea that e-learning has been associated with *distance education* for a long time. This is reflected in a number of reviews which attempted to make systematic observations about the impact of e-learning on academic performance (i.e. Orr 1997, Russell 2001, Bernard et al., 2004). Even in recent publications (Conole et al., 2007) the difference is still fuzzy and ill defined. The fact is that the use of ICT on- and off-campus, is very different. The same tools are used in different ways and for different purposes and the effectiveness of applications and implementations must be contextualised in the evaluation process.

Only recently some scholars started to talk about blended learning (Heinze & Procter, 2004) when referring to the particular case of a teaching model in which ICT is not substituting traditional teaching methods, but is offered to *enhance* or *augment* traditional techniques. In their words: “Blended learning is the effective combination of different modes of delivery, models of teaching and styles of learning.” This however is still not as specific as the definition of the Department for Education and Training in which blended learning is intended as “learning which combines online and face to face approaches” (DET 2003). This is a fundamental observation as it implies that the use of technology is not intended to be as a substitute to traditional education like authors coming from the strand of distance education might imply. We will endorse this view throughout the thesis.

E-learning should focus on learning, and learning technology offers a range of powerful tools which should become part of the toolbox of every practitioner. However, to evaluate modes of implementation and effectiveness of these tools, we will argue that it is essential to focus on three dimensions: pedagogical, organizational and technical.

The pedagogical aspects are fairly obvious, but it is important to highlight the fact that if a new ICT tool doesn’t have a specific pedagogical reason to be used, or it is not able to provide concrete, evidence-based benefits, then there is no real reason to introduce (or push) it into the curriculum. One of the most notable cases in recent times is the one of e-portfolios, which some might argue doesn’t add anything new to pedagogy, but technology facilitates its implementation (Cooper & Love 2007, in Buzzetto-More & Alade, 2006).

The organizational aspect is somewhat more complicated: in the first chapter we highlighted the frictions between top-down institutional changes and a pedagogy-driven approach in which practitioners and teachers demand the adoption of technology to support their teaching. This is a dimension which cannot be ignored: political pressure and economic demands have an essential interplay in the adoption and diffusion of ICT throughout the education system of a country and these have an obvious impact at the local and institutional level, as we noted in the case of the University of Edinburgh.

The technical aspect, instead, is the most controversial: on one hand there is a issue of technical expertise; on the other, one of pedagogical innovation. For example, is it the responsibility of the teacher to keep up with technological innovation (therefore maintaining a high technical expertise) to be able to use the best possible methods of teaching to deliver content, or should it be an organizational priority to offer the expertise at the service of teacher and pedagogists based on their teaching needs?

2.3.2. Technology in education: early adopters and good practice

No matter where, when, or how e-learning has been implemented there are two elements which are common: early adopters challenge the usual way of delivering instruction and often producing a ripple effect in cognate disciplines. In this section we will explore the fairly wide landscape of ICT and instruction and will attempt to show a shift from the system to the user (learner), which characterised implementations across the disciplinary spectrum, particularly in the past ten years. If one tries to build an ideal timeline of educational and learning technology, it would probably start around 300.000 BC with the use of cave paintings as a form of permanent communication. Some argue that because of their fairly inaccessible positions, these paintings did not have a purely decorative value. The introduction of the abacus and a variety of writing methods helped to provide the grounds for a shift from oral to written culture and transmission of knowledge. Such small steps continued throughout the centuries until the 18th century when technological innovation and scientific discovery suddenly increased exponentially the rate of change with today's pervasive communication tools, immediate information and just-in-time methods of communications which afford tele-presence or immersive environments. (a dynamic timeline using web 2.0 technology can be found here: http://www.xtimeline.com/timeline/History-of-Educational-Technology-2)

The most relevant sub-section of the timeline, however, is without doubt the period starting from the 1950s. In the first part of this chapter we already mentioned the teaching machine advocated by Skinner. The affordances of computing and miniaturization which lead to the introduction of personal computers and more recently mobile technologies, as well as communication devices from the telegraph and radio to satellite communications and the Internet in 1990s, allowed for a swift change in the way teaching and learning was envisaged and its role in the society. We have already seen how psychology of learning and education have developed in parallel to artificial intelligence with its promise of automated instruction.

One of the key issues mentioned then was that practitioners and instructors could not find an easy or universal answer to achieve better instruction from psychology or computing. However the impact of technology is now widespread, changing not only the way teaching is achieved, but to a large extent, forcing education to change because of the way we conduct our lives.

An obvious place to start understanding the context and address *evaluation methods* is in computing science, a discipline in which there have been a number of attempts to: 1) implement automatic, or machine-driven instruction; 2) communication technology is usually exploited first and 3) formal models to understand and interpret learning have been proposed.

In computing science and artificial intelligence, automated systems and CAI (computer-assisted instruction) or CBT (computer based training) have always been centre stage. The core idea was that because computers are incredibly more capable than humans in carrying out computations, it was assumed that then they would also be superior to human in teaching and adapting to the needs of learners. In practice this is still an illusion. Despite the massive progress in automation, computers can only address well-defined problems in specific contexts and struggle to cope with the complexity of ill-specified, but more realistic situations.

In 1989 some, like Ross & Morrison argued that:

> "(...) instructional technology researchers are still struggling, it seems, to find a foothold in identifying meaningful research questions and accepted paradigms for investigating them." (Ross, Morrison, & O'Dell, 1989, p. 19)

Issues regarding external validity, media replication and learner control have taken centre stage since the 1950s. As we learnt from the first section of this chapter, however, this was not so much because there was no progress, more because of the lack of a unitary framework to merge the vast amount of knowledge accrued in psychology and education with the technological know-how to implement instructional systems. (See Conole et al. 2007 for a commentary).

Oliver (Oliver, MacBean, Conole, & Harvey, 2002) also lamented that in most of the published literature practitioners draw from theoretical constructs (like constructivism for example) without explaining how their implementation embodies the principles and values of the approach taken in the design.

TML (technology mediated learning) has received a mixed reaction since 1997. A heated debate between Orr (1997, pro-technology) and Russell (2001, against it) sparked the interest of researchers as well as practitioners. Russell continues to maintain the popular ‘No significant difference phenomenon’ website, in which he collected over 400 papers which seemed to demonstrate that the use of technology made little or no difference to instruction.

Although McNaught (McNaught, 2003) attributed the lack of application models to the fact that practitioners, outside the field of education, struggle to make sense of the vast array of theories and models, Conole and colleagues (Conole, Dyke, Oliver, & Seale, 2004) thought that the use of ‘toolkits’ might help to make the application of formal models more systematic providing a clear and simple framework for applications. One solution is in the recent introduction of the *learning technologist* in educational institutions in an attempt to provide expertise at the crossing between disciplines, but often these professionals don’t have the necessary domain knowledge to maximise the impact of the implementations.

Ross and Morrison see this as an advantage:

> “These diverse characteristics of researchers appear to offer the study of instructional technology the advantages of varied talents, perspectives, and experiences with alternative research paradigms(…) At the same time, varying degrees of training (or commitment to) scientific research methodology and learning theory raise concerns about the preparedness to formulate the questions and find the answers that will bring about significant advancements.” (Ross & Morrison 1989, p20).

By their own admission, most of the research in instructional design had a predominantly behaviourist stance. Toward the end of 1980s this has become increasingly cognitive in nature (i.e. Bonner 1988), but has not changed much since. Scholars with a background in psychology or education looking at recent textbooks on instructional design are often disappointed in the realisation that the key theories reported are obsolete or inaccurate in their respective disciplines (an example is Gagné, Wager, Golas, & Keller, 2005, 5th edition). More recent efforts (i.e. Rothwell & Kazanas, 2008, 4th edition) make a specific attempt to integrate and contextualise human performance. Figure 2.4 shows the extent to which cognitive theories are contextualised in a more comprehensive model than the one proposed in Gagné et al. (2005). Multiple editions of these popular academic textbook, highlight how, even 20 years after the commentaries of Ross & Morrison (1989) or Driscoll (1984), it is still difficult to find a coherent framework or to identify a suitable and appropriate evaluation methodology.

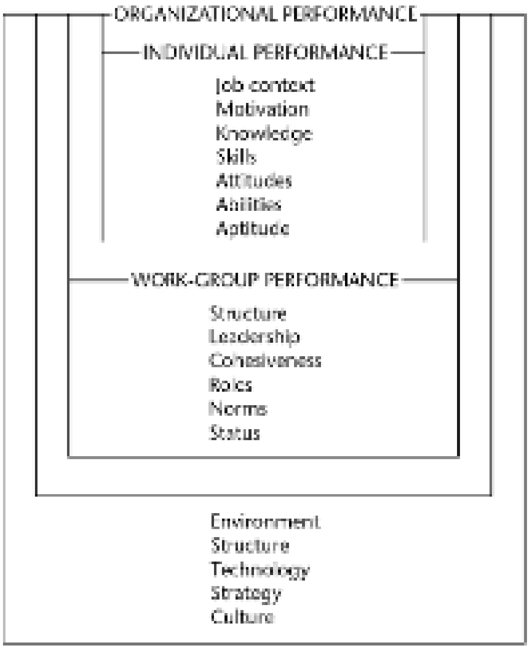


Figure 2.4. A comprehensive framework for instructional design.

Two prolific areas of research in the last 20 years have been medicine and computing science. Medical practitioners view learning technologies as an essential tool to support the development of the curriculum and promote a life-long approach to learning (i.e e-portfolios and problem-based teaching). Results are mixed with educational, instructional and psychological exploration of the tools (many presented in the Association of Learning Technology Journal published since 2004 and Computers and Education journal).
In IT and computing sciences, the tradition is more on mastery learning (a stream of systems have been successfully implemented to learn programming and system design from Logo to LISP and from basic electronics to completely automated security systems in airplanes).
In more recent times, the focus has shifted to enhance the curriculum with ‘soft’ or transferable skills (i.e. team work, communication, project management). A similar focus on learning technology has also occurred on GSS (Group support systems) which can support such dynamics in the class (i.e. wikis and blogs).

2.3.3. Instruction and e-learning: implementation models

Taking for granted the importance of technology in education, it is very important to consider that good course design should separate ends from means.

> "We are constantly making the mistake of specifying the means of doing something rather than the results we want. This can only limit our ability to find better solutions to real problems." (Gilb & Finzi, 1988)

Forms of e-learning vary according to the use and the role of technology as a tutor, as environment supporting learning and interaction or as a tool.

Evaluation does not always make a good link between methods and pedagogy. This means that we do not always get a good idea of the learning objectives or how learning took place e.g. Hiltz (Hiltz & Turoff, 2002). In contrast good examples of linking evaluation to an explicit model of pedagogy are provided by Duffy & Jonassen (1992) and Fowell and Levy (1995).

Dimension	Characteristics	Some issues for evaluation	Examples
Structure	Timetables, ground rules, tutor assessment, tutor direction, requirement to log on, assigned roles lead to high structure	Coherence of curriculum design? Match between design and learners preferred styles? Gap between intention and practice	Clarke (2002) an example of very low structure, Aviv, R. Erlich, Z., Ravid, G., Geva, A. (2003) provide two examples in a problematic paper
Communication and / or content	Content heavy - extensive predetermined course material, Communication heavy - high exchange of ideas and information	Communication reich leads to focus on messages, message analysis and relationship of learners to messages; Content heavy leads to study of imformation processing	McConnell, D. (2000) a good example of high communication low content
Independance / collaborative learning	Collaborative learning often involves teams discussing cases or scenarios and constructing a group response, cooperative collaborative contested terms here.	Collaborative leads to focus on group processes	Lockhorst, D., Admiraal, W., Pilot, A. and Veen, W. (2002) discuss case based approaches, Tsui, and Ki, (2002) discuss an open forum with independent contributions Dempster and Blackmore (2002) around student web publishing and critical analysis (see TELRI model below)
Online / off line	On line learning suggests that most learning literally takes place at the machine	much neglected - how do we capture what happens away from the machine	Henri and Pudelko (2003) discuss different types of online community Chen and Hung raise questions over personal and group knowledge

Dimension of Instruction	Definition
time	the timing of instruction and pace of learning and studying
space	physical location, meetings, class
technology	equipment provided to support teaching and learning
interaction	amount and type of interaction with instructors and peers
control	pace and way of engaging with instruction & material

Table 2.3. Two ways of mapping the dimensions of instruction to online
Sources: Dempster (2004) and Piccoli et al. (2001)

The summary table (2.3) is useful to identify some core dimensions of learning technology, its application to pedagogy and instruction and some issues relevant to evaluation purposes. There are overlaps between the two with the concepts of time and space in Piccoli et al. (Piccoli, Ahmad, & Ives, 2001) fitting the structure dimension in Dempster (2004).

However, the way in which e-learning satisfies the pedagogical goals is a rather different issue. McLuhan famous statement that ‘the medium is not the message’ comes to mind. Although not exhaustive, two different views on how learning technology can be integrated with pedagogy are suggested.

Laurillard (Laurillard, 2002, 2005) offered examples of how learning and teaching methods, media/tools contribute to a didactic or dialogic relationship between the lecturer and the students (Figure 2.5). Her conversational framework identifying the activities necessary to complete the learning process allows us to classify the use of learning technologies in the context of a specific model of instruction and therefore support the idea that e-learning is just another evolution of the medium rather than a different way of learning

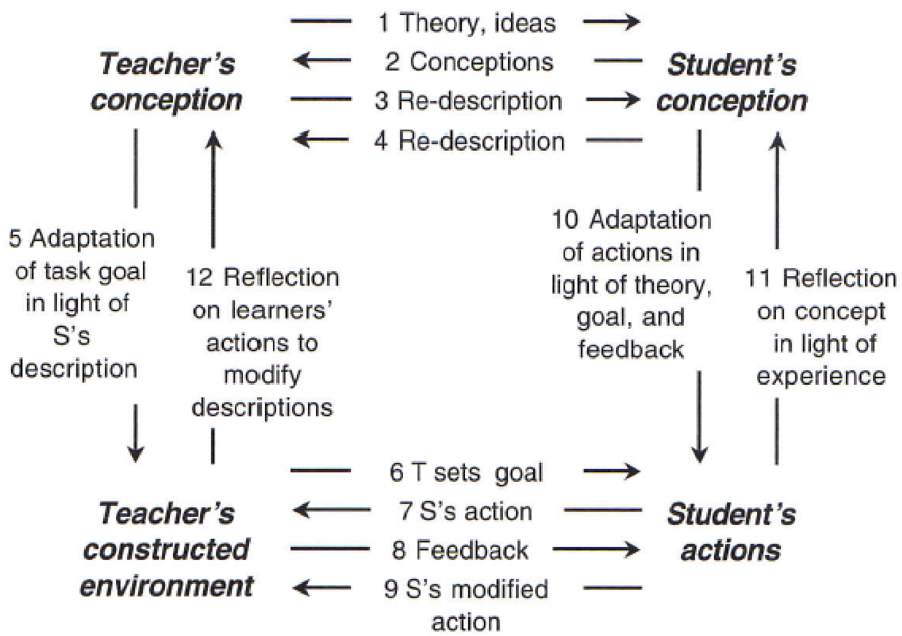


Figure 2.5. Laurillard’s conversational model of teaching from Laurillard 2003, pg 87

Using this model produces a course design profile by mapping the tools used to resulting learning outcomes based on the different kinds and directions of dialogue supported.

DeVilliers (De Villiers, 1999; De Villiers & Dersley, 2003), on the other hand, proposed what she called the Hexa-C metamodel. According to this integrative approach the various theories of learning can be interfaced to technology. The colourful representation of the Hexa-C model is appealing for a number of reasons: on one hand it attempts to pull together the core elements of very different traditions from psychology of learning, education and instructional design. In DeVilliers words,

> "No single paradigm is appropriate – 'no one size fits all' – but its elements can be translated into principles, design guidelines, and evaluation criteria for different domains and subject matter." (DeVilliers, p. 19 in (Buzzetto-More & Alade, 2006))

The second issue arising from this model is that the responsibility to ensure that technology is serving its purpose falls upon the instructional designer, e-learning practitioner or educational web developer who should make sure that "technology serves as a hub that delivers the message and does not distract or detract from the message". (ibid, p.19)

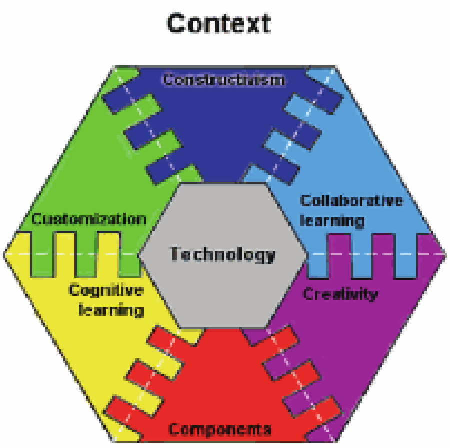


Figure 2.6. The Hexa-C model (from DeVilliers 1999) in which the technological element of e-learning directly 'plugs in' to the various aspects of design required for a successful application.

This is very different from Laurillard who implied that the teacher should be an expert in instructional design as well as knowledgeable in the domain to be taught and has obvious implication for the evaluation of the system at a macro, organizational or institutional level. Yet a different model is proposed by Conole et al. (2004, fig 2.7) with the octahedron presented in their toolkit for mapping e-learning to pedagogy which has only three orthogonal dimensions: Information vs Experience, Individual vs Social, and non-reflective vs reflective.

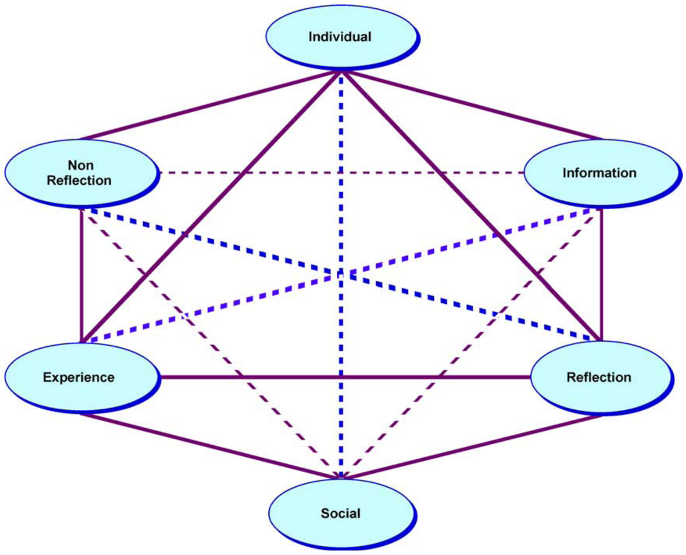


Figure 2.7. A toolkit to map learning theories and pedagogy in e-learning (from Conole et al. 2004). This representation allows to visualise the complex interactions between aspects of pedagogical theory.

According to this model any learning activity can be mapped using a combination of these simple dimensions and demonstrates its usage in relation to the theories (see table 2.4, for an example: Laurillard's conversational model is mapped to the octahedron focusing on the "non-reflection" bubble: Tutor describes concepts, Tutor Student Dialogue, Tutor adapts concepts, Tutor sets task, Student completes task, Dialogue on action, Student reflection)

Activity	Indv.–Social	Non-refl.–Refl.	Expr.–Info.
Brainstorming			
Seminar	--------X--	--X--------	--X--------
Online discussion	------X----	--------X--	--X--------
Online chat	------X----	--------X--	--X--------
Using a concept map	--X--------	--------X--	--X--------
Presentation of material			
Lecture	-----X-----	--------X--	---------X-
CAL tutorial	-X---------	------X----	---------X-
Searching the Web	-X---------	--X--------	-----X-----
Peer presentation	---------X-	--X--------	-----X-----
Assessment of level of competence			
1-to-1 tutor discussion	--------X--	--X--------	----X------
Peer assessment	--------X--	--X--------	-----X-----
CAA tool	--X--------	--------X--	---------X-
Marked assignment	-X---------	------X----	--------X--

Table 2.4. of activities to the three dimensions of the toolkit

Source: (Conole et al. 2004, p 30).

The toolkit allows easy classification of activities and even offers a quantitative measure to attribute and represent the value of each activity. Contrary to the previous models presented, the toolkit is very much a practical tool which enables practitioners to provide an explanation of the learning theories used, a mechanism to identify the key learning characteristics of the activities implemented (not necessarily in e-learning, as demonstrated by the table 2.4), and to provide a systematic tool for evaluating curriculum design.
An element which does not feature at all in this model is the consideration of the outcomes or the process of learning, which limits its usability to the design and implementation but leaves a big gap in the evaluation.

The take-home message of this section is therefore simple: even if there is evidence from a number of successful implementations across the spectrum of disciplines that a variety of tools, as well as the way in which they are used, can enhance the value of learning and teaching in higher education, it is paramount to individuate an appropriate evaluation strategy which considers the full cycle from the instructional theory, the implementation and consideration of if, and how, the learning occurs. This is what we attempt to achieve in the next section. Figure 2.8 takes the integrative framework proposed by Illeris and attempts to enhance it by specifying the element of interaction taken from the models cited in this section and offers a useful starting point to think about what should be considered when an e-learning system implementation is evaluated.

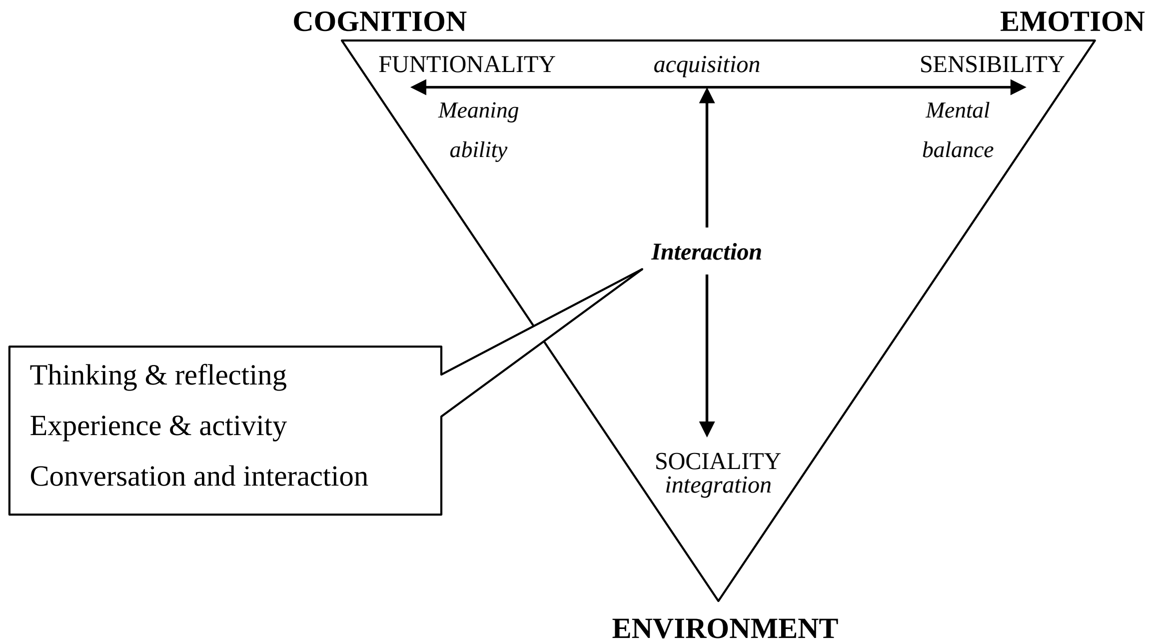


Figure 2.8. From the processes & dimensions of learning to a specification of the interaction in e-learning

2.4. E-learning evaluation framework and its limitations

Even though we now take computers for granted, it is important to remember that personal computers brought rapid technological advancements in an incredibly short time span. To understand the methods of evaluation of e-learning it is necessary to provide a brief overview of methods coming from both a system engineering and from a human-computer interaction points of view.

2.4.1. Evaluation frameworks

In engineering research, the most popular evaluation method is the DeLone and McLean model of information success. First published in 1992, the model was proposed as

> "a way to synthesize previous research involving IS [information system] success into a more coherent body of knowledge and to provide guidance for future researchers" (Delone & McLean, 2003, p. 10)

The model identified the multidimensional nature of IS success and most importantly that the selection of success measures should be contingent on the objectives and context of the empirical investigation. The original model and its updated revision of 10 years later are reproduced in figure 2.9.

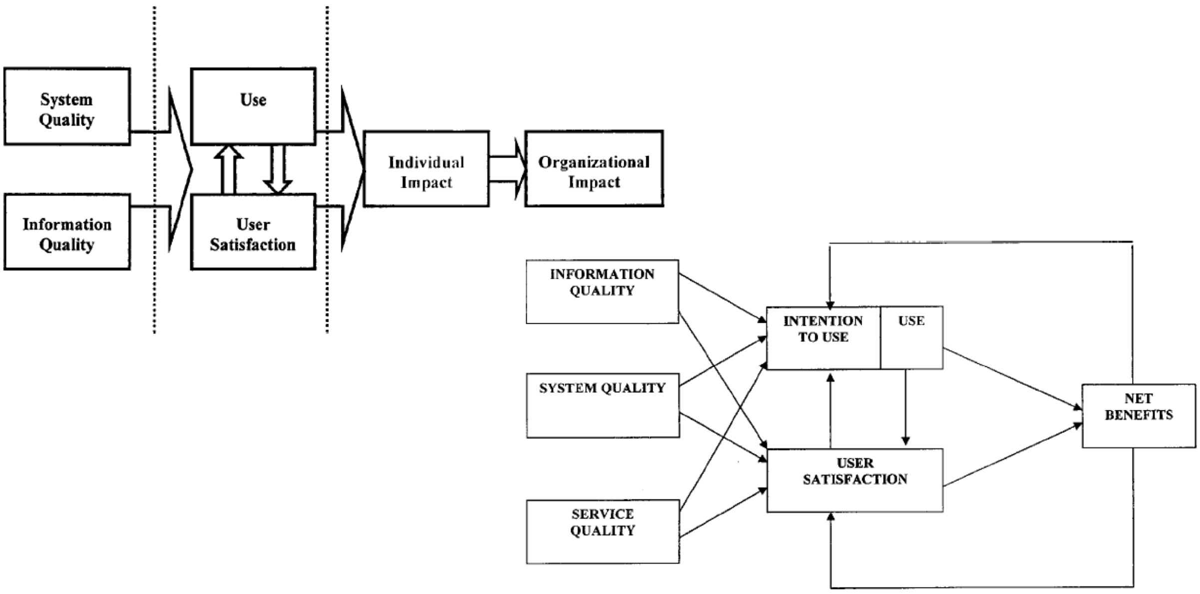


Figure 2.9. De Lone & Mclean IS success models: the original model from DeLone & McLean 1992) and the revised version (from DeLone & McLean 2003)

Any implementation can be assessed on a number of dimensions including “systems quality” which measures the technical success, “information quality” measuring semantic success or the success of the information of conveying the intended meaning, “use, user satisfaction and individual impact” and “organizational impacts” are the measures of effectiveness. In DeLone & Mclean review, the model was reported in 285 peer-reviewed articles in the 10 years since its first publication and even though the core principles have not changed, its revision included some specific measures of quality and the clear distinction between use and intention to use following the criticism by Seddon (1997) who tested empirically the temporal-causal path implied by the model and commented on the term ‘use’:

> “One should not assume that greater IS use per se is a good thing (although greater perceived usefulness and user satisfaction probably are)” Seddon 1997, p. 251)

In the 10 years since the first publication of the model, usability engineering also made great progress and there has been a dramatic shift from a system-centred design (or what the system can do) to a user-centred design, which focuses on the interaction between the user and the system and the involvement of the user directly in the design stages of the system (Nielsen 2003). Evaluation plays an essential role at all stages.

From the existing, and quite extensive literature on evaluation, it is possible to synthesize that there are at least four major dimensions by which evaluations can be characterized: (a) internal versus external validity, (b) formative versus summative, (c) qualitative versus quantitative, and (d) micro versus macro.

Proponents of internal validity argue for tight experimental control from which valid explanations of the specific phenomenon can be made (Cook, Campbell, & Peracchio, 1979; Shadish, Cook, & Campbell, 2002). The external validity perspective focuses more on the consistency of the findings across settings and the reproducibility.
The aim of the evaluation is also the essential. Harvey (Harvey, 1998) provided a very succinct and effective metaphor:

> “When the cook tastes the soup, it is formative evaluation; when the dinner guest tastes the soup, it is summative evaluation.” (Harvey et al. 1998, p.7).

The choice of techniques classed as quantitative and qualitative is a common ground for debate in the social sciences. Without going into this issue, it is generally agreed that hybrid or eclectic models enable the evaluators to determine appropriate methods and match them to specific questions and that qualitative data can often enrich the quantitative data.
The final dimension has to do with the scope and breadth of the analysis, and based on the focus (i.e. a single course/training programme or an institutional change) evaluation can have a confirmatory value or a policy changing impact for an organisation (Horton 2001)

2.4.2. An overview of evaluation processes and approaches.

In the previous section we mentioned the “Evaluation cookbook” compiled by Harvey and colleagues (1998) as part of a big project funded by the Scottish Higher Education Funding Council which included research from a variety of different universities in the UK.
In their review they identified a number of practical tips suitable to investigate the effectiveness of learning technology. In their “recipes” they list 16 different methods to support evaluation projects.

In the table 2.5 we reproduced their summary (on the left), which provides a good idea of the time cost for the different stakeholders, but we also expanded it specifying the main focus of each of the techniques used (on the right). This exercise provides evidence that a mixture of techniques is more appropriate depending on the aim of the evaluation.

Overall, it is possible to identify three main perspectives in the evaluation of e-learning; these are shaped by the relative emphasis placed on learning, theory, and evidence as the primary reason for undertaking an evaluation (European Evaluation Society, 2002).
A learning perspective supports the premise that stakeholders should learn from an evaluation. The basic thinking is that the evaluation should have a specific goal, which should also drive the evaluation from a research point of view. Alavi disputed that

> “Most would agree that the objective of using technology in learning should be to positively influence learning in one way or another, that is the student should either learn something that he/she would not have learned without the technology or learn it in a more efficient way.“ (Alavi & Leidner, 2001, p. 4)

A theory-based perspective maintains that evaluative information must be interpreted within a theoretical framework. In this sense, appropriate methodologies and the selection of adequate measures are essential to obtain useful research data.

Investigation methods	Time investment					Focus				
	preparation time	students' time	admin time	analysis	additional resources	Student, learner	System usability	Content	Instruction	Learning
Checklists	low-moderate	low	low	low	low		V	V	V	V
Concept maps	low	low	low	low	low			V		V
Confidence logs	low-moderate	low	low	moderate	none	V	V			V
Cost effectiveness	moderate-high	none	none	moderate-high	none				V	V
Designing experiments	high	low-moderate	low-moderate	low	low		V	V	V	V
Ethnography	low	low	high	high	moderate	V				V
Focus groups	low	moderate	moderate	low-moderate	moderate	V	V	V	V	
Interviews	moderate-high	moderate	high	moderate-high	moderate	V				
Nominal group techniques	low	low	low	low	low	V				
Pre- and post- testing	high	moderate	moderate-high	moderate	low				V	V
Questionnaires	moderate	low	low	moderate	none	V	V	V	V	V
Resource questionnaires	low	low	low	moderate	none	V	V			
Split-screen video	moderate	low	moderate	moderate	high		V		V	
Supplemental observation	low-moderate	moderate	moderate	moderate	moderate		V		V	
System log data	moderate-high	low	low	moderate	moderate		V	V	V	
Trials	moderate	high	moderate	moderate	low-high		V	V	V	

Table 2.5. Summary of methods to evaluate e-learning. Summary of methods to evaluate e-learning.

Focus and time investment of various types of investigation methods to assess the quality of e-learning implementations.

Evidence-based evaluation tends to use experimental methods in an attempt to determine what actually works. In this case it is essential to focus on the experimental design to identify suitable measures characterizing the outcomes.
The identification of key questions reflective of the needs of the stakeholders, decisions on appropriate constituencies, methodology, measures, data collection methods, and designs are the essential choices faced when an evaluation has to be done. These components must be adaptable because evaluation should be iterative and cyclical in nature (Mandinach, 2005).

Sheard and Markham (2005) pointed out that both Harvey et al. (1998) and the more recent Phillips et al. (2000) guides to evaluation of e-learning systems are very prescriptive and do not actually provide a specific *diagnostic* structure that allows the evaluator to explore the particular tasks/activities.

However Phillips (Phillips, Bain, McNaught, Rice, & Tripp, 2000) later proposed a model in which evaluation is mapped on the phases of development which takes into account the objectives, the stakeholders and gives hints on the use of specific methods at the different stages.

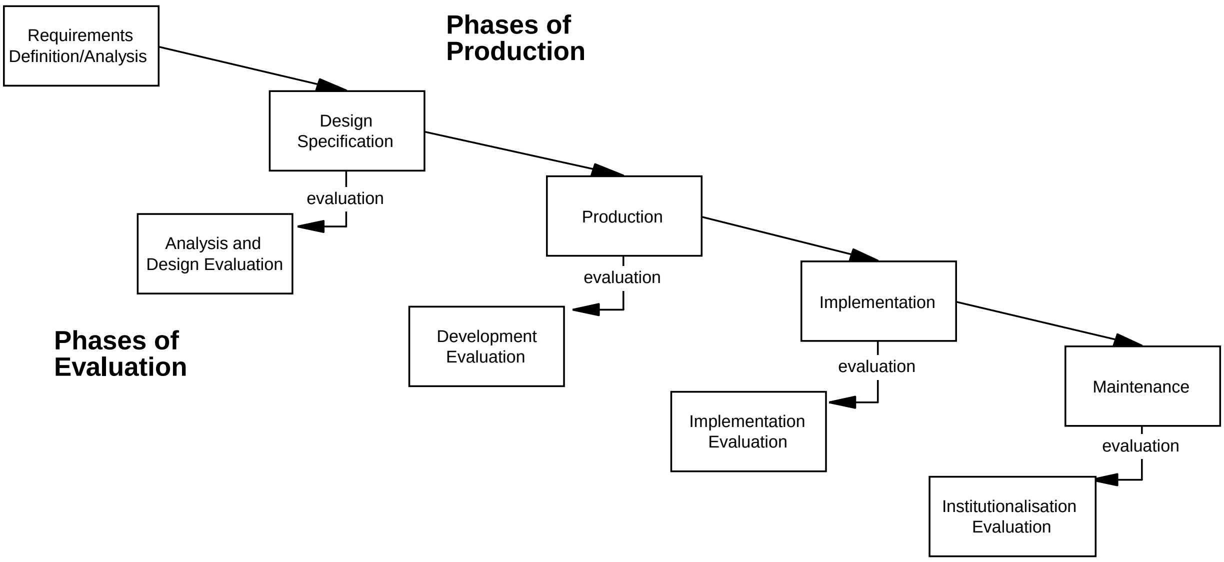


Figure 2.10. Mapping of the phases of an e-learning production process to corresponding evaluation phases.

One of the most difficult aspects of evaluations is to establish the parameters which the evaluation is supposed to use. Lorenzo & Moore (2002) listed ‘Five pillars of quality online education’: Learning effectiveness, student satisfaction, faculty satisfaction, cost effectiveness and access. Even if these were formulated predominantly for online and distance courses, the five dimensions can be easily juxtaposed to the DeLone & Mclean model of IS success. In this sense, it is possible to narrow down from the system engineering perspective to a the domain-specific implementation of e-learning.

A review and commentary of research on e-learning effectiveness can be found elsewhere (i.e. Holmes & J. Gardner, 2006; Horton, 2001; Levy, 2007). However, at this point in time, three interpretation models for researching the factors contributing to successful e-learning environments seem to be influential: Alavi & Leidner (2001) framework for technology-mediated instruction, Piccoli et al. (2001) and Johnson (Johnson, Hornik, & Salas, 2008). All are rooted in the tradition of the IS success model and can be integrated as iterative expansions. In particular Alavi et al. stressed the importance of course design in affecting the learning outcomes and the most effective outcomes occur when technology and pedagogy are integrated. Piccoli et al., on the other hand point their attention to the importance of the skills and characteristics of both students and instructors as the agents shaping the value of effectiveness measured by learning outcomes, satisfaction and self-esteem. Johnson et al. re-address some of the issues of failure of e-learning, especially in distance learning, by expanding on the concept of computer self-efficacy (personal dimension), social interaction and social presence, to stress that the medium is providing very different opportunities for interaction than the face-to-face classroom and such dimension is mediating outcomes.

In the context of this thesis, we borrow from Johnson et al. (2008) a specific conceptualization of effectiveness of learning technology as measured by:

- course performance (objective learning outcomes, discussed more in chapter 4)
- course instrumentality (students' belief about value of training and intentional engagement with the course as expressed by their behaviours rather than end-of-year feedback)
- course satisfaction (personal perception, only for group data)

Following on from the benchmarks suggested above, it is important to understand that the conceptualization of *effectiveness* is a complex task. To help map the conceptual framework to the data collected, it is useful to consider the following aspects:

- human dimension (computer self-efficacy, motivation and individual skills & characteristics)
- design dimension (user satisfaction, usability)

- perceived usefulness (what students think about the system)
- interaction (behavioural proxies of usage and communication)
- social aspects (engagement & communication)
- learning (performance measures as learning outcomes)

These will be specified in more details in chapter 5 and 6 and considered in the data evaluation in chapter 7 and 8.

2.5. Individual characteristics, e-learning and online behaviours

In this chapter we have already seen how individual skills and characteristics come into the frame of integrative models both for understanding learning and instruction and for evaluating the effectiveness of e-learning. Individual characteristics are closely related to the ability to learn and the propensity to take learning opportunities when these are offered. Already at the start of the last century Thorndike observed:

> "If, by a miracle of mechanical ingenuity, a book could be so arranged that only to him who had done what was directed on page one would page two become visible, and so on, much that now requires personal instruction could be managed in print."
> (Thorndike, 1911, p. 165)

This appears to be exactly what e-learning and learning technology seems to promise. However, even though most people conduct their lives immersed in technology, it has been noted that certain individual characteristics make some people use and work with technology more, and more efficiently than others.

It is useful to summarise some of the aspects which have been reviewed in the literature specifically in relation to learning technology: intellectual ability, personality, experience and self-efficacy with the medium (Johnson et al. 2008), personal willingness (Ertmer & Newby, 1996; Silber, 1998), learning styles (Kolb, 1984), work commitment (Tracey, Hinkin, Tannenbaum, & Mathieu, 2001) and motivation (Mathieu, Tannenbaum, & Salas, 1992).

The nature of individual differences affecting the interaction with technology will be expanded in the next two chapters, however we can organise such differences in a more systematic way, which will become useful in the treatment of the data collected. Zmud

(Zmud, 1979) classified personal characteristics related to technology use into three groups: demographic (age, sex and training for example), personality-based (attitudes, motivation), and *cognitive styles* (the way in which one handles information). We will focus in greater details on the research on intellectual and personality differences in the next 2 chapters, however, this fairly generic and broad distinction is useful to introduce the pragmatic issue of where the data actually comes from.
In the section on evaluation methods, for example, we have seen that the methods in the summary table can be classified in three types: observational (i.e. ethnographic or split screen methods), self-reported (i.e. questionnaires) or interactive (i.e. focus groups, interviews). Sheard (Sheard, Ceddia, Hurst, & Tuovinen, 2003) noted that in a conventional teaching environment it is a common practice to obtain feedback with end-of-year surveys and via continuous face-to-face interaction wherever possible. When students work in an electronic environment, however, this direct observation is not possible. In fact, by design, learning technologies are usually equipped with logging tools which track students' activity and produce automatically a very detailed trail of what each student does in the e-learning environment.

As Romero & Ventura (Romero & Ventura, 2007) pointed out, however, partially because of the complexity of the data extraction, partially because of the shear amount of data generated, this invaluable source has been largely ignored, and it is only in the last few years that a growing interests in the activity data has emerged. Another contributing element has been a swift development of data mining (Mobasher, Dai, Luo, & Nakagawa, 2002) and its application to e-commerce and web-technologies (Cooley, 2000; Srivastava, Cooley, Deshpande, & Tan, 2000).

Web Usage Mining (WUM) has seen a rapid expansion because of the incredible growth of e-commerce: managers and CEOs suddenly wanted to find out what customers actually do when they visit a website, but, as we will see in chapter 6, both data extraction and interpretation is still in its infancy. The application in the domain of education is an obvious application of WUM: learning technologists want to know *what* students are using, instructors want to know how effective is what students *do* online, faculty members want to know if their investment has an adequate return. Data mining can offer yet another source to understand how technology is affecting learning and instruction and it is a core mission of this thesis to relate the behavioural data emerging from the *virtual observation* of students' activity in the courses considered with a number of other, more traditional methods, with the ultimate aim to evaluate the effectiveness of e-learning.

2.6. Chapter summary

In this chapter we set out four goals: 1) select a number of useful theories from psychological research with a certain relevance to education, instructional design and learning technology; 2) review the terminology in reference to learning technology and e-learning and contextualise the use of e-learning for instruction in higher education, 3) present the current state of evaluation methods for learning technology; and 4) suggest the necessity of a differential perspective to approach the variability of students' learning.

Starting from behaviourist models we proposed an integrative approach which took into account learners from a cognitive, emotional and social perspective and reviewed some ideas about how instructors could use these factors into account to maximize the effect of instruction.

We also made the case for a differential psychology approach to learning as the combination of motives, preferences and abilities often produces unexpected results. It is necessary to recognize that learners might take very different paths to achieve their learning, some more effective than others, but this has to be evaluated considering not only the learning outcomes, but also the process to achieve such outcomes.

The review of evaluation methods of learning technology has also been presented to highlight some current practices, each of which provides only a partial view of the cost-benefit analysis. Considering a differential approach is essential to support other methods. We introduced WUM as another possible source to be included in the evaluation. As a very new field of investigation, exploring behavioural data using data mining techniques will offer a unique opportunity to portray the students' interaction with the system which can then be related to more traditional data collection techniques. This will offer a greater depth in both the evaluation of e-learning and our understanding of learning in HE.

Chapter 3. Understanding the need for students' profiles: a differential psychology perspective

In the last chapter we looked at learning mainly from a psychological and educational perspective and attempted to put the role of e-learning into context, noting the weaknesses of current evaluation methods and stressing where there are interesting overlaps between different approaches.

Despite the trends explored in the first chapter, which identified a tendency toward a *massification* of education, the belief that there is scope for *personalised* instruction in the modern educational system is the most important observation. E-learning could, in principle, fulfil one of its goals of delivering a more personalised learning experience grounded on strong empirical data.

In fact, e-learning systems do already store great amount of data about their users and their behaviours. Such data, as we will see in later chapters could provide an excellent window to a better understanding of *what*, *how* and *when* students are accessing the resources offered to them.

However, it is important to take a step back from expressed behaviours and consider if the presence of strong precursors, or preconditions typical of particular individuals might ultimately affect behavioural expression. It is essential to provide a more appropriate context to venture into the explanation of the reasons why students produce specific patterns of behaviour and we believe that a better knowledge of the individual differences between students may provide such contextualisation.

Research in individual differences has traditionally distinguished between *distal*, or dispositional trait-like constructs, and *proximal*, situational state-like constructs (Kanfer,

1990). In this chapter and the next one we will carefully consider research conducted in differential psychology to provide a theoretical framework supporting the creation of rich individual profiles. These are afforded by an in-depth analysis of students' intrinsic abilities, preferences, beliefs and motivations, and should encompass the whole spectrum from distal to proximal.
The statement that some individuals are more talented, gifted, bright, and clever or simply better suited to perform in specific situations or tasks does not raise much opposition. However, identifying in *which way*, or how an individual student is *better* than another (or performs better in higher education) is an entire different issue and can be controversial because of the value-laden nature of this representation.

In this chapter we tackle the question of what *kind* of student benefits the most from the interaction with education and will consider aptitude and abilities as two possible precursors to achievement. Achievement in the educational context is measured as academic performance (AP).
Students' performance at university might be affected by a variety of issues. Many are outwith the domain of academia and affect students' personal lives (i.e. adjustment to independent life, income and financial pressure, workload and time management, personal relations and exploration of social interactions etc). However it would be too ambitious to include everything related to student experience to account for the variations in academic success. Therefore, in the first section we will consider AP as an objective metric of achievement and contextualise this metric in a social and educational framework.

We will then look at the relations between individual characteristics and AP. We first explore the traditional association between AP intelligence followed by the analysis of the correlations personality and achievement. To make sense of the literature we will apply a meta-analytic approach allowing us to overcome the apparent inconsistencies between studies.
Finally we will attempt to interpret the relationships between intelligence and personality measures showing a complex theoretical framework.

The exclusive use of both intelligence and personality measures will be ruled out as inadequate to provide a *useful* characterization of a student. These, in fact are too deterministic to be pragmatically useful and the complexity of the interactions might actually confuse, rather than simplify the results.

For this reason, we will suggest that measures of style (which we will consider in detail in the next chapter), could be an alternative option because: 1) they draw from the full spectrum of distal to proximal constructs, 2) they identify personal modes of operation and preferences, rather than level of intellect or desirable traits of personality, and 3) allow to tap into a *typical* rather than *optimal* performance (Ackerman, 1996). These should provide a more useful characterization of the differences in the learning outcomes and usage behaviours.

3.1. Academic performance: social and educational issues

In the UK, the key criterion enabling progression to university is academic success measured by grades in secondary schools. Based on this metric, one could argue that taking into account pre-entry performance measures (grades) recorded at various assessment points (like in (Daniel Harris, 1940), it is possible to predict that good students will be able to obtain good marks and mediocre students will not improve. This prediction is based on the assumption that AP is not drastically affected by any other factor, and indeed, as we will see later in this chapter this might be a fair statement if AP represents ability and intelligence accurately. On the other hand, if we take a different view of ability and consider as an example an athlete progressing from local, to national and Olympic qualifications, we can easily grasp that not everyone will be able to win the gold medal.

Both examples put ability at the core. Although the latter example is taking for granted that ability is not infinite and there might be variation in quantity, which is highlighted by the selection process, the former example is quite cynical and openly critical of the *utility* of education.
Discussion about the scope of education goes beyond the scope of this thesis, however, driven by an implicit belief that education *should* contribute to personal development, encompassing personality, and not only intellectual abilities, the underlying argument poses a legitimate question about both the *context* and the *kind* of students who *can* succeed in the current educational framework.

The changing nature of education is a very important aspect. Until the mid-1800s instruction was an individual experience: pupils were taught individually by expert tutors and only the privileged few had access to such instruction. Class instruction is only a (relatively) recent introduction that was adopted to allow more efficient education in small communities. Mass instruction, with class sizes too big to allow for an individualised interaction with each student, is even more recent, taking the form of lectures, in which a transmission model, rather than one based on interaction between tutor and students has taken over instruction. Assessment of what has been learnt is therefore a core aspect of modern education. However, the concept of assessment requires some further elaboration as historically the term has been used in rather different ways.

Gipps (1999) identified two core aspects justifying assessment: selection and certification. Binet and Simon (Binet & Simon, 1905; Goddard, 1908) described a set of higher order mental tests which could be used for three classificatory purposes: medical (focusing on physiology and pathology), pedagogical (to determine intelligence, based on knowledge), and psychological (directed to observe intelligence based on abilities).

Standard examinations, however, are not completely new; examinations first appeared in China under the Han dynasty (206 BC to 220 AD) with the specific intent of selecting candidates for government service. Testing and exams have been adopted in western culture only around the 19th century when access to the professions shifted from the value of family history and patronage to the value of academic achievement and ability. Universities started to set up entry tests and examining boards in the 1850s and School certificates were introduced in the UK by the start of the 20th century. With more children brought into compulsory education, the publication of Spearman's and Binet's seminal works attracted the attention of those responsible for the efficient running of the state education system and the concepts of psychometric intelligence and IQ had obvious implications for the teaching models implemented. It should be noted that Spearman and Binet were critical of each other's work, especially regarding the development of theories of measurements for abilities and the way these develop over time. Gipps is critical of the social value of assessment and used the words of a sociologist to express her resentment:

> "Broadfoot (1996), argues that assessment in developed societies with mass education systems, whether for selection or certification, has a single underlying rationale: to control mass education and the nature of its goals and rewards. It operates to distribute, in a justifiable way, social roles that are not all equally desirable. Individuals are allowed to compete on an equal basis to demonstrate their competence. The provision of an apparently fair competition allows those who are not successful to accept their own failure (thus controlling resentment among the least privileged) and acquiesce in the legitimacy of the prevailing social order. Broadfoot cites IQ testing as a means of

> social control 'unsurpassed in teaching the doomed majority that their failure was the result of their own inbuilt inadequacy' (Broadfoot, 1979, p. 44)." (cited in Gipps 1999)

From an educational perspective, academic performance is not the centre, but merely the outcome of a process in which assessment *should* be used only as a tool to aid learning (i.e. Black, 1993). Especially in the more recent literature the focus shifted from the properties of specific forms of assessment to the effects that such forms have on the learning experience (l Black & William, 1998). Already back in the 1960s, Glaser (1991) promoted a shift from a psychometric type of testing to a "criterion-referenced measure". Glaser indicated that such methods could be used to assess student entry-levels and to determine the extent to which students had acquired the behaviours an instructional program was designed to teach. We should point out the stress on the "behaviours". Even if the idea was to move away from the psychometric value of testing, it is very difficult to eradicate old and well established beliefs that there is an objective test able to assess a specific skill or ability: this originates from associationism (Thorndike, 1922) and behaviourism (Skinner, 1963)and was applied in education and curriculum design with the strong influence of Gagné (1965), who is also considered one of the key figures in learning technology. The concept is grounded in classic learning theory:

> "The whole process of becoming competent in any field must be divided into a very large number of very small steps, and reinforcement must be contingent upon the accomplishment of each step. This solution to the problem of creating a complex repertoire of behavior also solves the problem of maintaining the behavior in strength (...) By making each successive step as small as possible, the frequency of reinforcement can be raised to a maximum, while the possibly aversive consequences of being wrong are reduced to a minimum." (Skinner, 1963, p. 94).

As identified by Shepard (2000) this view has had clear consequences upon curriculum design, forcing the instructors to identify teaching strategies which promote a tightly sequenced and hierarchical learning and to seek an objective form of assessment which would allow not only a "test-teach-test" form of instruction, but also a scaffold to motivate students (with grades) to achieve their potential. Gagné himself talks about learning hierarchies and goes as far as describing hierarchies as a 'route' for learning a topic. Instructors should determine (or diagnose) if subordinate abilities have been acquired by all students as a prerequisite for successful learning. (Gagné 1965) His examples of plans of instruction in math, science, English and foreign language acquisition showed how it is possible to create a systematic and rules-based approach to instruction and it constitutes one of the main reasons why this approach was popular in learning technology in which such rules acquisition could be automated.

This approach is highly criticised in more recent literature, and especially in the past 20 years researchers in education are more likely to take a constructivist approach to teaching and learning in which intellectual abilities are socially and culturally developed and learning is considered a process of constructing knowledge and understanding in which deep understanding is the mean allowing knowledge transfer (Entwistle & Ramsden, 1983; Entwistle & Entwistle, 2003) .

In this perspective, assessment should be designed to elicit higher order thinking processes and promote understanding rather than focusing on the benchmarks of achievement (David et al., 2009; Gibbs & Simpson, 1999). The greater attention to *formative* rather than *summative* assessment practices as better tools to aid the learning process is a focal point of very new research (i.e. REAP - Re-engineering assessment practice- project, (Nicol, 2006; Nicol & Macfarlane-Dick, 2006), but in practice grades remain the most essential aspect of educational practice as an indicator of performance and achievement. Interestingly, however, the pedagogical function attributed to specific forms of assessment is often not aligned with student's perceptions of what the assessment stands for. Instead, grades are what drives students' efforts and, to a great extent, also affects students' sense of achievement, success and, as a consequence, affects their perception of the self and satisfaction which characterise the experience of individual students.

John Biggs (Biggs, 1996) is one of those who stressed that teaching and learning can be improved through *constructive alignment*. A decade earlier and using a strong behaviourist approach, Cohen (Cohen, 1987; Cohen, 1984) indicated that:

> "*Instructional alignment* describes the extent to which stimulus conditions match among three instructional components: intended outcomes, instructional processes, and instructional assessment" (Cohen, 1987, p.16).

In both cases, even though their perspectives are quite different, learners must have a clear understanding of the educational outcomes and forms of assessment are implicitly reflecting expectations. In a very recent publication by the TLRP (teaching and learning research programme) David and colleagues depict the complex interactions of teaching and learning with ten core principles (Figure 3.1)

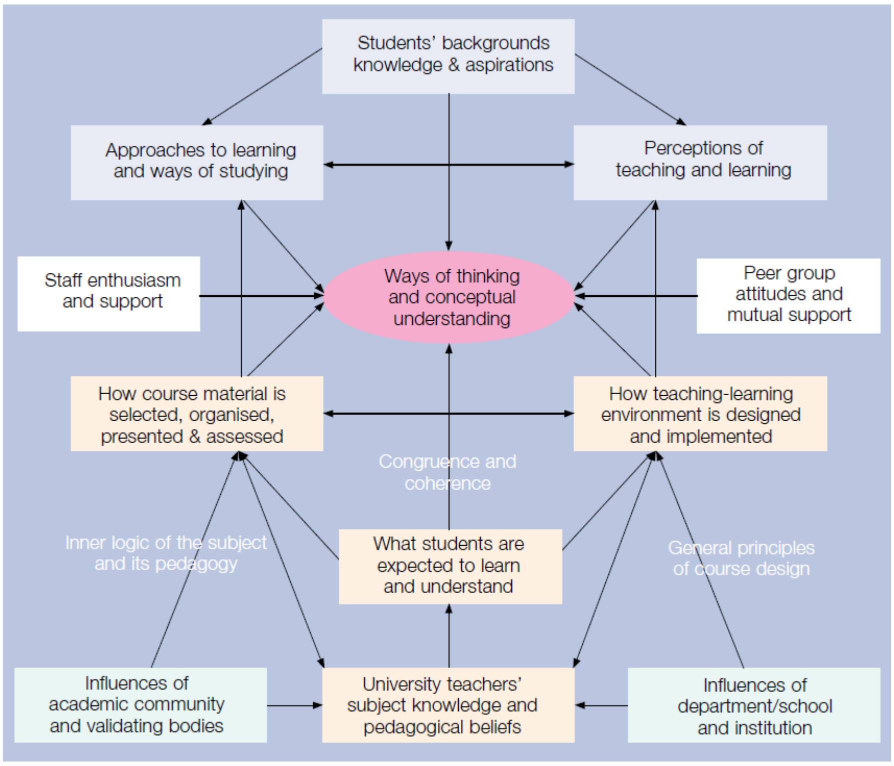


Figure 3.1. Ten evidence-informed principles to conceptualise teaching and learning in higher education (reproduced from David et al. 2009)

Ways of thinking and conceptual understanding are at the centre of this representation and assessment of the course material is only one aspect of the framework to understand teaching and learning. However, academic performance is still the *only* measure of achievement which is externally recognised (i.e. by employers) as a valid measure of one's learning (but not ability to learn!), and it is often misinterpreted as an index of ability. The statement is very much an oversimplification of the matter, but as we will see in the next section this fundamental assumption is overlooked.

3.2. Differential aspects of academic performance.

The following review of academic achievement, intelligence and personality, is far from exhaustive, but it will highlight some core aspects which seem to provide a certain overlap. We will refer in particular to a number of excellent reviews on the interrelations between these three concepts published over the last century. To make sense of the apparent inconsistencies we also use meta-analytic methods which will provide an effective overview of the research field.

Following from the warning in the last section about the fundamental difference between course grades and abilities, we can revisit the earlier definitions of learning which allow a better understanding of how the two concepts came close in the first place. According to de Raad & Schouwenburg:

> "Academic learning as a main component of education is basically a general process of information processing (Lindsay and Norman, 1972), in which perception, attention, memory, and thinking are involved. This process may be viewed as a chain of mental events leading from a stimulus of some subject matter to a response of reproduction or application in some examination behaviour. Non-cognitive personality factors may appear as moderators of the general process of learning, because they interact with-or moderate-successive stages of the information processing sequence." (Raad & Schouwenburg, 1996, p. 305).

De Raad is making a very clear distinction between cognitive and non-cognitive aspects affecting learning which Messick (1994, 1996) also refers to in the attempt to understand issues of differentiation in academic achievement.
These two aspects, however are not as clear-cut as one may think. In fact, using a very broad distinction, intelligence can be seen as an aspect of personality (i.e. Cattell, 1941) or as the cognitive component of personality (H. J. Eysenck, 1993, 1994, 1997), but in his case non-cognitive aspects are referred to as temperament) often intended as *ability*. The concept of *differentiation* was introduced in the 60's (i.e. Eysenck & White, 1964). Barton (Barton, Dielman, & Cattell, 1972) specifically indicated that:

> "On the average, such studies indicated a correlation of approximately .70 between IQ and achievement. As this figure shows that only about 50% of the variance in achievement scores can be accounted for in terms of IQ, several investigators looked elsewhere for other determining variables." (Barton et al., 1972, p. 398)

However it was only in the 90's that the empirical exploration of the relations between specific cognitive abilities and personality traits regained momentum (Ackerman & Heggestad, 1997; Austin, Deary, & Gibson, 1997; Austin, Hofer, Deary, & Eber, 2000; Furnham, Forde, & Cotter, 1998; Sternberg & Ruzgis, 1994).

Ackerman and Heggestad (1997) produced a systematic review and attempted to clarify the main results from an extensive body of literature in which results were often inconsistent and difficult to replicate because of the wide range of ability and personality measures used in the different studies.
This review was further enriched by an anthology on learning and individual differences by Ackerman (Ackerman & Rolfhus, 1999) in which he attempted to summarise a theory of

adult intellect and development according to which intelligence is differentiated into two core elements: intelligence as a process and intelligence as knowledge (the Process, Personality, Interest and intelligence as knowledge PPIK). We will review this theory toward the end of this chapter, however, in the next two sections we explore first the relations between academic success and intelligence focusing on two core ideas which attract a certain consensus: 1) there are strong relations between ability and academic success, but there is a certain amount of variance that cannot explain success alone; 2) there are consistent relations between certain aspects of personality and certain level of ability.

3.3. Intelligence and academic ability

The psychometric model applied to education, which Binet started from, must be taken into account when trying to understand the evolution of the concept of intelligence and the ways to measure it. Nowadays there seems to be a wide agreement about the existence of *g* (or general factor of intelligence, Deary, 2001; Spearman, 1904) as a result of the combination of performance measures in a number of tasks targeting different abilities. Using correlational methods, which take into account these different abilities measured by batteries of tests, allows to obtain an aggregated measure of intelligence (i.e. Kaufman, Raven, Wechsler). Cattell (Cattell, 1963; Cattell & Horn, 1978) promoted two dimensions of g: a crystallised (Gc) and a fluid (Gf) intelligence.

> "Crystallized ability loads more highly those cognitive performances in which skilled judgment habits have become crystallized (whence its name) as the result of earlier learning application of some prior, more fundamental general ability to these fields. Thurstone's Verbal and Numerical primaries, or achievement in geography or history, would be examples of such products". (Cattell & Horn 1963, p.3)

In this sense crystallised intelligence is the "accumulated knowledge of an individual." (Cattell 1978, p. 140).

> "Fluid general ability, on the other hand, shows more in tests requiring adaptation to new situations, where crystallized skills are of no particular advantage." (Cattell & Horn 1963, p.3).

Cattell also stressed that:

> "The principal distinction between Gf and Gc in the general theory, however, does not pertain to the type of tasks involved. Instead, it pertains to the kind of development that leads to the separation of two structures." (Cattell, 1979, p. 140)

Integrative models were also proposed such Vernon (1961) and Carroll (Carroll, 1993b, 1997) hierarchical models. The latter, in particular, is acknowledged as the most credible based on a review of over 400 research papers which is leading to a ‘three stratum model’ of human cognitive ability in which abilities are clustered and represented in a hierarchical system. Other theories criticise the “g-ocentric” model of intelligence and *cognitive* abilities: it is the case of Sternberg’s triarchic model (Sternberg, 1985) and Gardner’s multiple intelligences model (Gardner, 1993b).

Despite the different views, intelligence is drastically determined by its *instrumentalism* (Flynn, 2007); in Jensen’s words “intelligence, by definition, is what intelligence tests measure” (Jensen, 1972, p. 76) and this provides scope for criticism not only for the theoretical frameworks, but also undermines the credibility of psychology as a science as the measurement tools don’t seem to have the same level of precision of measures in physical sciences (Bond & Fox, 2007). Independent of the actual model or instrument used, quantifying abilities using measures of intelligence has been very appealing and it is difficult to extricate the close tie between intelligence and education.

As early as the 1940s, Harris, reviewing the literature available at the time, stressed that *intelligence* was the most essential determinants for academic success (Harris, 1931, 1940). In his second review he claimed that

> “High school grades combined with intelligence test score uniformly show higher correlations with grades than does either component separately (Finch & Nemzeck 1934, Johnston & Williamson 1934, Jones & Laslett 1935, Reitz 1935) ” (Harris 1940, p.127).

He concluded with the following statement:

> (...) the essential factors in student achievement are, in order of importance: (1) Ability (or intelligence, or scholastic aptitude, etc.). (2) Effort (or drive, or degree of motivation, etc.) (3) Circumstances (personal, social, economic, academic etc.)” (Harris 1940, p.151).

These beliefs have remained the cornerstones of folk psychology, but most worryingly, have been an implicit factor in shaping educational systems and job selection.

Linda Gottfredson is one of those who contest this strongly and claims that the

> “Predictive Validity of *g* is ubiquitous. The key observation here is that personnel psychologists no longer dispute the conclusion that g helps to predict performance in most if not all jobs (Hartigan & Wigdor, 1989). Rather, their disputes concern how large the predictive validities are, often in the context of deciding the appropriate composition of a personnel selection battery.” (Gottfredson, 1997, p. 81).

Gottfredson establish very specific and clear relations between ability and intelligence and the large samples summarised in table 3.1 with their relations to ability seems to make a compelling case.

At the other end of the spectrum, Sternberg is possibly the most active and vocal about the societal bias and educational weaknesses which is perpetuated by the traditional view of *g* and the test of individual abilities based on analytical skills. He believes that the ubiquitous relationship between *g* and measures of academic achievement (as in the wide review by Harris above) is partially attributable to the effects of schooling on analytical achievements and the relative neglect of practical and creative intellectual achievements.

NOTES:

a. Meta-analysis of 515 validation studies conducted by the D.S. Employment Service (Hunter, 1983; Hunter & Hunter, 1984),425 of job performance (32,124 workers) and 90 of training success (6496 workers).
b. E.g., retail food manager, fish and game warden, biologist, city circulation manager. DOT "data" equals 0 or I.
c. E.g., automotive mechanic, radiologic technician, automotive parts counterman, high school teacher. DOT "data" equals 2-4.
d. E.g., assembler, insulating machine operator, forklift truck operator. DOT "data" equals 5 or 6.
e. E.g., machinist, cabinetmaker, metal fabricator. DOT "things" equals O.
f. E.g., shrimp picker, corn-husking machine operator, cannery worker, spot welder. DOT "things" equals 6.

	Performance	Training
General job families		
High complexity[b]	.58	.50
Medium complexity[c]	.51	.57
Low complexity[d]	.40	.54
Industrial families		
Precision set-up[e]	.56	.65
Feeding/offbearing[f]	.23	—

Table 3.1. Predictive Validity of intelligence in jobs of different complexity. Reproduced from Gottfredson 1997)

This is demonstrated in cross-cultural studies in which he was able to show that the importance attributed to different abilities are partially exacerbated by cultural and societal pressures and, more importantly, that success in the academic sphere often does not equate to success in job performance (Sternberg, 2005; Sternberg, 1999). In his triarchic model of intelligence, Sternberg maintains that abilities are separate and partially independent. Specifically, he argued that intelligence could be understood in terms of a set of elementary information-processing components that accounted for individual differences (Sternberg, 1981) and in later works he expanded this analytical model of intelligence to include a creative and practical aspect of intelligence (Sternberg, 1996; Sternberg & Wagner, 1993). In 2005 he renamed the triarchic model the “Theory of successful intelligence”. After reviewing a large number of studies he provided fairly convincing evidence that

> “Some people are intelligent and creative, but foolish. That is they, are smart, but not wise.” (Sternberg 2005, p. 199).

In his lucid review he came to the following conclusion which also acknowledged his critics:

> “Some Psychologists will believe that the theory of successful intelligence departs too much from the conventional theory of General intelligence (Spearman 1904): Some disagree with parts of the theory (e.g. Brody, 2003a, 2003b) and some disagree with the whole thing, vehemently (Gottfredson 2003a, 2003b) Others believe the theory does not depart from conventional theory enough (Gardner 1983)”.

The most recent attempt to provide a unitary conceptual framework is the one by Flynn (2007) who asserted that researchers investigating intelligence *must* take an approach which considers the brain organization, individual differences and social trends as having an equal role into a coherent integral explanation of intelligence. His position is quite interesting, because he acknowledges these aspects not as new in research and does not attempt to propose yet another theory, but he is acutely aware that ignoring one of these aspects will not lead to a complete theory of intelligence. In his definition (Flynn 2007, pp 52-53) he explicitly refers to six aspects characterising a “pre-theory” of intelligence: mental acuity, habits of mind, attitudes, knowledge and information, speed of information processing and memory. He doesn’t exclude the likelihood that other factors can affect problem solving abilities and acknowledges Sternberg and Goleman (1995, who mentions emotional intelligence), however he is also reluctant to try to include everything.

No matter which view of intelligence is taken, or which interpretation is considered valid, Sternberg is right on one thing:

> “No doubt, there will be those who wish to preserve this [he refers to the theory of general intelligence] and related older theories, and those who will continue to do research that replicated hundreds and thousands of time that so-called general intelligence does indeed matter for success in many aspects of life. I agree. At the same time, I suspect it is not sufficient (…)”. (Sternberg 2005, p200).

Even though Gottfredson claimed that g is the best predictor for job performance and provided evidence for a clear causal relation between high level of g and proficiency in complex tasks (1997), the methodological weakness of using a measure of intelligence to provide an in-depth profile of students is self-evident: if we make the core assumption that intelligence can change over time and education can be used to improve abilities, when students reach university they are already *shaped* by the educational system to perform well on certain tasks having spent years in perfecting the necessary strategies and refining their abilities to achieve good grades and pass exams. According to Sternberg, at this stage the creative and practical abilities have been filtered out to focus on the analytical abilities which the education system is promoting. Those who do not, or cannot excel in school are not going to university at all or don't have an option to be accepted into *good*[5] universities: this is particularly the case in the British system in which the grades are largely determining entry. (i.e. the advertised minimum grade requirements effectively prune out the ambitions of prospective students even before they apply for a specific university and/or course). For these reasons we rule out the consideration of measures of intelligence as there would be little value or originality in finding that academic performance is predicted by intelligence scores *per se*.

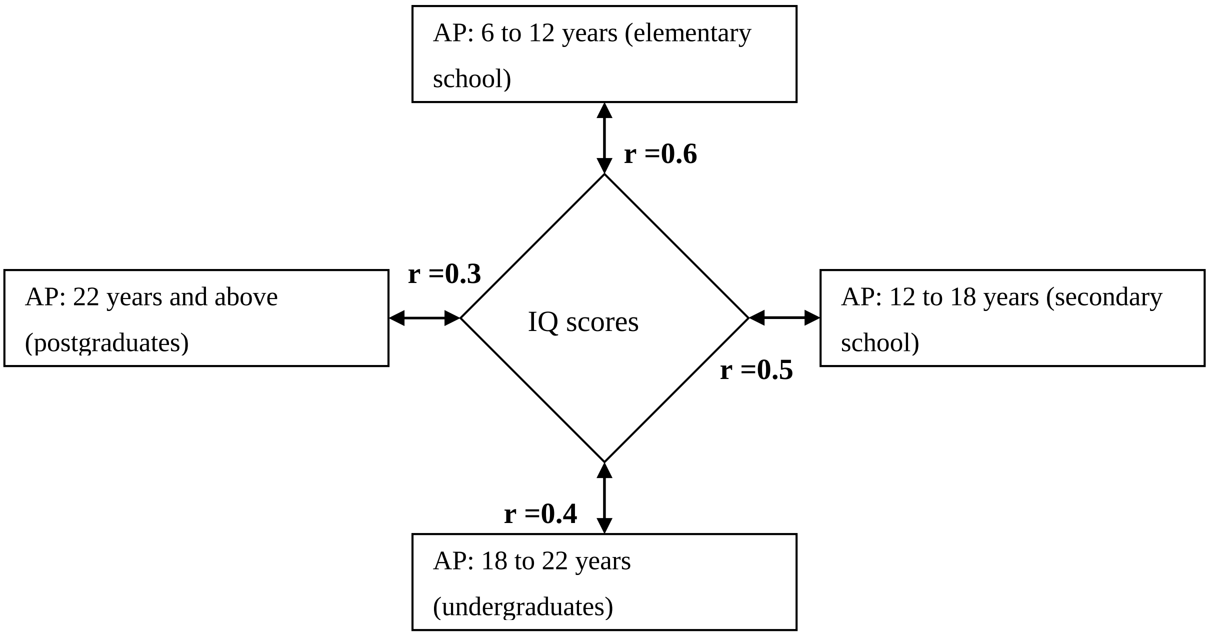


Figure 3.2. Adapted from Chamorro-Premunic & Furnham 2005.

Correlation between intelligence and Academic performance (AP). Note that the correlations are approximate based on Ackerman (1994), Boekaerts (1995), Hunter (1986) and Jensen (1980)

---

[5] Please take this with a warning as I'm not providing any indication of how *goodness* is measured, but I refer generically to common beliefs based on Universities league tables published by the Times and Guardian in 2007-08. These are different from the league tables produced by the RAE and from the National Student Survey, all of which provide very different views on each institution.

A different approach is to consider the relations between IQ and education level to assess the predicting power of IQ scores: Chamorro-Premuzic and Furnham (Chamorro-Premuzic & Furnham, 2005) summarised the correlations from studies over 20 years. (Figure 3.2)

The decrease in the strength of the correlation between IQ and AP, is a clear indicator that the predicting power of intelligence decreases with higher levels of education and that there are other factors affecting academic performance which become more relevant at higher levels of instruction.

There are two important aspects which arise from this discussion: the relation of intelligence with personality aspects, which will be addressed shortly after examining personality and academic achievement, and the idea that having or obtaining a measure of one's intelligence of ability, doesn't necessarily mean that this matches the individual self-belief of his/her intelligence or ability.

In fact, there is recent literature which stresses the importance of the mediating value of motivational factors (affected by self-concepts or perceived ability) between abilities and task completion (Chen, Gully, Whiteman, & Kilcullen, 2000; Dweck, 2005). According to these views, self-concepts or self-beliefs of intelligence strongly affect performance (possibly relatively to one's best abilities): in other words, if people believe not to be smart and hold a concept of intelligence which is given, rather than something they can work on and improve, they will tend to self-handicap themselves, avoiding challenging situations and fulfil their self-fulfilling prophecies of not being able to achieve much. (Dweck, 2005; Robins & Pals, 2002)

3.3.1. Personality and attainment

Theories and perspectives on personality evolved a great deal over the past century and a five-factors *model* of personality is now widely favoured over previous models (Costa & McRae, 2001; Goldberg, 1999). According to the five-factors model it is possible to identify five major bipolar factors: Openness to experience (or intellect), Extraversion, Agreeableness, Conscientiousness and Neuroticism (or emotional stability). As for the theories of intelligence, there are some who are quite critical about the bases upon which the model was built on. At the core of the discussion is the number of dimensions which can be considered sufficient to explain personality: even though there are overlapping concepts, using a biological/neurological explanation in contrast to a linguistic reduction of attributes, provided different solutions (i.e. Block, 1995; H. J. Eysenck, 1992; Zuckerman, Kuhlman, Joireman, Teta, & Kraft, 1993). Certainly, not everyone is happy to accept the five-factors unconditionally, but for different reasons (i.e. Block, Eysenck, Cattell).

In this brief review, we will focus mainly on the five-factors model as a scaffold to draw the relations between personality and academic achievement, but we will also relate the 5-factors to the corresponding constructs used in other research which focused on the 3-factors model (Eysenck, Eysenck, & Barrett, 1985) and the 16-dimensions in Cattel's inventory (Cattell & Bernard, 1946) which have been used widely in the existing literature.

The links between personality and performance or learning have a long history: in psychology Pavlov in his first experiments, already acknowledged *temperament* as a contributing factor in learning. Disposition and motivation have also been mentioned in relation to performance (Alexander, 1935; Webb, 1915; Wechsler, 1943). However it is only with the systematic study of personality that the relations between learning and academic performance were considered. Some authors were more concerned about the reciprocal effects of personality and achievement and considered personality a complicating factor. Entwistle, for example, suggested that

> "it is dangerous to assume wide generality in statements about the relationship between personality and academic attainment. Age, ability, sex, geographic area, classroom organization, class size, teaching methods and teachers' personality may all affect these relationships to some extent." (N. J. Entwistle, 1972)

Some even went as far as asserting that personality is not significantly related to academic achievement and concluded that the lack of consistent findings makes the value of personality of no real significance in educational settings (Allik & Realo, 1997; Dollinger & Orf, 1991; Green, Peters, & Webster, 2009; Mehta & Kumar, 1985; Rothstein, Paunonen, Rush, & King, 1994)

Nevertheless, methods to study individual differences and the amount of research invested clearly point in the opposite direction. In the last 2 years, the topic regained momentum and we could identify three meta-analyses (Noftle & Robins, 2007; O'Connor & Paunonen, 2007; Poropat, 2009) which we will try to condense at the end of this section. Earlier, in a detailed review of the literature on personality and academic achievement, De Raad & Schouwenburg (1996) pointed out that *types* of learners can be formed simply based on the characteristics of groups:

> “Analogously, the field of learning and education has produced a variety of promising and problematic learners: underachievers (Mandel and Marcus, 1988), procrastinators (Schouwenburg & Lay, 1995), self-actualizers, etc. Sometimes the types are constructed from an interpretation of characteristics of learners with problems, sometimes by taking the extremes of dimensions and/or theoretical ideals.“ (De Raad & Schouwenburg 1996, p 318).

Although in some respects the predicting power of personality traits for academic performance varies, the empirical evidence concerning the role each of these traits play in determining academic success is mixed and far from clear (Chamorro-Premuzic & Furhnam, 2008; Diseth, 2003a; Stevens, 2001). One of the core reasons for the variation seems to be age-related: for example, there are major differences in the way the trait extraversion affects how primary school children and university students manifest it in actual behaviours.

> “Many apparent contradictions in the literature are resolved by the recognition that stable extraverts tend to be successful in primary schools while introverts, and possibly even neurotic introverts, predominate among outstanding students. [in higher education]” (Entwistle, 1972, p147)

In support of the view that personality does have a valuable impact for educational research, it is possible to specify that some personality measures on their own are powerful enough to explain a small to moderate percentage of the variance in academic performance and we will review some of the studies highlighting the general consensus even if it is difficult to highlight specific patterns or combinations (Busato, Prins, Elshout, & Hamaker, 1999, 2000; Chamorro-Premuzic & Furnham, 2003; Diseth, 2003b; Laidra, Pullmann, & Allik, 2007).

To explain the value of the relations between personality and achievement we use Matthews’ cognitive-adaptive framework for personality (G. Matthews, 2000; Gerald Matthews et al., 2002). Traditionally, the behavioural expression of traits can be explained using one of two broad approaches: a psychobiological approach is taken when traits or dispositions are believed to cause a large portion of variance in individual differences (i.e.(H. J. Eysenck, 1967; Zuckerman et al., 1993) for a review). Alternatively approaches based on information processing theory are looking at differences in attentional demands of different tasks or individual limitations of cognitive systems (i.e. memory capacity, dual task performance or multitasking, or even reaction times; (Der & Deary, 2003; Gathercole & Baddeley, 1993; Logie, Cocchini, Delia Sala, & Baddeley, 2004). Deary and colleagues in particular, (Deary 2001, Der and Deary 2003) exemplified the complex non-linear relations between reaction times (RTs) and mental ability and called for some caution in the interpretation of the correlations reported in the literature.

Even though the psychobiological approach explains some of the physiological correlates of behaviours (i.e. related to arousal states), a crucial shortcoming is the inability of such explanations to make sense of complex behaviour. On the other hand, information processing theories allow to clearly identify "cognitive patterning" (Hockey, 1984; Matthews, 1992), or explain how a set of processing characteristics is associated with high level of specific traits. An example frequently used is that extraverts have good short-term memory but poor attentional span and vigilance whilst introverts are more successful in performing with highly demanding and prolonged tasks (Matthews 1992).

Both perspectives, however fail to keep into account the contextual aspects of behavioural expression and both Furnham (1992) and Matthews (Matthews & Wells, 1999) pointed out that people's lives differ in content, that is their experiences are different based on the environment in which they develop. For example Snyder (Snyder et al., 1991) suggested a social-cognitive perspective of extraversion: greater practice of social interaction and social regulation might be a stronger determinant to refine the level of extraversion of an individual. Austin (Austin et al., 2000), who looked into emotional intelligence, found that people with high levels of emotional intelligence were also more manipulative and termed this phenomenon the 'dark side of EI'. Matthews (1999) attempted to integrate these different perspectives and suggested that a cognitive-adaptive framework is more suited to explain behavioural manifestation. According to this theory, "Skills are learned processing routines for accomplishing specific work or social tasks that are 'tuned' to the environments associated with the tasks." (Matthews 1999, p254) The model identifies both a feed-back and feed-forward processes between cognition and personality which are mediated by self-beliefs, self-regulative processing and produced emotional reactions characterising the adaptation to the environment.
To understand the relations between traits and personality, we will examine some of the evidence present in the literature for each of the broad five traits in turn.

Neuroticism
One of the most studied personality traits in relation with academic performance is neuroticism. The key reason seems to be that both intelligence tests and academic performance are measured through maximal performance tests. As one would expect stressful situations create a certain degree of worry and anxiety which have evident psychophysiological effects. Neurotic people, who are already predisposed to react emotionally to situations, could be highly disadvantaged. Ackerman & Heggestad (Ackerman

& Heggestad, 1997) meta-analysis showed a significant negative relationship between stress reaction and knowledge achievement. Even though neuroticism and academic performance have been negatively associated with reported correlation coefficients ranging around .2 to .3 (Ackerman & Woltz, 1994; Hembree, 1988; Seipp, 1991), others found a positive relation (De Barbenza & Montoya, 1974; De Raad, 1998) or no significant relations (Busato et al., 2000; Furnham & Mitchell, 1991; Halamandaris & Power, 1999). Hence, it is natural to infer that any such association is more complex than it might initially seem. Chamorro-Premunic & Furnham, taking into account research on arousal and stress (Eysenck, Derakshan, Santos & Calvo, 2007; Lazarus, DeLongis, Folkman, & Gruen, 1985) and the fact that most of the existing literature is considering university students, suggested that actual and perceived competence interact to create a "neurotic feedback" which might lead to lower academic performance. An individual high in neuroticism is more likely to be affected by stressful situations: this enhances the effects on perceived confidence which is creating more instability and fuels further neurotic behaviours, leading to higher stress and lower performance. McKenzie (1989, 2000), also acknowledged the negative correlation between the trait and academic performance, but he also pointed out that N correlates positively with success at a the degree level for students who scores highly in Cattell's superego strength, which he termed the 'Furneaux effect'. In a similar way, but with a much bigger sample of non-students, Austin et al. (1999, 2001) examined more carefully the differential relations between levels of intelligence and personality traits and showed that the relations are indeed more complex (non-linear) in people with high abilities, displaying a greater differentiation at trait level, which we will consider in the next section.

Matthews (1999) associated the lack of emotional stability to anxiety and, using the cognitive-adaptive framework explained above, differentiated the layers of interaction between the two by including in his model the effects of the environment (perceived threat), motivation, self-knowledge and coping as all affecting basic processing and appraisal: all these elements have an essentially adaptive function of self-preservation.

Conscientiousness

Conscientiousness, defined by organization, persistence and motivation in goal-directed behaviours, is the trait which one would intuitively expect to closely affect academic performance; this finding is consistently supported in the literature with a positive association with academic performance. Many studies have replicated this relationship in school and at undergraduate or postgraduate level (Busato et al., 2000; Chamorro-Premuzic & Furhnam, 2003; Lounsbury, Sundstrom, Loveland, & Gibson, 2003; Nguyen, Allen, & Fraccastoro, 2005), however Farsides and Woodfield (2003) found that in a sample of 432 undergraduate students at the university of Sussex, conscientiousness was significantly and positively related to attendance but not the final grades. Diseth (2003) with a sample of 315 Norwegian students did not find a significant correlation, and, with smaller samples, Duff (Duff, Boyle, Dunleavy & Ferguson, 2004)with 146 undergraduate students in a Scottish University, and Burton & Nelson (Burton & Nelson, 2006) with 119 students at the University of Queensland, confirmed a lack of correlation with academic achievement. This evidence is not surprising as one would expect that the motivational drive of conscientious students leads to higher levels of focus on task and the implementation of better strategies for achieving in the educational context and workplace.

Interestingly, however, research conducted by Dweck and colleagues (Dweck 2005, Robins and Pals 2003, (Blackwell, Trzesniewski, & Dweck, 2007) indicated that depending on their idea of intelligence, students with a 'fixed' concept of intelligence (i.e. those who believe that their abilities are innate and cannot be modified), have the tendency of attributing failure to their inability. This puts them in a very difficult situation in which self-confidence is drastically affected by their perceived inadequacy rather than by other external factors such as effort or objective difficulty. In their analysis Robins & Pals (2003) clearly identify the strength of the relation between implicit self-theories, goal orientation (either performance or learning goals) and the effects that these have on self-esteem which is associated with either a 'helpless' or 'mastery response' behavioural manifestation.

Openness

There are logical reasons why openness should be able to predict academic performance: students who are more open to new learning, discovery and exploration should be more receptive and more effective in engaging with higher education. Openness is also associated with creativity, curiosity and flexible thinking and should facilitate students in dealing with novel material and situations. In the literature, a positive correlation between openness to experience and academic achievements have been found in a number of studies (Blickle, 1996; Cacioppo, Petty, Feinstein & Jarvis, 1996; De Raad & Schouwenburg, 1998; Goff & Ackerman, 1992; Lounsbury et al., 2003; Paunonen & Ashton, 2001; Rindermann & Neubauer, 2001; Wolfe & Johnson, 1995). However there are also studies in which this correlation was not supported (Busato et al. 2001, Bauer & Liang, 2003, Chamorro-Premuniz & Furnham 2003, Farside & Woodfield 2003,). In some cases (i.e. Fruyt & Mervielde, 1996 and Diseth 2003) different samples produced different correlations coefficients with Philosophy students showing higher or significant correlation not matched in other students.

Blickle (1996) when trying to identify a suitable explanatory model, justified the lack of direct correlations between O and C with academic performance because of the mediating effect of the cognitive dimensions of learning measured using the LIST inventory (Wild & Schiefele, 1994). Diseth (2003), which we will consider in more details in the next chapter, proposed a model in which learning styles are mediating the effects of personality traits (N, O and C). Interestingly, in the latest version of the path model proposed by Chamorro-Premuzic & Furnham (Chamorro-Premuzic & Furhnam, 2008) and which also account for intelligence, a deep approach to learning is a mediator with the same strength of openness between IQ and grades, but not between traits and exams. (figure 3.3)

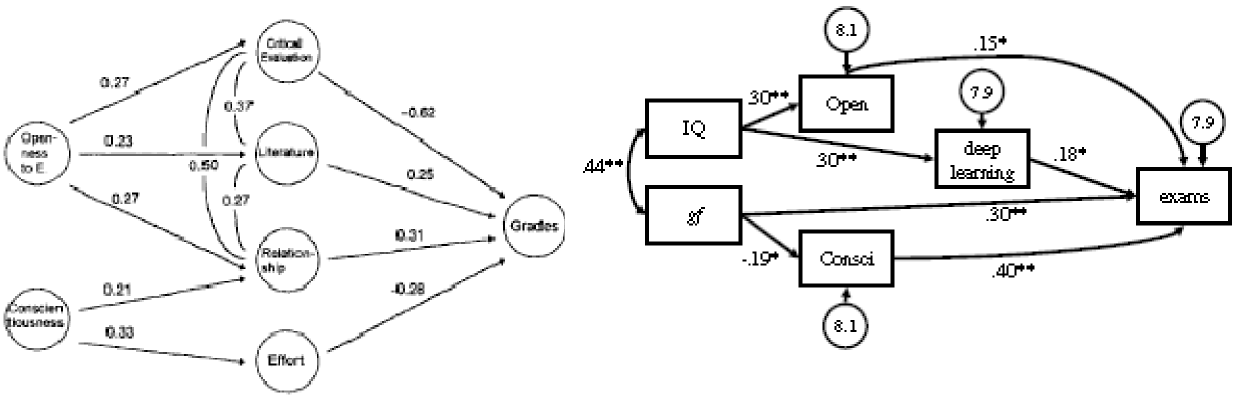


Figure 3.3. Path analysis of the relationships between grades and traits from Blickle and Chamorro-Premuzic.

Blickle (1996), left: note that GPA in the German system is inverted. i.e. 1=best grade, explaining negative correlations) and the same traits adding IQ (Chamorro-Premuzic & Furnham 2008).

Extraversion

Early research investigating extraversion has shown that introverts usually perform better at school, but this was not replicated in later studies, with a clear distinction between children and adolescents and younger adults. (Entwistle 1972) As observed earlier, Matthews explored the relations between personality and performance (not only academic). In particular, Matthews and colleagues were able to support the hypothesis that extraverted prefer environments which tend to overload attention and in which rapid responses and social interactions are required. In their research they also demonstrated that not only extraverted prefer these features in the environments they work in, but they are also performing better than introvert. Using cognitive patterning, Matthews identified that extraverted are normally superior in divided attention, resistance to distraction, retrieval from memory and short-term memory, however they were performing worse in vigilance, reflective problem solving and long-term memory. Other factors to be considered are the nature and difficulty of the task: in fact, Eysenck has shown that in experimental conditions, extraverts performed faster than introverts when the task was short and intrinsically interesting.

The existence and strength of the relation between E and academic achievement is very inconsistent. One of the possible reasons could be attributed to the fact that some learning activities which focus on individual abilities (i.e. passing a test) are posing very different demands compared with group projects just to give an example. In recent literature a small negative correlation with grades is reported in Bauer & Liang (2003, r=-.112, N=165), with the first examination at university in Busato et al. (2001, r=-.13, N=409), with projects (r=.14, N=338), but not with exams in Dollinger (Dollinger, Matyja, & Huber, 2008) and between E and work drive in Lounsbury et al. (2003, r=0.28, n=275).

From studies of activity and vocational interests Ackeman & Heggestad (1997) highlighted the fact that cognitive patterns do transfer in day-to-day behaviours and occupational preference and showed that extraverted people tend to prefer high-pressure jobs in which they can exercise their social skills and provide rapid action: the interaction with others, in combination with higher self-confidence and efficacy is further fuelled by the positive feedback from the relations and job satisfaction. These effects are also accounted for in Matthews' cognitive-adaptive framework (1999).

Agreeableness

Agreeableness is the trait which is associated with kindness, warmth and willingness to help, but it should be noted that Eysenck (1991) asserted that openness and agreeableness are only facets of psychoticism. Other facets of agreeableness are trust, altruism, compliance and modesty. De Raad & Shouwenburg (1996) observed that there might be a weak correlation between AP and A, but it is no surprise that very few of the papers cited thus far are reporting significant correlations with academic achievement. In particular, Farside & Woodfield (2003) reported a small but significant correlation between A and the final grade (r=.14, N=432) in their sample of British students. Chamorro-Premuniz & Furnham (2003) also reported a relation between A and the first year exam grades (r=.34, N=145) also on a sample of British students. Finally Dollinger et al. (2008) reported a small correlation between A and grade awarded for projects (r=.16, N=338) in a third year US sample.

3.4. A meta-analytic summary of the relations between AP and personality

So far we have been able to identify a number of relations between personality traits and academic performance, however findings are often inconsistent and it is possible that even when reasonable sample sizes are used the relations are actually confounded by non-measured mediators as it was the case for O and C in the path models proposed by Blickle (2003), Diseth (2003) and Chamorro-Premuzic (2008).

In the attempt to make the observation about these relations more systematic, and with the aid of two recent reviews (O'Connor & Paunonen 2007 and Noftle & Robbins 2007) we adopted a meta-analytic approach.

Coincidentally, a similar study by Poropat (2009) reached the publication stage allowing for a direct comparisons of the findings. The tables 3.2-3.4 in the next pages are useful to provide a full picture of the correlations reported with academic performance.
In this analysis we collated 48 studies (up to 2008 and specifically in tertiary education, excluding dissertations and theses) in which it was reported a correlation between a personality inventory and a measure of academic performance.
As not all the personality inventories could be reduced to the five factors model, in some instances there are empty values.

The crucial difference between Poropat's analysis and the one conducted in this thesis is the fact that AP was not intended solely as final grades, but we were particularly interest in identifying relations other indicators such as prior performance (table 3.5) various types of assessment (3.3) and other interesting proxy measures such as attendance and participation (table 3.4). For his reason we did not attempt to provide group summaries in the analysis.

Methodologically, some of the studies were just too different from each other to allow for a sensible integration, therefore we only relied on the transformation to Z scale values to allow for a direct comparison between the correlations. An exception was done for the final grades

and GPA values (table 3.2 and 3.3) in which we found it useful to compute aggregated correlation values reported at the bottom of each table. As there was not attempt to make methodologies or samples homogeneous in different studies a random effect model computation was used.
In the tables we reported the correlation values found in the original papers, Fisher's z scales and the relative adjusted p values.

Poropat (2009) reported 'generally modest scale-corrected correlations' (Emotional stability .06, Extraversion -.01, Openness .15, Agreeableness .01 and Conscientiousness -.03). As evident from the tables below, the figures are quite similar, although both C and O are smaller)

Even if this meta-analysis is quite interesting in its own right, for the purpose of this thesis it is not as important to determine the strength of the relations (which is quite small). However, it is fundamental to point out that such relations do exists and do have an impact on academic performance, and potentially job performance at later stages.

Personality traits account only for a very small variance in human performance, and it seems evident from the review carried out that the direct relations between traits and achievement could be mediated by other factors not considered in all studies (examples are Blickle, Diseth and Chamorro as well as Noftle & Robbins mentioned above). We will return on the issue of unaccounted variables in the final section of this chapter, but first it is necessary to review the models which consider the relations between intelligence, personality and performance.

GPA				Correlations reported						N		E		O		A		C	
Study authors	measure	criterion	Sample	N	E	O	A	C		Fisher Z	p value	Fisher Z	p value	Fisher Z	p value	Fisher Z	p value	Fisher Z	p value
Dollinger and Orf (1991)	NEO-PI	Course grade	118	-.01	.11	.2	.05	**.25**		-0.01	0.91	0.11	0.24	0.20	0.03	0.05	0.59	0.25	0.01
Chamorro-Premuzic & Furnham (2003) EJP	NEO-PI-R	Course grade	247	**-.16**	-.11	.02	.07	**0.36**		-0.16	0.01	-0.11	0.08	0.02	0.75	0.07	0.27	0.36	0.00
Lounsbury et al., 2003	PSI	Course grade	175	-0.11	.01	**.16**	-0.01	**0.18**		-0.11	0.15	0.01	0.90	0.16	0.03	-0.01	0.90	0.18	0.02
Paunonen and Ashton (2001a)	PRF	Course grade	717			-0.04		**0.21**		0.00				-0.04	0.28			0.21	0.00
Hair and Hampson (2006)	BFI	Course grade (exam avg)	236	.11	.15	0	.01	.18		0.11	0.09	0.15	0.02	0.00	1.00	0.01	0.88	0.18	0.01
Chamorro-Premuzic and Furnham (2003b) JofRP, Sample1	NEO-FFI	Course grade (exam tot)	70	**-0.35**	0.07	0.00	0.22	**0.39**		-0.37	0.00	0.07	0.57	0.00	1.00	0.22	0.07	0.39	0.00
Chamorro-Premuzic and Furnham (2003b) JofRP, Sample2	EPQ-R	Course grade (exam tot)	75	**-0.37**	0.13	**-0.29**				-0.39	0.00	0.06	0.62	-0.29	0.01				
Furnham et al. (2003)	NEO-PI-R	Course grade (Exam)	93	.14	**-0.29**	-0.16	.06	**0.40**		0.14	0.18	-0.20	0.10	-0.16	0.19	0.06	0.62	0.40	0.00
Furnham et al. (2003)	NEO-PI-R	Course grade (Exam_Y2)	93	.08	**.22**	.09	.01	**.34**		0.08	0.45	0.13	0.27	0.09	0.44	0.01	0.93	0.34	0.00
Chamorro-Premuzic & Furnham (2008)	NEO-PI-R	Course grade (Exams)	158	-.05	.16	**.21**	.02	**.37**		-0.05	0.53	-0.29	0.00	0.21	0.04	0.02	0.85	0.37	0.00
Diseth (2003)	NEO-PI-R	Course grade (Exam-S1)	151	-0.03	-0.10	.03	.12	0.06		-0.03	0.72	0.22	0.03	0.03	0.78	0.12	0.25	0.06	0.57
Diseth (2003)	NEO-PI-R	Course grade (Exam-S2)	164	.2	-.07	**.22**	**-.21**	-.10		0.20	0.01	0.16	0.04	0.22	0.01	-0.21	0.01	-0.10	0.21
Chamorro-Premuzic and Furnham (2003b) JofRP, Sample1	NEO-FFI	Course grade Y2	70	**-0.31**	0.06	0.06	0.06	**0.34**		-0.32	0.01	-0.10	0.22	0.06	0.46	0.06	0.46	0.34	0.00
Chamorro-Premuzic and Furnham (2003b) JofRP, Sample1	NEO-FFI	Course grade Y3	70	**-0.32**	-0.20	0.15	0.03	**0.34**		-0.33	0.01	-0.07	0.37	0.15	0.06	0.03	0.70	0.34	0.00
		tot N	**2437**						**Random model effects**	**-0.08**	**-1.59**	**0.01**	**0.87**	**0.05**	**0.12**	**0.02**	**0.39**	**0.25**	**0.00**
Lounsbury et al. (2003)	PSI	Course grade	175	0	0	+	0	+	course grade										
Lounsbury et al. (2005)a	APSI	Course grade	434	- -	+	+	+	+	course grade										
Ridgell & Lounsbury (2004)	PSI	Course grade	140	-	0	0	0	0	course grade										
Wolfe & Johnson (1995)	*BFI*	*GPA*	*201*	*0*	*0*	*0*	*0*	+++	GPA										
Langford (2003)	*BFM*	*GPA*	*203*	*0*	*0*	*0*	*0*	+++	GPA										

Table 3.2. Meta-analysis of the relations between personality traits and academic success measured using GPA, course grades and exams.

GPA				Correlations reported						N		E		O		A		C	
Study authors	**measure**	**criterion**	**Sample**	**N**	**E**	**O**	**A**	**C**		**Fisher Z**	**p value**	**Fisher Z**	**p value**	**Fisher Z**	**p value**	**Fisher Z**	**p value**	**Fisher Z**	**p value**
Barchard (2003)	IPIP, NEO-PI	GPA	150	-0.13	.050	.171	**.142**	.327	GPA	-0.13	0.11	**0.05**	**0.54**	0.17	0.04	0.14	0.08	0.33	0.00
Conard (2006)	NEO-FFI	GPA	289	-.06	.0	-.02	.11	**.35**	GPA	-0.06	0.31	0.00	1.00	-0.02	0.74	0.11	0.06	0.35	0.00
Dollinger, Matyja and Huber (2008)	BFI	GPA	338	0.01	.02	.03	.02	.26	GPA	0.01	0.85	0.02	0.71	0.03	0.58	0.02	0.71	0.26	0.00
Duff & al (2004)	16PFI	GPA	146	-0.14	0.06	0.07	0.12	**0.21**	GPA	-0.14	0.09	0.06	0.47	0.07	0.40	0.12	0.15	0.21	0.01
Oswald & al (2004)	IPIP BFM	GPA	644	0.7	-.03	.03	.10	0.21	GPA	0.70	0.00	-0.03	0.45	0.03	0.45	0.10	0.01	0.21	0.00
Bauer and Liang (2003)	NEO-FFI	GPA	265	0.00	-.175	-.016	.059	.216	GPA	0.00	1.00	-0.18	0.00	-0.02	0.80	0.06	0.34	0.22	0.00
Farsides and Woodfield (2003)	NEO-FFI	GPA	432	0.03	0.00	**0.26**	**0.14**	0.09	GPA	0.03	0.53	0.00	1.00	0.26	0.00	0.14	0.00	0.09	0.06
Goff and Ackerman (1992)	NEO-PI	GPA	147	-0.09	**-0.17**	0.00	0.03	**0.17**	GPA	-0.09	0.39	0.04	0.71	0.00	1.00	0.03	0.78	0.17	0.11
Oswald et al. (2004)	IPIP/BFM	GPA	636	0.00	0.00	0.00	0.00	0.00	GPA	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00
Paunonen (1998) (Study 1)	NEO-FFI	GPA	96	0.18	-0.15	.08	**-0.24**	0.06	GPA	0.18	0.00	0.04	0.42	0.08	0.11	-0.24	0.00	0.06	0.23
Paunonen (1998) (Study 2)	NEO-FFI	GPA	92	0.03	-0.02	.19	.03	0.20	GPA	0.03	0.55	0.04	0.42	0.19	0.00			0.20	0.00
Ridgell & Lounsbury (2004)	PSI	GPA	140	0.00	0.00	0.00	0.00	0.00	GPA	0.00	1.00	0.02	0.59	0.00	1.00	0.03	0.55	0.00	1.00
Busato & al 2000	5PFT	GPA (AcadY2)	409	0.00	-0.05	0.02	0.06	**0.21**	GPA	0.00	1.00	0.02	0.83	0.02	0.83	0.00	1.00	0.21	0.02
Busato & al 2000	5PFT	GPA (AcadY3)	409	0.03	-0.07	0.01	0.02	**0.18**	GPA	0.03	0.53	-0.17	0.00	0.01	0.84	0.06	0.51	0.18	0.00
De Fruyt and Mervielde (1996)	NEO-PI-R	GPA (Final grades)	741	-0.09	0.02	-0.09	0.05	**0.28**	GPA	-0.09	0.06	0.00	1.00	-0.09	0.06	0.02	0.68	0.28	0.00
Phillips et al. (2003)	NEO-FFI	GPA (Final Results)	125	0.04	-0.04	**.19**		**0.26**	GPA	0.04	0.41	-0.15	0.00	0.19	0.00	0.05	0.30	0.26	0.00
Farsides and Woodfield (2003)	NEO-FFI	GPA (Major Yr1)	432	0.03	0.04	0.11	0.04	**0.17**	GPA	0.03	0.53	-0.02	0.68	0.11	0.02	0.04	0.41	0.17	0.00
Farsides and Woodfield (2003)	NEO-FFI	GPA (Major Yr2)	432	0.00	0.00	**0.17**	0.11	0.06	GPA	0.00	1.00	0.00	1.00	0.17	0.00	0.11	0.02	0.06	0.21
Farsides and Woodfield (2003)	NEO-FFI	GPA (Major Yr3)	432	0.02	0.04	**0.24**	**0.15**	0.08	GPA	0.02	0.68	-0.05	0.30	0.24	0.00	0.15	0.00	0.08	0.10
Farsides and Woodfield (2003)	NEO-FFI	GPA (School Yr1)	432	0.04	0.04	0.14	0.05	0.06	GPA	0.04	0.63	-0.07	0.40	0.14	0.09	0.05	0.55	0.06	0.47
Farsides and Woodfield (2003)	NEO-FFI	GPA (School Yr2)	432	0.07	0.02	**0.24**	0.14	0.01	GPA	0.07	0.08	0.02	0.61	0.24	0.00	0.14	0.00	0.01	0.80
Farsides and Woodfield (2003)	NEO-FFI	GPA (School Yr3)	432	0.07	0.02	0.09	0.15	0.05	GPA	0.07	0.50	-0.04	0.70	0.09	0.38	0.15	0.14	0.05	0.63
Gray and Watson (2002)	NEO-PI-R	GPA (university)	300	0.00	-.09	.19	.15	.36	GPA	0.00	1.00	-0.09	0.12	0.19	0.00	0.15	0.01	0.36	0.00
Rothstein et al. (1994)	PRF	GPA sample1	225	-0.08	.09	.05	**-.23**	.04	GPA	-0.08	0.23	0.09	0.18	0.05	0.46	-0.23	0.00	0.04	0.55
Rothstein et al. (1994)	PRF	GPA sample2	225	-0.08	.05	**.17**	-.16	**.14**	GPA	-0.08	0.23	0.05	0.46	0.17	0.01	-0.16	0.02	0.14	0.04
Noftle & Robins (2007)	BFI	GPA_uni	10497	**0.04**	-.02	**.06**	**.03**	**.22**	GPA	0.04	0.00	-0.02	0.04	0.06	0.00	0.03	0.00	0.22	0.00
Noftle & Robins (2007)	NEO-FFI	GPA_uni	475	-0.08	.02	**.13**	.1	**.19**	GPA	-0.08	0.08	0.02	0.66	0.13	0.00	0.10	0.03	0.19	0.00
Noftle & Robins (2007)	HEXACO	GPA_uni	470	0.10	-.11	.05	-.03	**.2**	GPA	0.10	0.03	-0.11	0.02	0.05	0.28	-0.03	0.52	0.20	0.00
Noftle & Robins (2007)	NEO-PI-R	GPA_uni	425			**.13**		**.18**	GPA					0.13	0.01			0.18	0.00
Lievens et al. (2002)	NEO-PI-R	GPA_Y1	607	0.06	**-.12**	.09	-.05	**.24**	GPA	0.06	0.14	-0.12	0.00	0.09	0.03	-0.05	0.22	0.24	0.00
Lievens et al. (2002)	NEO-PI-R	GPA_Y2	413	0.03	-.02	.08	-.08	**.17**	GPA	0.03	0.54	-0.02	0.69	0.08	0.10	-0.08	0.10	0.17	0.00
Lievens et al. (2002)	NEO-PI-R	GPA_Y3	341	0.03	-0.04	**.15**	-0.10	**.19**	GPA	0.03	0.58	-0.04	0.46	0.15	0.01	-0.10	0.07	0.19	0.00
		tot N	**21775**						**Random model effects**	**0.03**	**0.30**	**-0.02**	**0.03**	**0.10**	**0.00**	**0.03**	**0.09**	**0.17**	**0.00**

Table 3.3. Meta-analysis of the relations between personality traits and academic success measured using GPA sort of metrics.

Course Grades/Participation				Correlations reported						N		E		O		A		C	
Study authors	measure	criterion	Sample	N	E	O	A	C	category	Fisher Z	p value	Fisher Z	p value	Fisher Z	p value	Fisher Z	p value	Fisher Z	p value
Conard (2006)	NEO-FFI	Attendance	186	-.16	.01	.05	.22	.34	participation	-0.16	0.03	0.01	0.89	0.05	0.50	0.22	0.00	0.34	0.00
Dollinger, Matyja and Huber (2008)	BFI	Attendance	338	-.04	-.07	.05	-.0	.05	participation	-0.11	0.14	-0.07	0.34	0.05	0.50	0.00	1.00	0.05	0.50
Farsides and Woodfield (2003)	NEO-FFI	Absenteism	432	-0.13	0.07	-0.02	-0.14	-0.16	participation	-0.04	0.46	0.07	0.20	-0.02	0.71	-0.14	0.01	-0.16	0.00
Oswald & al (2004)	IPIP BFM	Absenteism	644	-.05	.1	.04	-.05	-0.27	participation	0.05	0.36	0.10	0.07	0.04	0.46	-0.05	0.36	-0.27	0.00
Rothstein et al. (1994)	PRF_s2	Classroom performance	225	-.14	.17	.18	-.16	.1	participation	-0.13	0.01	0.17	0.00	0.18	0.00	-0.16	0.00	0.10	0.04
Rothstein et al. (1994)	PRF_s3	Classroom performance	225	-.04	.20	.14	-.26	-.01	participation	0.03	0.53	0.20	0.00	0.14	0.00	-0.26	0.00	-0.01	0.84
Stevens & Ash (2001)	NEO-PI-R	week of participation	253	.08	.08	.16	-.05	-.17	participation	0.04	0.41	0.08	0.10	0.16	0.00	-0.05	0.30	-0.17	0.00
Stevens & Ash (2001)	NEO-PI-R	no activities	253	-.07	.34	.05	-.01	.19	participation	0.07	0.15	0.34	0.00	0.05	0.30	-0.01	0.84	0.19	0.00
Farsides and Woodfield (2003)	NEO-FFI	Tutor Grade	432	0.07	0.00	0.07	0.11	0.22	participation	-0.05	0.21	0.00	1.00	0.07	0.08	0.11	0.01	0.22	0.00
Stevens & Ash (2001)	NEO-PI-R	hours of work	253	.07	.02	-.01	-.07	.03	participation	0.15	0.00	0.02	0.61	-0.01	0.80	-0.07	0.08	0.03	0.45
Chamorro-Premuzic and Furnham (2003b) JofRP, Sample1	NEO-FFI	Project	70	-0.25	-0.10	-0.30	0.13	0.36	coursework	-0.14	0.04	-0.10	0.13	-0.30	0.00	0.13	0.05	0.36	0.00
Chamorro-Premuzic and Furnham (2003b) JofRP, Sample2	EPQ-R	Coursework	75	-0.09	0.27	-0.27			coursework	-0.04	0.55	0.27	0.00	-0.27	0.00				
Dollinger and Orf (1991)	NEO-PI	Project	118	-.11	.22	-.08	-.07	.24	coursework	0.03	0.65	0.22	0.00	-0.08	0.23	-0.07	0.30	0.24	0.00
Dollinger, Matyja and Huber (2008)	BFI	Project	338	.05	.14	-.08	.16	.19	coursework	-0.07	0.30	0.14	0.04	-0.08	0.23	0.16	0.02	0.19	0.00
Hair and Hampson (2006)	BFI	Coursework	236	00	.12	00	.06	.13	coursework	0.07	0.27	0.12	0.06	0.00	1.00	0.06	0.34	0.13	0.04
Rothstein et al. (1994)	PRF_s1	Written performance	225	.03	-.12	.05	-0.07	.11	coursework	-0.07	0.27	-0.12	0.06	0.05	0.43	-0.07	0.27	0.11	0.08
Rothstein et al. (1994)	PRF_s2	Written performance	225	-.07	-.06	-.06	-.07	.07	coursework	0.08	0.20	-0.06	0.34	-0.06	0.34	-0.07	0.27	0.07	0.27
Busato & al 2000	5PFT	Acad1	409	0.00	-0.07	0.03	0.05	0.16	firstExam	-0.28	0.02	-0.07	0.57	0.03	0.81	0.05	0.68	0.16	0.19
Busato & al 2000	5PFT	First exam	409	0.06	-0.13	0.03	0.00	0.16	firstExam	-0.25	0.04	-0.13	0.28	0.03	0.81	0.00	1.00	0.16	0.19
Chamorro-Premuzic and Furnham (2003b) JofRP, Sample1	NEO-FFI	exam (essay) Y1	70	-0.28	0.05	-0.60	0.34	0.33	firstExam	-0.09	0.44	0.05	0.67	-0.60	0.00	0.34	0.00	0.33	0.00
Conard (2006)	NEO-FFI	mid-term	186	-.11	-.06	.11	.17	.31	firstExam	-0.12	0.20	-0.06	0.52	0.11	0.24	0.17	0.07	0.31	0.00
De Fruyt and Mervielde (1996)	NEO-PI-R	Grades first exam period	741	-0.16	0.10	-0.16	-0.01	0.35	firstExam	0.10	0.28	0.10	0.28	-0.16	0.08	-0.01	0.91	0.35	0.00
Dollinger and Orf (1991)	NEO-PI	mid-term	118	.1	.0	.3	.1	.21	firstExam	-0.11	0.24	0.00	1.00	0.30	0.00	0.10	0.28	0.21	0.02
Farsides and Woodfield (2003)	NEO-FFI	School Yr1	432	0.04	0.04	0.14	0.05	0.06	firstExam	0.00	1.00	0.04	0.54	0.14	0.03	0.05	0.44	0.06	0.36
Farsides and Woodfield (2003)	NEO-FFI	Major Yr1	432	0.03	0.04	0.11	0.04	0.17	firstExam	0.09	0.17	0.04	0.54	0.11	0.09	0.04	0.54	0.17	0.01
Furnham et al. (2003)	NEO-PI-R	Exam_Y1	93	.18	.36	.19	.10	.44	firstExam	0.00	1.00	0.36	0.00	0.19	0.00	0.10	0.04	0.44	0.00
Dollinger and Orf (1991)	NEO-PI	Exam Essay	118	-0.12	-.04	.06	.18	.17	exams	0.06	0.23	-0.04	0.42	0.06	0.23	0.18	0.00	0.17	0.00
Furnham and Chamorro-Premuzic (2004)	NEO-FFI	Average STATS exam	91	.04	-0.24	-0.06	-0.04	0.25	exams	-0.16	0.00	-0.24	0.00	-0.06	0.10	-0.04	0.28	0.25	0.00
Hair and Hampson (2006)	BFI	STATS exam	236	.09	.04	.1	.07	.14	exams	0.08	0.03	0.04	0.28	0.10	0.01	0.07	0.06	0.14	0.00
Bauer and Liang (2003)	NEO-FFI	PGI (predicted grade index)	265	.048	-.064	-.023	.086	.125	other	0.18	0.08	-0.06	0.54	-0.02	0.83	0.09	0.41	0.13	0.23
De Fruyt and Mervielde (1996)	NEO-PI-R	Number of re-exams	741	0.08	-0.06	0.08	-0.04	-0.25	other	0.04	0.71	-0.06	0.57	0.08	0.45	-0.04	0.71	-0.25	0.02
Goff and Ackerman (1992)	NEO-PI	acad comfort	147	-0.17	-0.09	0.42	-0.08	0.10	other	0.05	0.44	-0.09	0.14	0.42	0.00	-0.08	0.19	0.10	0.10
Goff and Ackerman (1992)	NEO-PI	hard work	147	-0.25	0.21	0.65	-0.13	0.27	other	-0.17	0.04	0.21	0.01	0.65	0.00	-0.13	0.12	0.27	0.00
Oswald & al (2004)	IPIP BFM	self-rating BARS	644	.15	.24	.35	.37	0.30	other	-0.25	0.00	0.24	0.00	0.35	0.00	0.37	0.00	0.30	0.00

Table 3.4. Meta-analysis of the relations between personality traits and academic success measured using mid-term exams, coursework and attendance.

SAT/pre-entry/Other									N		E		O		A		C	
Study authors	measure	criterion	Sample	N	E	O	A	C	Fisher Z	p value	Fisher Z	p value	Fisher Z	p value	Fisher Z	p value	Fisher Z	p value
Conard (2006)	NEO-FFI	SAT	271	-.02	.02	.15	-.06	-.05	-0.02	0.74	0.02	0.74	0.15	0.01	-0.06	0.33	-0.05	0.41
Chamorro-Premuzic & Furnham (2008)	NEO-PI-R	IQ	158	-.05	-.08	.02	.22	.05	0.11	0.17	-0.08	0.32	0.02	0.80	0.22	0.01	0.05	0.53
Chamorro-Premuzic & Furnham (2008)	NEO-PI-R	Gf	158	.11	.14	.01	.01	-.19	-0.05	0.53	0.14	0.08	0.01	0.90	0.01	0.90	-0.19	0.02
Dollinger, Matyja and Huber (2008)	BFI	verbal ability	338	.03	-.20	.21	-.12	-.01	0.03	0.58	-0.20	0.00	0.21	0.00	-0.12	0.03	-0.01	0.85
Lounsbury et al., 2003	PSI	IG (OTIS-Lennon)	175	.03	-.1	.12	-.0	.01	0.03	0.69	-0.10	0.19	0.12	0.11	0.00	1.00	0.01	0.90
Lounsbury et al., 2003	PSI	work drive	175	-.3	.28	.4	.3	.53	-0.30	0.00	0.28	0.00	0.40	0.00	0.30	0.00	0.53	0.00
Moutafi Furnham & Crump (2006)	NEO-PI-R	Gf	2658	-.01	.03	.09	.01	-.11	-0.01	0.61	0.03	0.12	0.09	0.00	0.01	0.61	-0.11	0.00
Reeve, Meyer and Bonaccio (2006)	IPIP	(6) Sociability	219	0.24	0.69	0.06	0.50	0.06	0.64	0.00	0.69	0.00	0.06	0.38	0.50	0.00	0.06	0.38
Reeve, Meyer and Bonaccio (2006)	IPIP	(7) Social Sensitivity	219	0.11	0.16	0.33	0.79	0.25	-0.04	0.56	0.16	0.02	0.33	0.00	0.79	0.00	0.25	0.00
Reeve, Meyer and Bonaccio (2006)	IPIP	(8) Impulsiveness	219	0.13	0.36	0.13	-0.01	-0.27	0.09	0.18	0.36	0.00	0.13	0.05	-0.01	0.88	-0.27	0.00
Reeve, Meyer and Bonaccio (2006)	IPIP	(9) Vigor	219	0.27	0.45	0.27	0.41	0.27	0.26	0.00	0.45	0.00	0.27	0.00	0.41	0.00	0.27	0.00
Reeve, Meyer and Bonaccio (2006)	IPIP	(10) Calmness	219	0.64	0.07	0.33	0.29	0.28	0.69	0.00	0.07	0.30	0.33	0.00	0.29	0.00	0.28	0.00
Reeve, Meyer and Bonaccio (2006)	IPIP	(11) Tidiness	219	-0.04	0.00	0.07	0.18	0.79	0.24	0.00	0.00	1.00	0.07	0.30	0.18	0.01	0.79	0.00
Reeve, Meyer and Bonaccio (2006)	IPIP	(12) Culture	219	0.09	0.20	0.51	0.51	0.29	0.24	0.00	0.20	0.00	0.51	0.00	0.51	0.00	0.29	0.00
Reeve, Meyer and Bonaccio (2006)	IPIP	(13) Leadership	219	0.26	0.50	0.38	0.23	0.15	0.11	0.10	0.50	0.00	0.38	0.00	0.23	0.00	0.15	0.03
Reeve, Meyer and Bonaccio (2006)	IPIP	(14) Self-confidence	219	0.69	0.60	0.26	0.12	0.06	0.13	0.05	0.60	0.00	0.26	0.00	0.12	0.08	0.06	0.38
Reeve, Meyer and Bonaccio (2006)	IPIP	(15) Maturity	219	0.24	0.17	0.37	0.49	0.70	0.27	0.00	0.17	0.01	0.37	0.00	0.49	0.00	0.70	0.00
Mottus, Allik and Pullmann (2006)	EPI-NEO	Cognitive Abilities Test	154	-.11	.23	.42	-.18	-.19	-0.11	0.17	0.23	0.00	0.42	0.00	-0.18	0.03	-0.19	0.02
Mottus, Allik and Pullmann (2006)	EPI-NEO	Cognitive Abilities Test	154	-.17	.21	.32	-.14	-.10	-0.17	0.03	0.21	0.01	0.32	0.00	-0.14	0.08	-0.10	0.22
Mottus, Allik and Pullmann (2006)	EPI-NEO	Cognitive Abilities Test	154	-.14	.23	.38	-.23	-.12	-0.14	0.08	0.23	0.00	0.38	0.00	-0.23	0.00	-0.12	0.14
Mottus, Allik and Pullmann (2006)	EPI-NEO	Cognitive Abilities Test	154	-.14	.24	.42	-.20	-.16	-0.14	0.08	0.24	0.00	0.42	0.00	-0.20	0.01	-0.16	0.05
Farsides and Woodfield (2003)	NEO-FFI	Iqv	432	-0.05	0.00	0.16	0.17	-0.07	0.09	0.06	0.00	1.00	0.16	0.00	0.17	0.00	-0.07	0.15
Farsides and Woodfield (2003)	NEO-FFI	Iqs	432	-0.09	0.00	0.01	0.02	-0.11	-0.09	0.06	0.00	1.00	0.01	0.84	0.02	0.68	-0.11	0.02
Farsides and Woodfield (2003)	NEO-FFI	A' Points	432	0.09	0.12	0.24	0.09	0.07	-0.05	0.30	0.12	0.01	0.24	0.00	0.09	0.06	0.07	0.15
Bauer and Liang (2003)	NEO-FFI	WGCTA	265	-.084	-.167	.147	.086	-.23	-0.21	0.00	-0.17	0.01	0.15	0.02	0.09	0.16	-0.23	0.00
Bauer and Liang (2003)	NEO-FFI	CSEQ_acad	265	-.213	.173	.458	.065	.332	-0.16	0.01	0.17	0.00	0.46	0.00	0.07	0.29	0.33	0.00
Bauer and Liang (2003)	NEO-FFI	CSEQ_pers	265	-.16	.443	.213	.047	.154	-0.08	0.17	0.44	0.00	0.21	0.00	0.05	0.45	0.15	0.01
Goff and Ackerman (1992)	NEO-PI	perfectionism	147	-0.26	0.25	0.33	0.03	0.54	-0.26	0.00	0.25	0.00	0.33	0.00	0.03	0.72	0.54	0.00
Furnham et al. (2003)	NEO-PI-R	BAI (beliefs about intelligence)	93	.07	.15	.09	.01	.29	0.07	0.51	0.15	0.15	0.09	0.39	0.01	0.92	0.29	0.00
Gray and Watson (2002)	NEO-FFI	PRE-GPA	300	.00	-.05	.01	.11	.22	0.00	1.00	-0.05	0.39	0.01	0.86	0.11	0.06	0.22	0.00
Hair and Hampson, 2009	BFI	alcohol	138	.22	.06	.03	.29	.27	0.22	0.01	0.06	0.49	0.03	0.73	0.29	0.00	0.27	0.00
Hair and Hampson, 2010	BFI	impulse	238	.20	.11	.07	.31	.70	0.20	0.00	0.11	0.09	0.07	0.28	0.31	0.00	0.70	0.00
Phillips et al. (2003)	NEO-FFI	Intent	125	0.0	.17	.13		.46	0.00	1.00	0.17	0.06	0.13	0.15			0.46	0.00
Noftle & Robins (2007)	BFI	SAT_verb	10497	-.05	.02	.2	-.03	-.01	0.03	0.00	0.02	0.04	0.20	0.00	-0.03	0.00	-0.01	0.31
Noftle & Robins (2007)	NEO-FFI	SAT_verb	475	.03	-.15	.2	-.05	-.09	0.04	0.38	-0.15	0.00	0.20	0.00	-0.05	0.28	-0.09	0.05
Noftle & Robins (2007)	HEXACO	SAT_verb	470	-.02	.07	.26	-.1	.05	0.05	0.28	0.07	0.13	0.26	0.00	-0.10	0.03	0.05	0.28
Noftle & Robins (2007)	NEO-PI-R	SAT_verb	425			.26		.00					0.26	0.00			0.00	1.00
Noftle & Robins (2007)	BFI	SAT_MATH	10497	-.07	-.06	.05	-.06	-.07	-0.07	0.00	-0.06	0.00	0.05	0.00	-0.06	0.00	-0.07	0.00
Noftle & Robins (2007)	NEO-FFI	SAT_MATH	475	.03	-.08	.02	-.05	-.03	0.03	0.51	-0.08	0.08	0.02	0.66	-0.05	0.28	-0.03	0.51
Noftle & Robins (2007)	HEXACO	SAT_MATH	470	-.08	-.04	.04	-.03	-.03	-0.08	0.08	-0.04	0.39	0.04	0.39	-0.03	0.52	-0.03	0.52
Noftle & Robins (2007)	NEO-PI-R	SAT_MATH	425			.05		-.06					0.05	0.30			-0.06	0.22
Noftle & Robins (2007)	BFI	GPA(high school)	10497	.03	.03	.01	.1	.22	-0.05	0.00	0.03	0.00	0.01	0.31	0.10	0.00	0.22	0.00
Noftle & Robins (2007)	NEO-FFI	GPA(high school)	475	.04	-.09	.03	.06	.10	0.03	0.51	-0.09	0.05	0.03	0.51	0.06	0.19	0.10	0.03
Noftle & Robins (2007)	HEXACO	GPA(high school)	470	.05	.03	.02	.11	.26	-0.02	0.67	0.03	0.52	0.02	0.67	0.11	0.02	0.26	0.00
Noftle & Robins (2007)	NEO-PI-R	GPA(high school)	425			.04		.25					0.04	0.41			0.25	0.00
Duff & al (2004)	16PFI	HPts (prior grades)	146	0.22	0.01	-0.01	-0.03	0.09	0.22	0.01	0.01	0.90	-0.01	0.90	-0.03	0.72	0.09	0.28
									-0.01	**0.02**	**0.02**	**0.00**	**0.10**	**0.00**	**0.05**	**0.00**	**0.06**	**0.00**

Table 3.5. Meta-analysis of the relations between personality traits and academic success measured using a variety of prior performance and Intelligence metrics.

3.5. Extricating the relations between intelligence and personality

As we mentioned in the first section regarding the relations between intelligence and academic performance, it is very difficult to find studies which do not find any correlation between intelligence and academic performance. However, as we've discussed in the last section, personality seems to provide a further explanation for the variability of human performance which is not accounted for by intelligence alone.

In the table 3.5 we summarised the results of a meta-analysis conducted to identify the strength of the relations between personality traits and some pre-cursors of academic success and some intelligence measures (this review is limited to studies which did not study only IQ and AP).

From this review and earlier literature, it is possible to summarise three main perspectives on the associations between personality and intelligence (Chamorro-Premuzic & Furnham, 2005; Reeve, Meyer, & Bonaccio, 2006): a traditional perspective in which a complete independence in assumed due to zero-order correlations: this demonstrates that there is very little overlap between the dimensions (i.e. Webb 1915). A second view is that personality and intelligence are conceptually independent, however the measures used to quantify them showed a certain degree of correlation (i.e. T. Chamorro-Premuzic & Furnham, 2004), Eysenck 1967). A third alternative posits not only that personality and intelligence are highly correlated, but that the way in which they correlate has a substantive impact on information processing and learning (i.e. Matthews).

In the 1990s the topic became an interesting area of debate and various studies reviewed the relations between varying degree of intelligence and dimensions of personality. For example, work carried out by the differential psychology group in Edinburgh brought up core data to explore such interactions. Austin, Deary and colleagues identified not only relations between intelligence and personality traits, but also allowed to quantify the strength of the relations in very big samples.

In this section we would like to focus on three studies which support the third view and which will be relevant for the discussion in the next chapter: Ackerman & Haggerstead (1997) produced a comprehensive review of the relations between intelligence, personality and interests; Austin et al. (1999, 2000) in which the authors quantified the deviations of varying degrees of the two dimensions, and Chamorro-Premuzic & Furnham (2003) who proposed an integrative perspective in differential psychology to account for the relations between personality and intelligence which differs somewhat from Matthews' cognitive-adaptive framework.

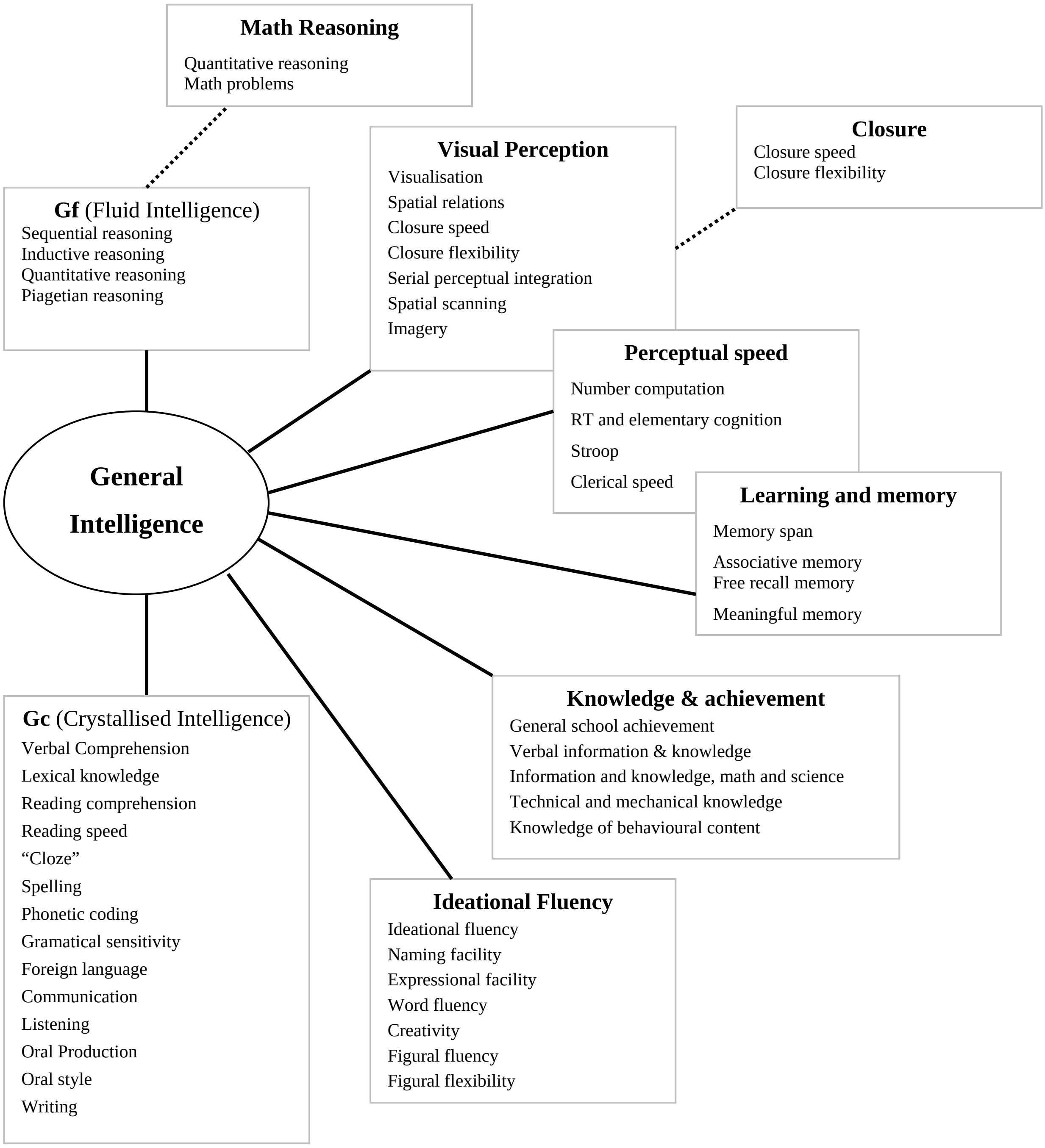


Figure 3.4. Hierarchical organization of the intelligence constructs.

Third-order constructs = General Intelligence; second-order constructs shown with solid lines; first-order construct shown with dotted lines.

3.5.1. A review of the interactions

In the 1990s, Philip Ackermann pursued vigorously a systematic research of the relations between classic IQ measures with ability and personality. This review (Ackermann & Heggersted 1997) still remains one of the key references in the field. The authors take an historical stance and look at the evolution of the concept of intelligence as a paradigm to test for abilities. In their discussion they also consider the issues of maximal versus typical performance and try to identify the core aspects of ability measured by standard tests.

This topic is investigated further by yet another 'heroic' (Oberauer, Schulze, Wilhelm, & Süss, 2005) review of the relations between intelligence and memory (Ackerman, Beier, & Boyle, 2005). Before delving into the summary of their findings, it is important to mention that Ackermann (1996) postulated that in the earlier stages of education IQ and achievement seem to be highly correlated while the strength of the relation (if any) diminishes for measures of higher level of educational achievement and occupational success (also in Sternberg), which might be because performance takes place in a 'typical' environment' rather than in test-conditions. Therefore intelligence as a typical performance measure should be more highly associated with crystallised abilities. In contrast, intelligence as maximal performance should be more related to fluid intelligence.

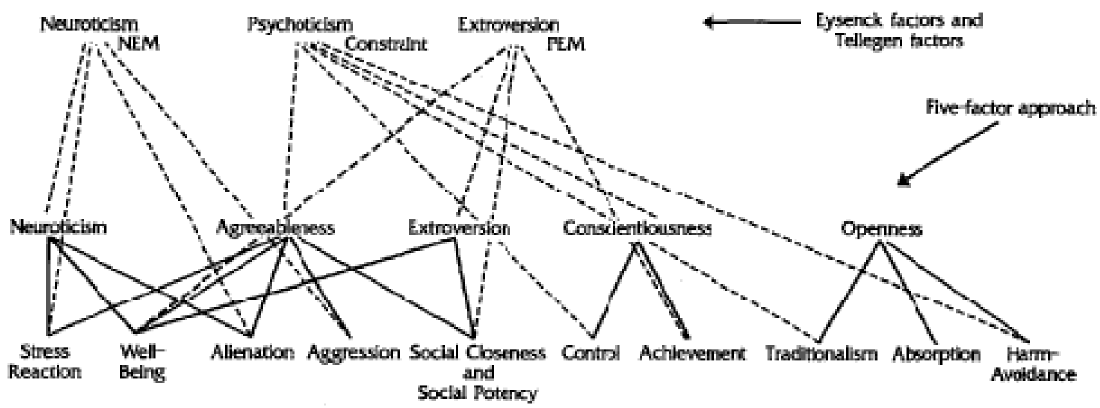


Figure 3.5. Personality constructs and their relations.

Lowest order constructs (from Tellegen, 1982); five-factor approach (FFA) constructs (from Costa & McCrae, 1992c; Digman, 1990; and others); highest order constructs (from H. J. Eysenck, 1970; and Tellegen & Waller, in press). Lines indicate both positive and negative correlational (not necessarily causal) relations. Solid lines indicate relations between Tellegen and FFA constructs. Dotted lines indicate relations between Tellegen and FFA constructs and H. J. Eysenck constructs. NEM = Negative Emotionality; PEM = Positive Emotionality. From: Ackerman: Psychol Bull, Volume 121(2).March 1997.219–245

In the meta-analysis Ackerman & Heggestad first identified a model of intelligence, which was derived from Carroll (1993). In the model, components of intelligence are listed in Figure 3.3 and are organised within the hierarchical structure as task-related abilities. The most interesting aspect of this model is that Visual Perception, Perceptual Speed, Learning and Memory, Knowledge and achievement and Ideational fluency are all listed as second order constructs in parallel to Gc and Gf. We will not argue against this model, even if there is evidence that the way in which abilities are presented might not be the best possible (see also Ackerman 2005).

The authors subsequently used three core theoretical constructs to classify personality dimensions. In particular they reduced Tellegen's 'lowest order traits' to three core factors advocated by Eysenck (Figure 3.4). Again, this model will be taken for granted and use it for its historical value. As discussed above, however, the number of traits is still a subject of dispute and the model suggested by Ackerman is only one possible interpretation.

The high number of variables involved in the meta analysis makes it quite difficult to produce a satisfactory representation, but the results of their efforts allowed to identify a number of patterns. The most notable are that personality traits within the Neuroticisms/NEM and Psychoticisms/Constraint sphere are generally negatively correlated with ability whilst those in the Extraversion/PEM tend to have positive correlations with intellectual abilities.

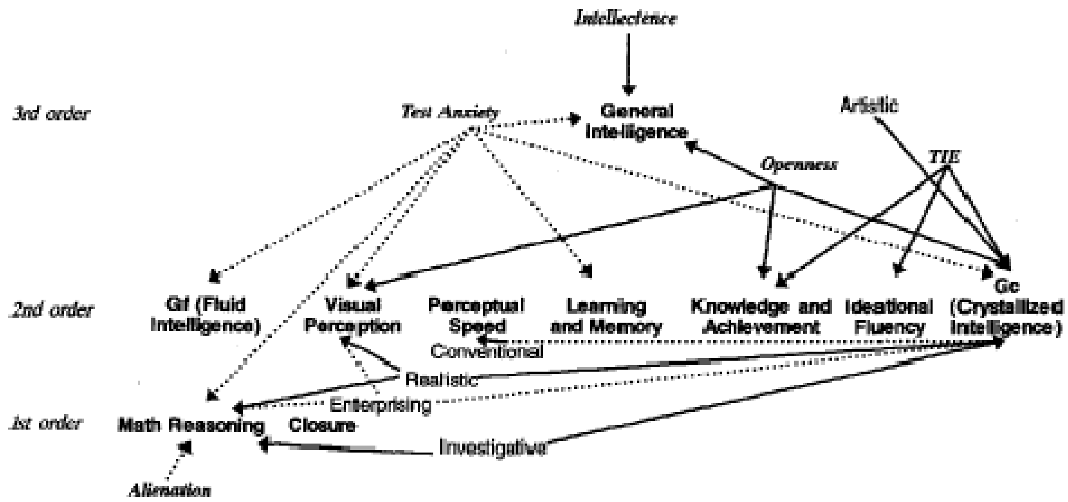


Figure 3.6. Associations between abilities (bold), interests (regular), and personality (italic) traits. Solid lines = positive correlations; dotted lines = negative correlations. TIE = typical intellectual engagement. From: Ackerman: Psychol Bull, Volume 121(2). March 1997.219–245

The authors also explicitly identified the relations between interests (TIE) and abilities with a strong link between openness to experience. Another interesting observation was that test anxiety, even though highly correlated with neuroticism didn't fit well within the taxonomic personality structure they used, but also that it correlated negatively with abilities.

From the meta analysis presented, we can extrapolate two core points for discussion. Firstly the historical evolution of the concepts of both personality and intelligence makes it quite difficult to individuate reliable patterns, but patterns do exist and need to be investigated further. There is evidence that personality and intelligence are not completely independent and consistent, but small effect sizes are present in the correlations between IQ and personality. Austin et al. (Austin et al., 1997; Austin et al., 2002; Austin et al., 2000) summarised the relations based on Ackerman's review as follows: N and P have a small negative correlation with magnitude of 0.1; E and O have small positive correlation with intelligence with magnitude of respectively 0.1 and 0.3.

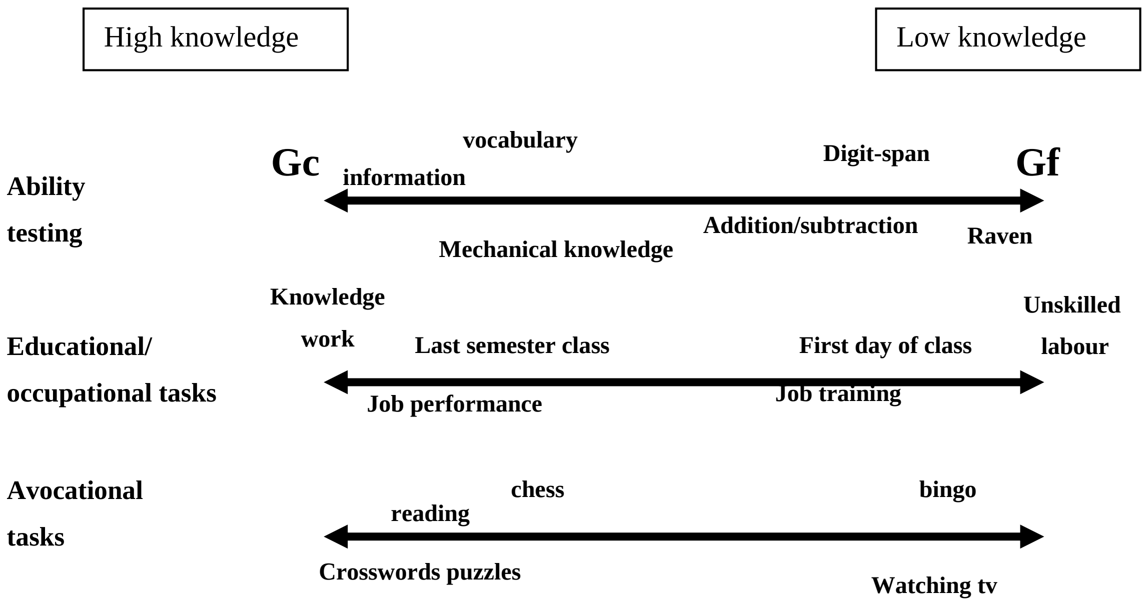


Figure 3.7. High knowledge and low knowledge across different domains. From Ackermann 1999.

Secondly, subject to the criticism of which model is used to reduce the complexity of the relations, it seems that there are a number of mediating factors which neither Intelligence nor personality can clearly pin down. These factors should be taken into account when evaluating the amount of variance explained by the relations between personality and intelligence.

Matthews (1999) attempted to include these factors as content of experience, which is obviously idiosyncratic and might be able to account for the adaptivity of the relations between cognition and personality. Ackerman (1999) attempted to put together traits and knowledge as determinants of learning by separating the domain of knowledge as parallel streams. (Figure 3.6)

Based on the three domains identified above, Ackerman continues by proposing a model which includes all the different aspects and tested it in a series of studies conducted with Rolfhus (Ackerman & Rolfhus, 1999; Rolfhus & Ackerman, 1996). The PPIK and the research carried out to test the idea are interesting for two main reasons: on one hand Ackerman & Rolfhus clearly identified differences in abilities and knowledge domains between their younger and older group, but also showed that maturation affects the different aspects of knowledge content differently.

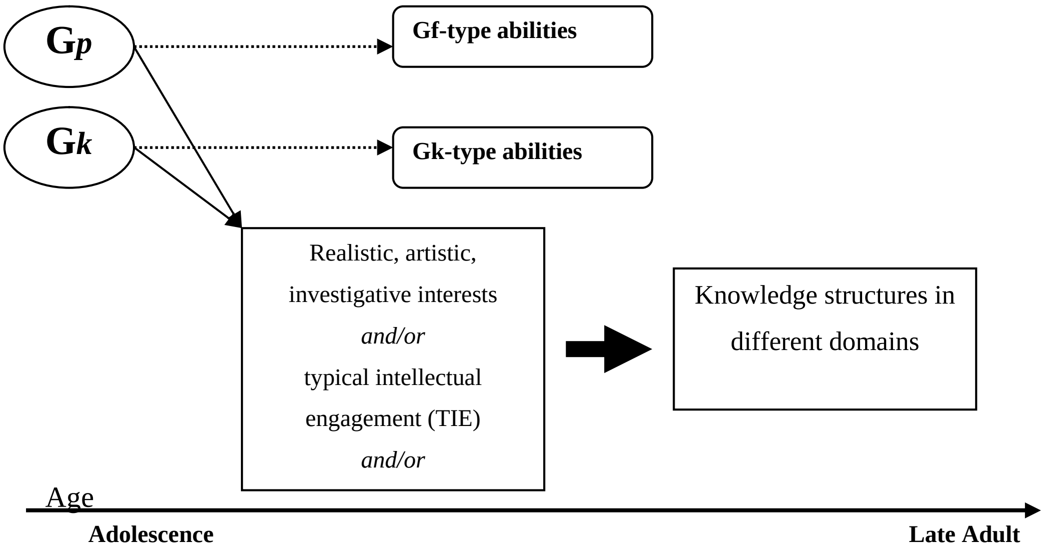


Figure 3.8. A theory of adult development, adapted from Ackermann 1996.

Note the stress on G associated to knowledge and process rather than crystallised and fluid intelligence, also note that the mediation interests and personality traits is simplified in this graph.

On the other hand, their research showed that there are correlations between personality and motivational traits and different knowledge domains. The most striking are the fairly strong negative correlation between extraversion and knowledge (range -.17 statistics, to -.36, American government), the positive one between openness and knowledge (range .1, chemistry and .46, American literature) and the relations between TIE (typical intellectual engagement) and knowledge (range .19, statistics, to .58 world literature). Even though these findings might not be surprising, they bring attention to the value of knowledge content and knowledge processes, rather than the measurement of maximal performance which IQ measures focus on.

3.5.2. A Statistical approach to individual differentiation

In 2002 Austin and colleagues published a replication study based on a decade of work conducted in Edinburgh. Specifically, they looked at the relationships between intelligence and personality mapping the differentiation of personality based on levels of g. In their research they hypothesised that the shape of the distributions of scores using the big 5 traits is dependent, rather than simply related, to the level of intelligence. Here we focus on two core publications which identified the progress of this approach.

The first study (Austin et al., 1997) was carried out with Scottish farmers (n=210) who offered a reasonably representative sample of distributions of intelligence. In this article Austin and colleagues suggested three possible hypotheses. The first one was derived predominantly from the literature which reported inconsistent findings about the relations between neuroticism and intelligence. Earlier in this chapter we already discussed the Neurotic cycle proposed by McKenzie (1989 and detailed in Chamorro & Furnham 1998) which highlighted that the interrelations between neuroticism and performance might actually be more complicated than in a simple linear correlation.

Austin et al. believed that personality in general is more differentiated at higher rather than lower levels of mental abilities. They also indicated that mental abilities are more differentiated at lower levels of N. The third hypothesis was that intelligence affects the correlation for specific pairs of personality measures.

Even though evidence supporting personality differentiation based on intelligence is sparse, Brand, Egan & Deary (1994) demonstrated that there are regular relations between intelligence and personality as measured by the MBTI. For example they contested that whilst IQ is represented with a U-shaped function mapped with extraversion-introversion, judgement-perception and judgement by thinking-feeling, the relations between perception by sensing or feeling is more of a linear trend with high scores of perception by feeling correlated with higher scores of IQ. Their second and third hypotheses that abilities are

structured differently at different level of personality dimensions was founded on the lack of linear relations between the constructs purported by Eysenck and Brand and colleagues.
The results presented by Austin et al. are quite interesting. Firstly they found that there were no significant differences in the average personality scores in the high and low performing groups based on the intelligence scores. They did show that instruments were biased in their measurements and indicated that in the lower performing group all reliability indices were lower which may suggest that the NEO is not appropriate for lower levels of IQ. Secondly, when the intelligence scores were split into two separate scales (NART for Gc and Raven for Gf) and related with N, results identified a quadratic relationship which fitted the expected U-shaped function. Openess, which is known to have similar theoretical commonalities with the MBTI sensing/intuition, was found to be non-linear with an expected trend with higher IQ with higher level of O. Finally, contrary to expectations, Austin et al. were not able to identify patterning of correlations of pairs of trait with IQ, but noted that this might be due to the confounding effects of the variability between the levels of IQ and personality.

In the second study Austin et al. investigated the relations between personality, measured using the Cattell 16PF and intelligence, using the culture fair test in over 30.000 participants (about half policemen and half felons). One of the reasons why the Cattell 16PF is interesting is that one of the second order factors is normally termed ability, and in this case direct comparisons can be drawn between self-reported ability and the scores from the CFIT. In fact, Austin et al. were able to show that the correlations between these two dimensions were consistently affected by the level of N across all groups as expected. However, the interpretation of the changes in structure of ability based on personality (and N in particular) is more controversial.

The research conducted in Edinburgh is useful for two core reasons: on one hand the idea that there is a simple linear correlation between intelligence and personality traits was challenged and that different levels of personality (N and O in particular) map on different levels of ability. On the other hand the methods used to investigate the relations is very useful as it highlighted the potential methodological flaws behind simple correlational studies.

3.5.3. A ‘third way’ in differential psychology: two integrative models and a possible alternative for all

Chamorro-Premuzic & Furnham (Chamorro-Premuzic & Furnham, 2005; 2006), carefully considered the possible relations between intelligence (as cognitive abilities), personality and self-assessed intelligence (as non-cognitive) with academic performance and offered yet another alternative in which the concept of Intellectual competence (IC) should be pursued by researcher to better characterise differences in human performance. IC also referred to as intelligent personality is encompassing three core components: performance, development and confidence.

According to this model, performance can be measured with objective tools such as timed power test, reactions times and measures of fluid intelligence; developmental aspects can be assessed using real life achievements, test of crystallized intelligence, academic or occupation performance. Confidence can be subjectively assessed using self-report inventories or core self-evaluation (Chamorro-Premuzic & Furnham 2006).

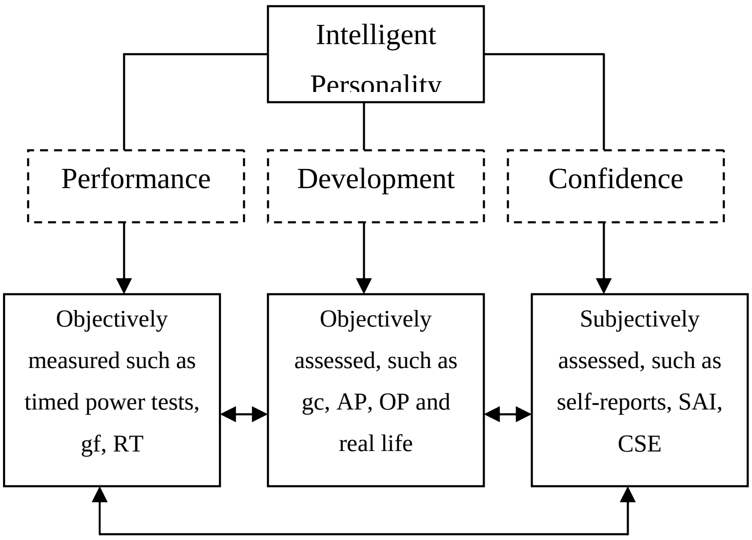


Figure 3.9. Intellectual competence and intelligent personality; from Chamorro-Premuzic & Furnham 2006, p 260.

Abbreivations gf=fluid intelligence, RT=reaction time, GC=crystallised intelligence, AP=academic performance, OP=occupational performance, SAI=self-assessed intelligence, CSE=core self-evaluation

According to this explanation, however, it might be difficult to identify homogeneous traits to determine academic performance: for example, extroverts might be advantaged when assessed orally or in group projects, whilst introverts outperform their peers in tasks which require extended period of independent studying. In the meta-analysis compiled in the previous section, we also identified the conflicting evidence of the relations between certain traits and certain forms of assessment (or performance criterion). For example A and E had an impact in predicting performance when class participation or group projects were used as criterion for AP, but didn't have any value in the prediction of exam performance. According to this model academic performance (AP) is correlated with both IQ and self-assessed intelligence (SAI) and AP is influenced by specific processes causing individual differences. Furthermore, intelligent personality is characterised by three main paths and it is mediated by non-cognitive elements. (See Figure 3.8)

For example, we have seen how N and E, for example, mediate the correlation between actual ability and test performance (i.e. N determining higher levels of test anxiety, or E determining test-style preference). Intelligent personality can also partly explaining differences in confidence or self-assessed abilities which we found to be relevant in predicting performance independently from ability. In their assessment, the authors are cautious when considering the inclusion of the assessment of intelligent personality in the real world:

> "Admittedly, when the intelligent personality is extended to real-world achievement criteria, the construct may become even more diversified —in fact, so diversified it may seem meaningless to conceptualize a predefined combination of traits underlying the intelligent personality. However, the fact that the same combination of personality traits may not be extrapolated across different performance criteria should not discourage researchers from conceptualizing and assessing the personality determinants of success in each context or setting. Rather, a robust and methodical examination of the different performance criteria will shed light on to the different algorithms that represent the intelligent personality in different circumstances or environments, which merely requires identifying moderating variables in the relationship between the Big Five and different indicators of performance. As long as the processes determining individual differences in a specific outcome are defined, the "right" combination of personality traits can be identified, too." (Chamorro-Premuzic And Furnham 2006, p 261)

The model proposed by Chamorro-Premuzic and Furnham has the major advantage of proposing an open approach to mediating variables and contrary to some restricted views of IQ and abilities, also attempt to include non-cognitive abilities in the equation like we observed in Ackerman's and Matthews' models. One interesting element of this model is the particular use of the terms *assessment* and *measurement* as well the specific inclusion of *self-assessed measures* (hence subjective) as opposed to objective measures in the prediction model. Austin and colleagues (2002) already pointed out that the self-assessment of personality might be affected by the level of abilities simply because the tests are standardised with high ability groups in the first instance. (Mõttus, Pullmann, & Allik, 2006; Paunonen, 1998), however, were not able to observe major differences in the self- and other-level of reported personality at the domain level, but did find some in the facets (as these were identified in an Estonian sample, it is possible that the differences are simply caused by the translation of the items).

3.6. Avoiding theoretical entrenchment

From the evidence presented in this chapter it is clear that both personality and intelligence constructs have a moderate ability in predicting academic performance at university, but also that the power of prediction decreases over time (or growth). Using either one of them or both, would allow a researcher to draw a complex picture of the students. However, what emerges from the literature is that the combination of feasible interpretations produces a picture that might be *too* complex to be usable for the individual student and that the selection of measures or constructs comes with a set of implicit assumption which brings no better clarity in explaining the manifestation of complex behaviours. Going back to the example of the athlete competing at different levels, other elements affect performance to the higher levels and, if all had equal talent/ability, the nature of their training as well as motivation and attitudes could equally contribute to achievement.

Earlier we mentioned the problems with the use of IQ tests; in particular it would be unreasonable to ask students to complete a full range of intelligence tests encompassing both measures of fluid and crystallised intelligence before they start. Focusing too narrowly on specific abilities would not allow accounting for individual variability. In addition, the sample under examination involves a selected group of highly performing students prior to admission. Even though this idea will be tested in more detail, from the existing literature we could easily infer that there is little variability between high school grades and performance at university as a group. Furthermore, even if we actually did present the students with a large IQ testing battery, this would not tell us anything about what they think about their abilities and how this affects their capacity to perform well under stress. Another issue to keep in mind is that even though we will use some objective measures of AP in maximal performance conditions (i.e. examinations) the usage of e-learning is in the domain of typical performance and is a continuous activity.

With regards to assessing personality, one is entitled to ask which model should be used? In our review we specifically focus on the big-five model, however, even in the review we've seen that Eysenck's EPQ or Cattell's 16PFI were used not helping to clarify the differences in performance.

Although there is some coherence in the links between the differential levels of neuroticism/emotional stability and academic achievement as determined by test scores in exams, and between levels of conscientiousness and openness and extraversion with some aspects of academic success, the meta-analysis carried out also allowed us to observe that personality measures actually account for a very small amount of individual variability. Often the sample under consideration is the main cause of such differences. Furthermore there is an overwhelming amount of evidence that the relations at domain level are much weaker than facets-level with subscales of the NEO and IPIP showing a greater power in predicting academic performance than the trait measure they belong to.
The analysis also identified an inextricable relations between N and O with performance, especially when results are looked at with a focus on maximal performance (i.e. exam grades/GPA). The path models taken into consideration placed the trait O as mediator between IQ measures and performance, but it was also shown that O could be mediated by other factors, of which the deep approach to learning was the most relevant. The latter relation is of particular relevance to this argument as it showed that the weight of a personality trait and a measure of style or preference carried the same weight (and prediction power) in the path model.

There are two further reasons why measures of personality should not be used. The appropriateness of the instruments has been criticised because of the existence of internal variability in measures of traits and levels of ability. In particular the effect of 'faking good' might become an issue with our particular sample of psychology students. Because they have some knowledge of personality they might tend to match desirable patterns and invalidate the usefulness of the instrument. Additionally we have seen from the model proposed by Matthews and the self-criticisms in Chamorro-Premuzic & Furnham, that patterning, or the correlation of multiple traits and specific behaviour might be just too complex, especially when considering facets rather than traits.

All these factors contributed to shape the methodology of this research and lead us to rule out the use of both measures of intelligence and standard personality inventories as not practical nor usable tools in identifying relations between individual differences and behaviours.

Nevertheless we don't intend to rule out the important aspects learnt thus far and in particular the relations between self-assessment, preferences and interests and performance in a typical engagement with domain knowledge.

To achieve this, Ackerman's focus on processes and knowledge seems quite useful and Matthews' interrelations between cognition, dispositions and context allows to re-address the discussion on processes and behaviours. In the literature in individual differences, measures of styles might be the answers to find out methods which are less entrenched, go from distal to proximal and focus on the applications and practical aspects making them more affordable, more usable and more importantly more efficient for students' learning. Styles measures are preferences, they don't need to be completely stable over time and are targeting 'typical' performance.

In the next chapter we will take into account this literature, place styles over the backdrop of intelligence and personality research and uncover the reasons why styles were selected to identify individual differences in our sample.

Chapter 4. A matter of styles: using ‘preferences’ to understand behaviour

In the first two chapters we considered how learning technology could be used to aid students’ learning in higher education and suggested that the creation of individual profiles based on psychological theory could be used to improve our understanding of the interaction between students and modern education, of which e-learning is a small part. In the last chapter it was made clear that although intelligence and personality moderately affect students’ achievement, in the effort of providing a detailed picture of the student, profiles containing standardised measures of intelligence and personality might ultimately lead to an overly complex and unusable profile with a fairly low prediction power.

However, informed instructional decisions in the design of e-learning tools must be supported by a deeper and more structured knowledge of the individual students, and we suggested that using *measures of styles* to explore individual differences might be a possible alternative. This also emerged from the practical applications in machine learning reviewed in chapter 2.

In this chapter, we take on the challenge to detail what is meant by *measures of styles*. We also attempt to provide a unitary framework of interpretation. This is partially built on the analysis of a number of well-established integrative approaches, and partially from a novel approach to the exploration of the literature using data mining and information visualization techniques (InfoVis). This approach aims to be more comprehensive and unbiased than the previous attempts, takes into account individual differences as inferred from behaviours and from the nexus between personality and abilities, but always maintains a firm practical relevance for the context of education and potential applications in learning technology.

Finally, because some measures of styles have been used with instructional technology in successful applications of personalization we will review a number of cases of good practice. The theoretical and practical review provides the evidence in support of the selection of four measures of style (which will be described in more detail in chapter 6) and an essential conceptual map to steer the discussion of data in later chapters.

4.1. Why measures of styles?

In the last chapter we were able to conclude that using personality traits as predictors for academic success may account for some of the variance in performance, however, especially with Openness and Conscientiousness we identified a couple of models in which traits were mediated by other factors, not accounted for by either intelligence or personality. The path analyses presented by Chamorro-Premuzic or Blickle (reproduced in figure 3.2) are two examples.

Duff (2004) suggested that there are two core models which underlie the research on styles: on one hand Furnham (Furnham & Medhurst, 1995) proposed a model for the role of styles in relation to intelligence, personality and achievement (Figure 4.1), predominantly drawing from a cognitive psychology perspective and exploring how learning takes place as students process information. On the other hand the 3Ps model advocated by Ramsden (1992) which we have looked at in one of the latest revision by Biggs (1999) in chapter 2 (Figure 2.3) was described as “a model of student learning in context”.

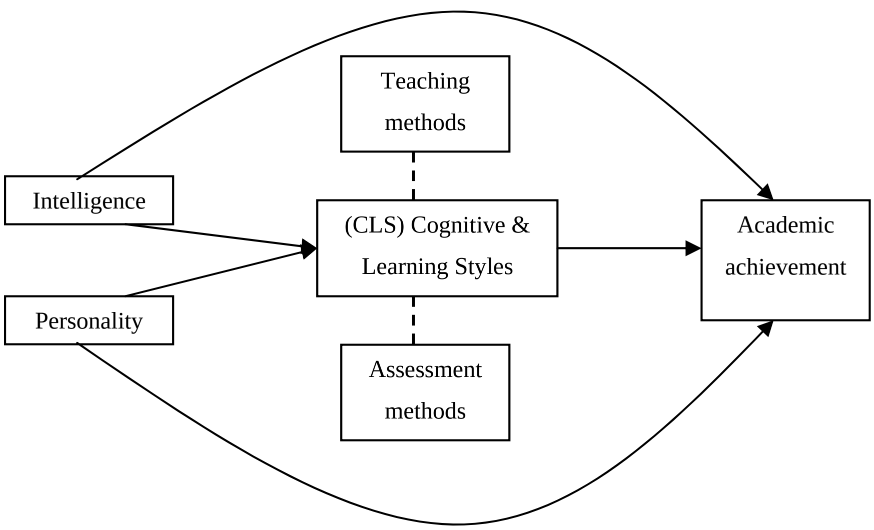


Figure 4.1. Furnham’s model of the predictors of performance.

Both models will be called into the argument in the review of styles, but we would like to focus particularly on Furnham's model here, as it allows discussing of three controversial points which have been argued in the literature. Firstly, Furnham implied that intelligence and personality are independent predictors of achievement (as indicated in chapter 3, however, assessment issues should be considered carefully). Secondly, both intelligence and personality are predictors of cognitive and learning styles (CLS), which then moderate academic achievement. This is possibly the most contested point as the relations between intelligence, personality and CLS are far from clear and some authors argue that personality aspects coincide with CLS. Thirdly, teaching and assessment methods are independently related to CLS and academic achievement. Furnham, however, endorsed the intuitive idea that a fit between one's preferred CLS and teaching/assessment methods should be beneficial for student learning.

Based on the review in the last chapter it was argued that the 'personality component' produces a more complex relation: for example, it is unlikely that a shy and introverted individual would perform as well in a group project as in an individual assignment. In another ideal example, an intelligent, but not so conscientious student will have less difficulty than an overzealous, but not so intelligent student, in achieving a good grade. However, if a student is able to identify effective strategies to cope with the workload, based on their preferred ways of processing, learning and thinking -therefore maximising what suits them best-, by factoring out both IQ and conscientiousness we might find that students with slightly lower IQ and little C might perform as well as more intelligent or more conscientious individuals. Some might attribute the success to effective teaching, what we have noted as 'aligned teaching' (i.e. Biggs in chapter 2): in the 3Ps model, a lot more emphasis is given to the interaction occurring in the learning context. However, the very concept of alignment implies that there must be a reference to align to; even if literature on alignment superficially refers to a balancing act between elements with a *contextual* rather than *universal* value, throughout this thesis we will argue that measures of styles, with their intrinsic value of self-appraisal and their limitations, might be the best candidate to provide such a reference for instruction, learning and applied to instructional systems using learning technology.

The interaction between the CLS component, teaching and achievement is fairly intuitive. However, as Furnham (1995) also indicated, the nature of the interaction has to be tested more thoroughly. Recently, Lori Breslow (2009) pointed out a striking concern which emerged from students' interviews in her research evaluating the effectiveness of intervention on studying and learning strategies: she intimated that students from both MIT and Cambridge in her studies reported that they knew what grade they could achieve before they had even started a course. Any form of teaching and learning implemented which could make students' life easier was very much welcome and increased their overall opinion and satisfaction of the courses, but, when one looked at grades, there was no discernible difference. From a cynical stance, one would question the value of pedagogical efforts and curriculum changes: if the underlying approach characterising students' effort is not leading to *better* performance of *better* learning, then the principles upon which evaluation is carried out are undermined.

Certainly these observations suggest that Furnham's model might be an oversimplification of the relations between abilities, personality aspects, CLS and academic performance. However the centrality of CLS is arguable as the value of styles remains to be ascertained. Hence, it will help to look at the conceptualization of styles in more detail.

According to Sternberg (Sternberg & Grigorenko, 1997), a reason for the sense people have that styles exist, may be:

> "(…) that they account for variation in performance that abilities do not account for, and that they may be important in a variety of real-world settings, such as the school, the workplace, and even the home" (Sternberg & Grigorenko, 1997, p. 147).

This idea that intelligence and personality do not explain the full range of performance has been presented in the last chapter. Motivation has also been used in the attempt to justify the interaction between abilities and achievement. Mayer (1955), and Vroom (1966), for example believed that *general performance* is a multiplicative function of intelligence and motivation, where motivation can be conceptualized in terms of personality characteristics (Rindermann & Neubauer, 2001). However, as indicated earlier, motivation is not equating to personality (even though it might be strongly driven by Conscientiousness or Openness). This interactive model, and the integrative reviews in the last chapter, suggest that both intelligence and personality comprise salient individual differences, which influence performance: generic intelligence/abilities (or what a person can do) afforded by specific abilities, facilitates understanding and learning. Broad personality factors (how a person will do it) through certain traits enhance or handicaps the use of these abilities.

Even though research on styles can be identified as far as Galton (Galton, 1883) who investigated tendencies to use imagery strategies as opposed to verbal strategies at recall, it is with Gordon Allport, who introduced the term cognitive style, that research in this area proliferated with a variety of models and definitions, often with little coherence or clear structure. Allport and Odbert (Allport & Odbert, 1936) defined cognitive styles as a person's innate, habitual or preferred mode of information processing. Witkin detailed:

> "cognitive styles are concerned with the form rather than the content of cognitive activity. They refer to individual differences in how we perceive, think, solve problems, learn, relate to others, etc. (…) cognitive styles are pervasive dimensions. They cut across the boundaries traditionally-and, we believe, inappropriately-used in compartmentalizing the human psyche and so help restore the psyche to its proper status as a holistic entity." (Witkin, Moore, Goodenough, & Cox, 1977), p 15).

Messick (1976) further expanded this definition by adding that cognitive styles are:

> "representing consistencies in the manner or form of cognition, as distinct from the content of cognition or the level of skill displayed in the cognitive performance." (p. 5)

The first systematic study of the variation cutting across the boundaries between personality and intelligence resulted from the work of Klein (1951) and Witkin (Witkin et al., 1954). Klein distinguished two types of individuals: *sharpeners*, who notice contrasts and maintain a high degree of stimulus differentiation, and *levellers*, who notice similarities among stimuli and ignore differences (Klein, 1951; Klein & Schlesinger, 1951). Witkin also classified two types of people in his perceptual tasks: *field dependent* (FD) people who exhibit high dependency on the surrounding, and *field independent* (FI) who exhibit a low dependence. It is notable that Witkin's types can be considered as opposite poles of a continuum, and it was pointed out by Witkin and colleagues that a large number of individuals didn't fall distinctively in either category. It was also found that FI and FD people present significant relations between perceptual tasks, personality characteristics and social behaviour. For example people in the FD group made greater use of external social referents and were more attentive to social cues than people in the FI group (Witkin & Goodenough, 1981).

These first two major contributions set the stage for most of the following research: they inspired other researchers to fill the gap in the overlapping areas of personality and intelligence already advocated by Cronbach (1957), but also left some major issues open for debate and confusion, which characterised the following evolution of the research on styles. In particular, both authors agreed that styles were stable over time and should be related to personality (also fairly stable). Witkin's ideas strongly impacted upon later research in cognitive styles arguing that a 'style' is a "broad bipolar dimension", which, in principle, allows for a value-free construct/categorization. Klein, on the other hand, even if supporting

the idea of a value-free construct, seemed to imply that two separate orthogonal dimensions were more suitable to measure the extent of the presence of styles.
In much the same way in which the optimal number of personality traits dominated personality research, the 1950s were characterised by a proliferation of measures of styles capable of capturing individual differences in performance. Such trends continued up to the 1980s and became reason for concern for many researchers (Curry, 1983b; Messick, 1976; n Miller, 1987; Riding & Cheema, 1991) to name a few). Other notable measures (both for citations and replications) are the Kagan's impulsivity-reflexivity dimensions (Kagan, 1964, 1966), Kirton's adaptors-innovators (Kirton, 1976, 1994) and Paivio's verbaliser-visualiser (1971).

The creation of new measures of styles was mainly characterised by a number of independent and often entrenched interpretations. In her review, Kozhevnikov is critical of the fact that there were very few attempts to integrate the numerous cognitive styles:

> "the main experimental paradigm was as follows: a simple task with two or more possible ways of solving it was offered to a subject. In situations of uncertainty about the "right way" of performing the task, the subject would choose his or her preferred way. Because all ways of solving the task were considered to have equal value, it was assumed that the subject's choice revealed a preference, not an ability. A group of subjects was then divided on the basis of their performance via a median split, forming two opposing poles of a particular style. This approach led to a situation in which as many different cognitive styles were described as there were researchers who could design different tasks." (Kozhevnikov, 2007, p. 466)

To a certain extent, this is still true today, but it must be acknowledged that by the end of the 1970s and 1980s a number of reviews started to appear in the attempt to make the field more coherent. For example, just to provide a timeline, Messick counted 19 styles in 1976, Keefe listed 40 in 1988, Coffield et al. selected 78 in 2004, which demonstrates the core problem of this field.

Despite these integrative attempts (Coffield, Moseley, Hall, & Ecclestone, 2004a; Curry, 1983b; Messick, 1976; Miller, 1987; Riding & Rayner, 1991; Zhang & Sternberg, 2005) to mention some influential ones which are reviewed in this chapter), the question of what the difference between styles and personality or intelligence is (or what the relations between these constructs are, to turn the question around) remains still very open. For the purposes of this thesis, it is not essential to justify the theoretical underpinnings of the definitions, but it is necessary to address the problem to clarify the terminology used and provide a coherent stance with no margin for interpretations, as *styles* mean very different things for different researchers.

Definitions Cognitive Style	**% ***	**Definitions Learning Style**	
Cognitive styles are individual differences in processing that are integrally linked to a person's cognitive system. More specifically, they are a person's preferred way of processing (perceiving, organising and analysing) information using cognitive brain-based mechanisms and structures. They are partly fixed, relatively stable and possibly innate preferences.	66.0%	Learning styles are an individual's preferred ways of responding (cognitively and behaviourally) to learning tasks which change depending on the environment or context. Therefore a person's learning style is malleable.	40.9 %
Cognitive styles are complex, multifaceted psychological variables that affect the way a person prefers to process information. In particular, they refer to the way people solve problems, make decisions and undertake tasks. They are not tied to a particularly cognitive mechanism or structure. They are partly fixed, relatively stable and possibly innate preferences.	34.0%	A learning style is an individual's psychological repertoire of preferred learning processes and strategies that are used when learning. These preferred processes can be cognitive, affective, motivational and behavioural and they shape the social and personal aspects of an individual's learning performance.	36.4 %
Cognitive styles are relatively stable super-ordinate psychological structures and processes (possibly innate) that determine a person's preferred way of thinking.	23.4%	Learning styles are individual differences in the way a person processes information (i.e., their cognitive style) which determines their typical or preferred response (cognitive and behavioural) in a learning context. A person's learning style is relatively stable	27.3 %
Cognitive styles are trait-like individual differences in the way people think. They are strongly linked, or possibly the same as, personality traits.	15.2%		

Table 4.1. Summary of definitions given for cognitive and learning style. Source: Peterson, Rayner and Armstrong 2009. The survey and the votes received in round 1 of the Delphi study. *percentage of the sample (N=44) that strongly or mostly agreed with these definitions

The term styles, in fact, is rarely used by itself, and is mainly associated with the labels *cognitive*, *intellectual*, *thinking* and *learning* -styles. Depending on the model considered, the overlap with aspects or facets of personality and levels of abilities is more or less assumed or implied (see for example (Riding & Cheema, 1991; Zhang & Sternberg, 2000), often ignoring the clear definition of the terms, which also contributes to the growing confusion (Desmedt & Valcke, 2004) named it the 'styles jungle'). Nevertheless, there seems to be a certain consensus that styles equate to *preferences* and could be more or less stable for a specific individual across the lifespan (Messick 1976, Curry 1983, Peterson, Rayner, & Armstrong, 2009; Sternberg, 1999).

4.2. A first take on the definition of styles.

A useful starting point in understanding the slight differentiation between the meanings of styles is to look at what researchers in the area commonly refer to as *styles*.

In a recent review, Peterson et al. (2009) conducted a Delphi study to explore the level of agreement on what expert mean with the various terms. Table 4.1 is useful in identifying the core features of the definitions and to highlight the variations suggested. Some of the elements of these definitions follow straight from the original wording of Allport and Messick cited above, however the most interesting aspect of this paper is that whilst there seems to be a certain majority consensus, in the definition of *cognitive* and *learning* styles, not everyone means the same thing. This finding demonstrates the difficulty in summarising the area with a coherent theoretical framework. The same problem will feature across this chapter, as different studies have very different assumptions and understandings of the terms used.

Some comfort derives from empirical evidence in research like Sadler-Smith (2001) and Zhang & Sternberg (Zhang & Sternberg, 2000; Zhang & Sternberg, 2005), who showed that measures of *cognitive* styles and *learning* styles (at least for the instruments they used to measure them) were independent constructs. This lends support to the idea that careful attention must be paid to the definition of terms and the conceptual frameworks from which the instruments to measure styles were developed.

It also makes it impossible to go from a systematic review to meta-analytic studies as we did in the last chapter as methods used are different, samples are not homogeneous and the theoretical frameworks are beset by an ever growing number of models, theories and instruments.

An alternative starting point is to identify variations in the *use* of different terms/labels. For example, Peterson (2002) summarised key differences between cognitive style, cognitive ability, learning strategies and cognitive skill integrating work from Messick (1984) and Tiedermann et al. (1999). This approach provides a further differentiation grounded on a more detailed specification (table 4.2). However it is lacking in a number of aspects: the dimensions and labels were condensed, and the categories for comparisons don’t seem adequate enough to convey subtle differences. For this reason we propose an enhanced version (table 4.2) which takes into consideration a wider spectrum of sources and which also contextualises some of the aspects of the literature review in chapter 3.

Based on the definition of cognitive and learning styles in Peterson (2002) and Peterson et al. (2009) as well as the reviews carried out in the past 5 years (particularly Zhang & Sternberg 2005, Coffield et al. 2004, Desmedt & Valcke 2004 and also Miller 1987, Rayner & Riding, 1997; Riding & Rayner, 1991), her original representation (2002) can be expanded by specifying a number of categories: in particular it should be possible to visualize the relations with each other and create inter-dependencies between constructs both at the level of *measurements*, *scope* and *functions*. At the interface level we refer specifically to the meta-analysis conducted in the last chapter to identify where the concept may be mapped in a more comprehensive framework of differential psychology and we used Ackerman's hierarchical model of cognitive abilities as a general guideline (see figures 3.3 and 3.5).

The most important feature of this table is the clear semantic differentiation of *cognitive*, *learning* and *thinking* styles as well as the associated terms *styles*, *abilities* and *skills* (or *strategies*) which all appear in the literature, but never together. Even though in the literature published in the past couple of years the concept of 'learning patterns' has also emerged in the open discussion at the ELSIN (European Learning Styles Network) meetings, it seems to me that this label could easily fit in any of the three categories already proposed, therefore we will ignore it in this thesis.

Approaches to learning (as in Entwistle et al.) are also problematic to place as some aspects belong more to 'style' whilst others can be identified as 'strategy' –the differentiation will become more clear in the treatment of the ASSIST as a tool to measure them.
The distinctions proposed in this table will be useful in the next section when we point out some common themes in various integrative attempts, in the last section of this chapter to identify suitable measures and later in the thesis to specify appropriate student profiles.
This table is also crucial in the attempt to cluster the concepts/models in a meaningful way and at a later stage we will be able to test some hypotheses emerging from such organization.

	Cognitive Styles	Cognitive Ability	Learning Strategies	Cognitive Skill
Question	the manner or mode of processing	what is processed	general skills used to help processing	general skills used to help processing
Measure	typical performance measured on a bipolar scale	maximal performance unipolar scale (usually 0 upward)	general strategy use unipolar scale (usually 0 upward)	specific strategy unipolar scale (usually 0 upward)
Scope	value differentiated cannot be learned cuts across domains	value directional can be learned domain-specific	value directional can be learned cuts across domains	value directional can be learned domain-specific
Functions	as an organising variable	as enabling variable	as enabling variable	as enabling variable
Examples	wholistic or analytic	verbal or spatial abilities	mind mapping, goal setting	acrostics, puzzles

	Cognitive Styles	Cognitive Ability	Cognitive Skill	Thinking Styles	Learning Styles	Learning Ability	Learning Strategy
Question	the manner or mode of processing	what is processed	general skills used to help processing	the manner or mode of processing	general preferences for learning	general skills used to help processing	specific skill or strategy to achieve learning
Measure	typical performance	maximal performance	specific strategy	typical performance	preferences for methods of learning	general strategy use	general strategy use
	mediates BOTH IQ and personality measured on a bipolar scale or categorical exclusive	component of IQ unipolar scale (usually 0 upward) partially exclusive	affected by IQ and cognitive styles unipolar scale (usually 0 upward) non-exclusive	mediates BOTH IQ and personality unipolar scale (usually 0 upward) or categorical (presence/absence) partially exclusive	mediated by cognitive styles unipolar scale (usually 0 upward) non-exclusive	affected by IQ, mediated by styles and personality categorical non-exclusive	affected by cognitive skills and mediated by learning styles and strategies unipolar scale (usually 0 upward) non-exclusive
Scope	value differentiated cannot be learned cannot change over time cuts across domains no metaknowledge	value directional can be partially learned can change over time domain-specific partially aware (I'm better at…)	value directional can be learned can change over time domain-specific aware (I'm good at)	value differentiated cannot be learned can change over time cuts across domains no metaknowledge	value differentiated can be learned can change over time cuts across domains no metaknowledge	value directional can be learned can change over time cuts across domains partially aware (I'm better at…)	value directional can be learned can change over time domain-specific aware (I'm good at)
Interface	2nd order	3rd order	behaviour	2nd order & behaviours	2nd order	behaviour	behaviour
Functions	as an organising & control variable affects both modes and outcomes	as enabling variable affects outcomes	as enabling variable affects outcomes	as an organising & control variable affects both modes and outcomes	as BOTH organising and enabling affects both modes and outcomes	as enabling variable affects outcomes	as enabling variable affects outcomes
Examples	wholistic or analytic	verbal or spatial abilities	acrostics, puzzles	field dependent, oligarchic style	deep, surface approach	mind mapping, goal setting	memory strategies, note taking

Table 4.2. A summary of labels used in styles research.

At the top Peterson's (2002) summary. At the bottom a reviewed summary expanded and updated categorization based on a wider set of sources.

4.3. Integrative frameworks

Despite the proliferation of styles and the numbers of interpretations of what the concept of styles actually entails, we already mentioned that with the 1980s a number of researchers started to feel the need to put some order in the field and a number of reviews were published in the attempt to provide integrative models of the conceptualisation of styles. Here we consider six different models, selected partly for their prominence in the literature (i.e. number of citations), partly because the different aspects touched in each review are useful to lay the foundation for the selection of specific measures of styles which are proposed at the end of the chapter. The shortcomings and strengths of each approach will become apparent and at the end we propose a different a more comprehensive way of summarising the literature.

4.3.1. Curry's onion model

The first model considered in this thesis is Curry's 'onion' model (1983b, 1987). Curry organised theories and dimensions of cognitive and learning styles into three strata resembling the section of an onion: the innermost layer includes models of *cognitive personality*, the middle one models of *information processing* and the outer layer the *instructional preference* indicators. In her words:

> "learning behaviour is fundamentally controlled by the central personality dimension, translated through middle strata information processing dimension and, given a final twist by interaction with environmental factors encountered in the other strata" (Curry 1983, p.)

It is implied by this organization that styles are more stable in the inner layer and more changeable in the outer layer. Examples of models attributed to the different layers are Witkin's field dependence and the Myers-Briggs Types Inventory (MBTI) in the inner layer, Kolb's learning styles inventory in the middle layer and Stritter & Friedman Instructional preferences questionnaire in the external layer. Curry's work was influential for two reasons: on one hand this was the first attempt to provide a clear and concise specification of the terms and definitions as well as providing an integrative overview of the theory; on the other it was a simple model, which practitioners could easily understand and use in their everyday practice to improve instruction. Curry criticised openly the weakness of the psychometric properties of some of the instruments used to study styles and her attempt to evaluate the strength of the relations reported was a welcome move.

If we look back at the schematic representation of the definitions of styles (table 4.2), Curry contributed to forge the core distinction between *learning style* and *learning strategy.* She also argued for a coherent process in which learning is the product of the interaction between internal and external factors, but mediated by a specific cognitive interface.

4.3.2. A tighter connection between cognitive processes and styles

In 1987, Alan Miller set out to provide yet another integrative model; in his view,

> "One proven strategy for resuscitating a moribund field is to link it with a more vigorous one. In the case of cognitive styles, an obvious candidate […] is that of mainstream cognitive psychology." (Miller 1987, p 252).

Even if such a statement might be arguable, in his paper, Miller attempts to re-connect styles research to cognitive psychology by arguing that measures of styles are able to differentiate individual performance in the various sub-components of an information processing model. Such an attempt is not new: the original positions of both Witkin and Kagan were to identify differences which are *cognitive* in nature. However, Miller systematically layered the impact of styles on cognition. In his opinion,

> "all cognitive styles are subordinate to and reflect, a broad superordinate stylistic difference" (Miller 1987, p 253).

In the figure 4.2 his representation is made clear: the three core cognitive processes are specifically associated with stylistic patterns which influence cognitive processes and the way in which the three core modules interact to generate a response.

In a later paper Miller (1991) also attempted to integrate the information-processing model with an affective dimension and a more specific relation between styles and personality traits which allows to identify three core dimensions of styles-personality interactions: the objective-subjective, holistic-analytic and a third which encompass low to high emotional stability.

Another researcher who proposed a theoretically grounded model of styles within the information processing theory is Nosal (1990, in Kozhevnikov 2007). Nosal proposed a matrix-style organization in which the vertical axis represents different ways of information processing, the horizontal axis represents different levels of processing. The metadimensions of the vertical axis are four hierarchical levels (from simple perception to decision making) and on the horizontal axis four methods (from automatic encoding to conscious allocation of mental resources). With this matrix Nosal was able to place 12 cognitive styles (see Figure 4.3).

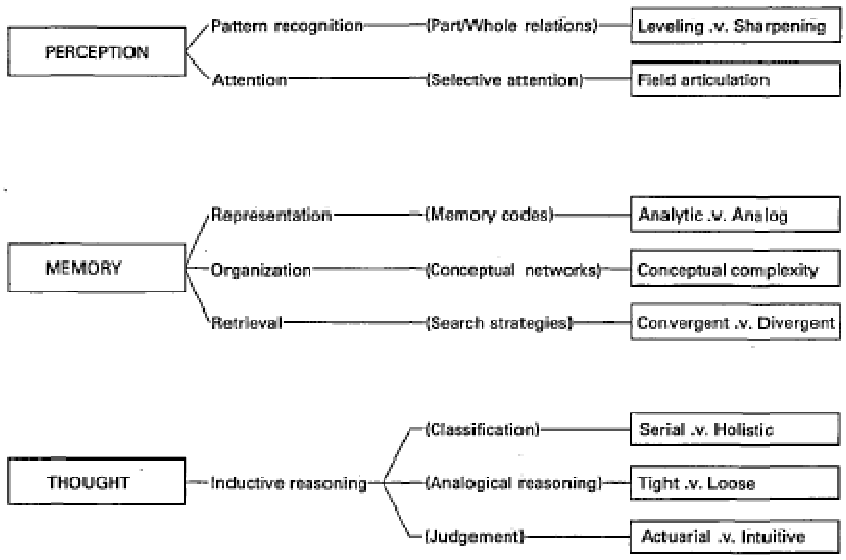
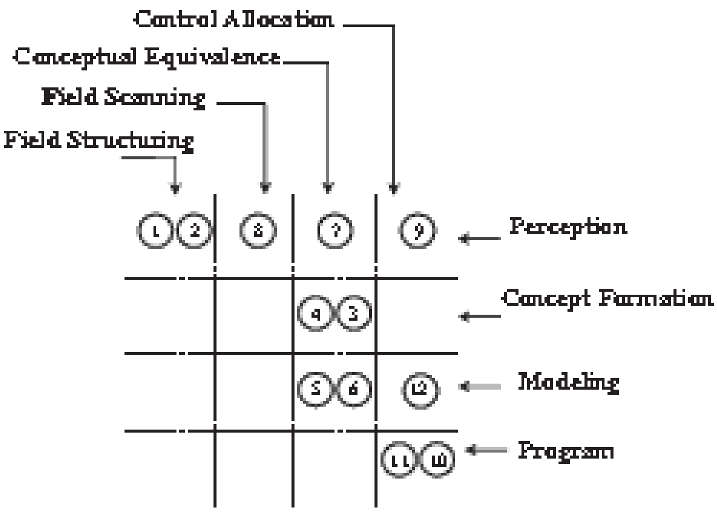


Figure 4.2. . Miller's hierarchical organization of cognitive functions and their relations with styles.



1 = field dependence–independence;
2 = field articulation (element vs. form articulation);
3 = breadth of conceptualization;
4 = range of equivalence;
5 = articulation of conceptual structure;
6= tolerance for unrealistic experience;
7 = leveling–sharpening;
8 = range of scanning;
9 = reflectivity–impulsivity;
10 = rigidity–flexibility;
11 = locus of control;
12 = time orientation.

Figure 4.3. Cognitive styles in relation to metadimensions and levels of information processing according to Nosal's theory.

This systematic approach to the classification of styles, based on a *stronger* cognitive perspective, provides the basis for a hierarchical classification of styles both in terms of their relations with each other and their specific interrelations with the processes of cognitive control and regulation.

Such an approach is useful for three reasons. Firstly, at the time of their writing, the mainstream study of cognitive abilities ignored the effect of individual differences: a systematic attempt to map the impact of individual differences on cognitive processes intentionally steered and re-focused styles research toward a more experimental approach in which the variation of performance could be partially explained by the strength of one's cognitive styles.

Secondly, even though both Miller and Nosal confidently produced a hierarchical model with a superordinate level or with metadimensions, in which certain styles operate at the superordinate level, Moran (1991) warned that the research on styles should avoid the quest for general 'knowledge-seeking' styles: in fact, after failing to pinpoint generic cognitive abilities, cognitive theories demonstrated the crucial relevance of specific and domain-related knowledge and expertise. In other words, researchers should be cautious in providing styles constructs which try to explain all functioning. In both models, the impact of general thinking styles is still mediated by basic cognitive abilities and domain-specific knowledge.

Miller (1991) also shed some light on the relations between students' styles and teaching methods and reflected on the application of styles in the classroom, critical for both practitioners and teachers. In particular he presented the risks associated with mismatched styles and expectations: in his teaching he believed that equipping students with more effective styles was essential to maximise learning. It transpires from Miller's writing that he considered a deep or strategic approach to be *better* for students: this view is well supported by the literature in education which shows that a strategic approach to study and learning generally led to better understanding and potentially better performance. However it is important to consider that this might not be the best approach for every student on every course. He attempted to make students more aware of their own preferred styles, but this generated a quite hostile reaction in his students. His attempt to encourage styles versatility in students highlighted the fact that some students simply do not benefit from a higher meta-cognitive evaluation of their own styles. Styles might operate at a higher, but automatic level: by eliciting a self-appraisal of styles and reflection, Miller demonstrated that some individual might experience a negative, emotionally driven experience, rather than a constructive one, which certainly raises some questions about the educational scope and applications of the utility of styles.

4.3.3. Riding's categorisation of styles

In 1991, another influential review was published which attempted to put order into the field of styles. Riding & Cheema (1991) summarised the empirical results of a decade of research comparing different dimensions and styles constructs. They contended that in many cases the dimensions used by different theories and instruments, were actually only different labels with major conceptual overlap. The core position was that cognitive styles and *learning strategies* are quite different:

> "[styles are] a fixed characteristic of an individual, while strategies are the ways that may be used to cope with situations and tasks" (Riding & Cheema 1991, p 195).

Based on this definition they postulated three core categorization of styles: a *wholistic-analytic cognitive style* family, *a verbaliser-imager cognitive style* family and a set of learning strategies constructs. Reflecting the research he conducted in Birmingham, Riding mapped the two cognitive styles dimension as orthogonal axis. Table 4.3 is useful to identify the mappings with other dimensions.

	Wholistic	**Analytic**
Witkin (1962)	Field dependence	Field independence
Kagan 1965	Impulsive	Reflective
Holzman & Klein 1954	Levellers	Sharpeners
Hudson 1966	Divergers	Convergers
Pask 1972	Holistics	Serialists
Das 1988	Simultaneous	Successive
	Verbaliser	**Imager**
Bartlett (1932) sensory modality	verbal	visual
Paivio (1971) ways of thinking	verbal	visual
Mental rotation tasks, number comparisons etc (Paivio 1971, Macleod 198x)	Propositional	Mind's eye

Table 4.3. The Mapping of the two cognitive styles dimensions and existing models. Based on Riding & Cheema (1991)

According to Coffield et al. (2004) Riding never defined the core of learning styles and criticised the fact that quite different perspectives were added under the label *learning strategies* without further specification. Experiential learning models (i.e. Kolb), orientation to study and instructional preferences (i.e. Biggs, 1979; Dunn & Dunn, 1975)) and even the MBTI (Myers, McCaulley, Quenk, & Hammer, 1999) were all allocated under the same umbrella category.

This might have been true for the first review (Riding & Cheema 1991), however such criticism is unfounded after a careful reading of Riding and Rayner (1998) which shows that, not only the definition was clearly specified, but also that the authors tried to systematically organise the literature on learning styles into four different categories, in a similar vein to the above distinction for cognitive styles. Table 4.4 clarifies this organization.

Styles based on:	Dimensions	Literature
Learning process	Concrete experience/reflective observation/abstract conceptualization/active experimentation	Kolb 1976
Orientation to study	Activist/theorist/pragmatist/reflector learners Meaning/reproducing/achieving orientations later developed into deep/surface/strategic	Honey & Mumford 1986, 1992 Entwistle 1979, Entwistle & Tait 1994
	Surface-deep-achieving and motivational orientation	Biggs 1978, 1985
	Synthesis/analysis, elaborative processing/study methods	Schmeck & al. 1977
Instructional preference	Environmental/sociological/emotional/physical/ psychological aspects	Price & al. 1976-77, Dunn & al. 1989
Cognitive skills development	Participant-avoidant, collaborative-competitive, independent-dependent Visualization/ verbal symbols/sounds/emotional feelings	Grasha & Riechmann 1975 Reinert 1976
	Cognitive skills/ perceptual responses/study and instructional preferences	Keefe & Monk 1986, Keefe 1989a,b, 1990
	3 types based on analytic-global continuum	Letteri 1980

Table 4.4. Models and features of learning styles. (adapted from Riding & Rayner 1998, p. 53)

The only potentially unclear distinction emerges from the last category (cog. Skills development) which has a clear overlap with the overall WA and VI dimensions, but the intended focus is on the *development* of skills rather than the categorization of an individual's styles. However there is no doubt that learning styles and learning strategies are very different from cognitive styles:

> "Individuals develop learning strategies to deal with the learning material which is not initially compatible with their cognitive style. Strategies can be learned and modified while style is a relatively fixed core characteristic of an individual. (…) Learning strategies are formed as part of a response within the individual to meet the demands of the environment. Learning strategies may thus be seen as cognitive tools which for the individual are particularly helpful for successfully completing a specific task. This approach leads to the concept of the strategic learner." (Riding & Rayner 1998, p. 79)

The result of this integration of cognitive styles under two broad dimensions was the Cognitive Styles Analysis (CSA), an instrument which measures styles on these two broad dimensions. The major impact of this work was the specific attention to the reliability of the test used to identify stable profiles. The CSA prompted a lot of research and Peterson (2002), in her doctoral research, clearly identified a number of flaws with the test and specifically with the validity and reliability of the instrument. Nevertheless, the mode of organising styles in such way was testable, simple, and provided a useful way of categorising cognitive styles. Riding (1997) argued that the construct validity of styles dimensions needs to meet 6 key criteria:

- Be objectively measured
- Be independent of one another
- Be separate from intelligence
- Be independent of personality
- Be related to observed behaviours
- Be related to physiological measures

Research using the CSA framework was conducted attempting to satisfy all the aspects listed and provided a simple paradigm to benchmark other measures. Contrary to other measures of styles, the CSA is a task-based rather than self-reported measure and we will consider the implications at a later stage.

4.3.4. Mental theory of self-government and the threefold model of intellect

Sternberg is one of the most cited authors in intelligence research: in the last chapter we have also seen that he is one of the most vocal about the shortcomings of theories of abilities and intelligence as good predictors of achievement. Acknowledging that abilities are not telling the full story about how well people do in life, Sternberg was very concerned about styles, especially with attention to education and learning. He associated the term styles with *thinking*, which puts his contribution in between the definitions of cognitive and learning styles as characterised in the summary table of definition (table 4.2). In his words:

> "a style is a preferred way of thinking. It is not an ability, but rather how we use the abilities we have. We do not have a style, but rather a profile of styles." Sternberg (1999, p. 19)

Working roughly in parallel to Curry and Riding, Sternberg and his collaborators (in chronological order: Wagner, Grigorenko and Zhang) progressively refined his conceptualization of styles based on the so-called theory of mental self-government (MSG).

To sum it up,

> “The basic idea of the theory of mental self-government is that the forms of government we have in the world are not coincidental. Rather, they are external reflections of what goes on in people’s mind. They represent alternative ways of organising our thinking.” (Sternberg, 1998, p. 19)

We will provide a more comprehensive description of the model in a later section when we discuss the Thinking styles questionnaire, but basically the MSG is founded on three *functions* of governments (legislative, executive and judicial), four *forms* of governments (monarchic, hierarchic, oligarchic and anarchic), which are further specified in *leanings* (liberal and conservative), *levels* (global and local) and *scopes* (internal or external).

It is very easy to notice that such organization is very different from the two types of organization presented by Curry or Riding and colleagues. In fact, Sternberg approaches and relates to the pre-existent literature after he detailed his model: in his view styles research can be categorised in *cognition-centred*, *personality-centred*, *activity-centred* and *teaching styles* (Sternberg 1998), but he justifies the usefulness of the government metaphor echoing some of the theme already mentioned in Curry and Riding. In particular he advocates the need for a unifying theory of styles that allows for a convergence of methods and measurements, but also allows differentiating styles from both abilities and personality and, more importantly proves to be *useful* and *compelling* for its validity (Sternberg 1998).

In line with other researchers, however, by 2001 Sternberg also clarified the difference in the terms used and started to make a clearer distinctions between the terms thinking, cognitive and learning styles and in Zhang & Sternberg (2005) the three terms were integrated in the term *intellectual styles*. In this new definition,

> “An intellectual style refers to one’s preferred way of processing information and dealing with tasks. To varying degrees, an intellectual style is cognitive, affective, physiological, psychological and sociological.” Zhang & Sternberg (2005, p. 2)

This definition is much broader than any previously used and shows how the original conceptualization evolved in the context of the progress in the styles research field. However, it might also look like the concept of styles has become an umbrella term.
In their latest works, Zhang & Sternberg produced yet another integrative model termed the “threefold model of intellectual styles”. This was created with the specific purpose of exploring three core issues unresolved by the previous integrative models. The controversies are about: 1) the trait vs. state nature of styles (i.e. are styles unchangeable in nature and

therefore more trait-like?), 2) the value-laden or value-free conceptualization of styles (i.e. are the opposites of a bi-polar dimension equally good to solve a particular problem?); and 3) the correspondence between constructs and tools (i.e. are different labels of styles really representing different constructs or overlapping dimensions?).

The threefold model, based on empirical evidence in studies conducted over the 1990s, categorises a number of different constructs into three types. Table 4.5 provides a simple characterisation of the mappings between the types and the constructs.

Table V. Intellectual Styles

	Style type	Type I	Type II	Type III
	[a] Learning approach	Deep	Surface	Achieving
	[b] Career personality type	Artistic	Conventional	Realistic, Investigative, Social, Enterprising
	[c] Mode of thinking	Holistic	Analytic	Integrative
	[d] Personality type	Intuitive, Perceiving	Sensing, Judging	Thinking, Feeling, Introversion, Extraversion
Style construct	[e] Mind style	Concrete random	Concrete sequential	Abstract random, Abstract sequential
	[f] Decision-making style	Innovation	Adaptation	
	[g] Conceptual tempo	Reflectivity	Impulsivity	
	[h] Structure of intellect	Divergent thinking	Convergent thinking	
	[i] Perceptual style	Field independent	Field dependent	
	[j] Thinking style	Legislative, Judicial, Global, Hierarchical, Judicial	Executive, Local, Conservative Monarchic,	Oligarchic, Anarchic, Internal, External

Note. Theoretical foundations: [a] Biggs's theory of student learning, [b] Holland's theory of career personality types, [c] Torrance's construct of brain dominance, [d] Jung's theory of personality types, [e] Gregorc's model of mind styles, [f] Kirton's model of decision-making styles, [g] Kagan's model of reflectivity-impulsivity conceptual tempo, [h] Guilford's model of structure of intellect, [i] Witkin's construct of field-dependence/independence, [j] Sternberg's theory of mental self-government.

Table 4.5. The organization of intellectual styles (from Zhang & Sternberg 2005, p. 38)

Methods of instruction	**Styles most compatible**
Lecture	Executive, hierarchical
Thought-based question	Judicial, legislative
Cooperative (group) learning	External
Problem solving of given problems	Executive
Projects	Legislative
Small groups: students answering factual questions	External, executive
Small groups: students discussing ideas	External, judicial
Reading	Internal, hierarchical

Table 4.6. Thinking styles and methods of instructions (adapted from Sternberg 1998)

As in the other reviews previously mentioned, the methods for selection and inclusion of styles were rather idiosyncratic, exemplified in the following excerpt:

> "First, the models selected are among those commonly considered to be influential in the styles literature. Second, the styles constructs defined in the models are operationalised and thus empirically based. Finally, the style construct defined in a model has been tested against at least one other style construct." (Zhang & Sternberg 2005, p 19)

Another major contribution from Sternberg's MSG was his specific attention to instruction and the methods of delivery and assessment. These are formalised in the table 4.6, and its intuitive mapping of instructional tasks and styles resembles the pedagogical mapping reviewed in chapter 2, providing a practical relevance for the application of styles in education and instructional design.

4.3.5. The 'Coffield review': families of styles.

From the fairly brief and coarse account of the styles literature given so far, it is evident that there are numerous problems, not only with the terminology used (i.e. Desmedt and Valcke 2004, Peterson et al. 2009), but also regarding the conceptualization of styles and the issues about validity raised by almost all authors who attempted an integration or review of styles (i.e. Curry 1983, Miller 1987, Riding & Cheema 1991, Grigorenko & Sternberg, 1998; Sternberg & Grigorenko, 1997, Zhang & Sternberg 2001, 2005).
Coffield et al. (2004) intended to provoke researchers in the field of styles in their critique:

> "The field of learning styles suffers from almost fatal flaws of theoretical incoherence and conceptual confusion; for example, you can read about left-brainers versus right-brainers, pragmatists versus theorists, and globalists versus analysts. We collected thirty such pairings – the logo for the learning styles movement should be Dichotomies R Us. There is no agreed technical vocabulary and after thirty years of research, there is no consensus." (Coffield et al. 2004, p 18)

However, the points raised by Zhang & Sternberg and noted in the last section, already provide a deep reflection on the problems. One of the key features, also pointed out in all reviews cited so far, as well as by Desmedt and Valcke (2004), was the way in which styles were selected by the respective authors. This was not systematic, but rather *selective* to say the least (most say that they referred to 'influential' models). Because each author uses independent and unrelated methods to justify the eligibility of measures, the results obtained are often very difficult to integrate. The lack of systematicity prompted Desmedt & Valcke to use an alternative organization, which focused on the scientific value of the research used, and capitalised on citation rates as an objective index. This method was useful to provide a visual representation of the research and lend some support to Riding & Rayner's organization of styles into two families of cognitive and learning styles as well as a number of smaller clusters around influential researchers, which provided an empirical way of evaluating 'popular' measures.

The popularity of styles demonstrates the interest and potential utility of these constructs in applied settings (both classroom and workplace) and according to Cheminais (2002) and Reid (2005), learning styles and psychometric assessment (Rayner, 2007) should be an essential part of research in education which promotes diversity and inclusion. Hall & Moseley (2005) argued very effectively that

> "There is enormous intuitive appeal in the idea that teachers and course designers should pay closer attention to students' learning styles: by diagnosing them, by encouraging students to reflect on them and by designing teaching and learning interventions around them. The shift of focus on the learner, rather than on the subject matter which may (or may not) be learned would, it is argued, have a considerable motivational effect both on students, who feel valued, and on teachers, who feel that they are engaging directly with learners' needs rather than delivering a prescribed curriculum." (Hall & Moseley 2005, p. 248)

In the UK the views supported by the Labour government were instrumental in introducing policies that shifted the focus from the class to the individual needs in teaching. In 2002 the government, via the Learning and Skills Development Agency commissioned a comprehensive and systematic review of the theory and practice of learning styles (the 'Coffield review' hereafter).

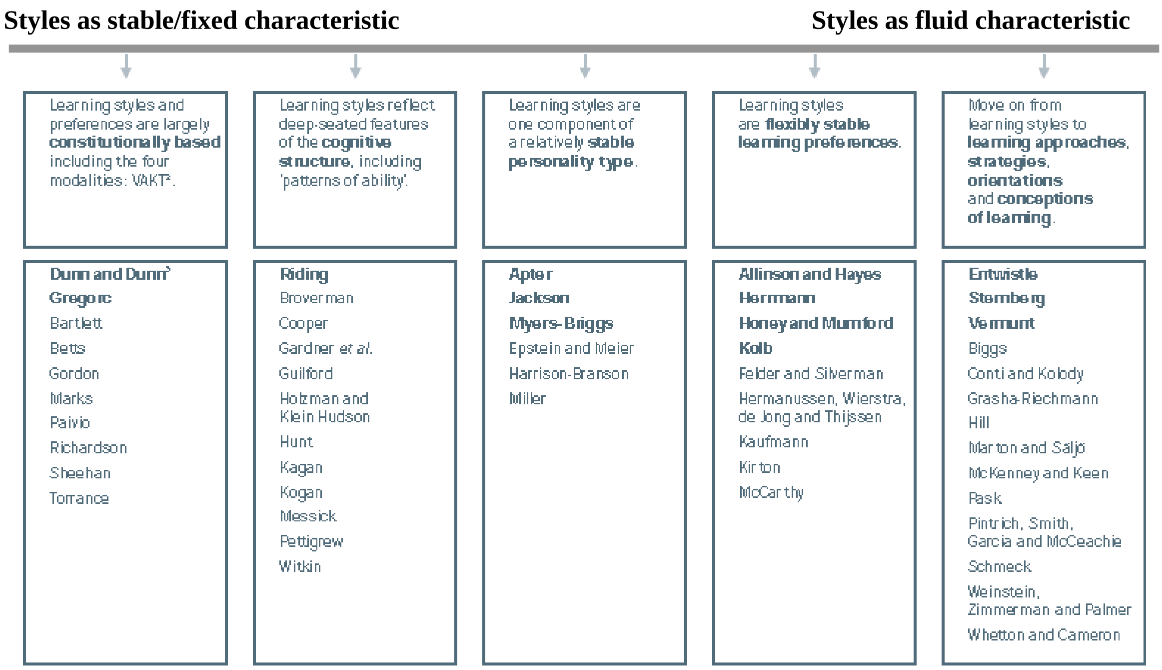


Figure 4.4. The Families of Learning Styles (from Coffield et al. 2004, p.20)

The project produced the first systematic review of its kind in the field. As well as proposing an integrative model organised around ‘families’ of styles (figure 4.4), one of the most controversial points of the review was the selection of 13 “potentially influential models of learning styles” which the authors selected from a list of 71 models published in the literature between 1902-2002 (Coffield et al. 2004, Hall & Moseley 2005).

The families of learning styles are aligned on a continuum of stability in which at one end styles are considered as either stable/fixed characteristic of individuals or, at the other end, as showing fluid characteristics. Figure 4.4 provides an overview of the whole organization. The publication of the review created a ripple effect in the learning styles community. Even if scholars were invited during the review process to defend their positions, the outcomes were received in quite different ways by different researchers.

The common denominator was an increased awareness of the shortcomings of different constructs. Nevertheless, a generalised resentment was felt and most notable was the response by Rayner (2007) who put forward the views of the members of the European Learning Styles Network (ELSIN). In response to the Coffield review, Rayner argued that:

> “Much of this conceptual infra-structure is unattributed, deals in secondary sources, reflects a shaky basis for an interpretation of psychometric judgements, largely emulates the structure but not the process of Curry’s review of learning styles completed two decades earlier (Curry, 1987), and is summarily justified by a declaration that leading academics in field were given an opportunity to defend their work. The sharpest criticism in this review, however, is linked to a concern that psychometric assessment should serve no or little place in education. Allied to this is the criticism of a divided field that has produced a tradition of contested research evidence regarding the predictive validity, reliability, and utility of various tools for psychological assessment.” (Rayner 2007, pp 25-26)

It is evident from both Coffield and Rayner’s comment that the strident tone and carefully worded criticisms animated the academic debate.

On one hand the ‘Coffield review’ attempted to provide an unbiased overview of the research available on styles: its success is arguable, but their effort did create some doubts as to the credibility of a number of styles constructs and provoke scholars to defend their positions. On the other hand, it became the primary reference for practitioners and instructors alike, offering a clear practical guide to make sense of the big number of constructs. Even with its flaws, to date, the number of citations in Google Scholar for the review and subsequent reports (Coffield et al., 2004a, 2004b; Hall & Moseley, 2005) is over 250[6], with most of the citing works specifically addressing teaching, pedagogy or e-learning, which demonstrates the practical relevance of this review and the potential bias which Rayner strongly criticised.

4.3.6. A 'learning styles' perspective

As we have seen from the various integrative models, the focus on the classification of cognitive styles often obfuscates the definition or inclusion of the concept of learning styles. In chapter 2 we covered a number of issues about learning and its relations with instruction: in particular we talked about experiential learning and the resulting model of approaches to learning with the three basic strategies (surface, deep and strategic). At the start of this chapter we also mentioned Ramsden 3P model. These examples are all framed into an educational research context in which abilities and personality play a lesser role than previous experience and instruction.

Entwistle and colleagues as well as Biggs demonstrated how students apply different strategies in different contexts, with different tasks and in different subjects. This locates learning strategies (or approaches to learning) at a specific extreme of the models cited so far and clearly not overlapping with cognitive styles.
One model, which strongly influenced the development of measures of learning styles, is Kolb's model of experiential learning. Here, learning is intended as a continuous process that is strongly associated with specific operations occurring in the brain. The most recognizable visual representation is the one in Zull (2002) reproduced above (figure 4.5), which attempts to localise functions with a specific brain topology.

In Kolb's model, learning styles are measured using the Learning Styles Inventory (LSI) which provides an insight into one's preferred styles based on two dimensions (abstract/concrete and activity/reflection).

---

[6] To put this into context, Desmedt & Valcke (2004) indicated 340 citations for Witkin and 172 for Kolb over a span of 50 years

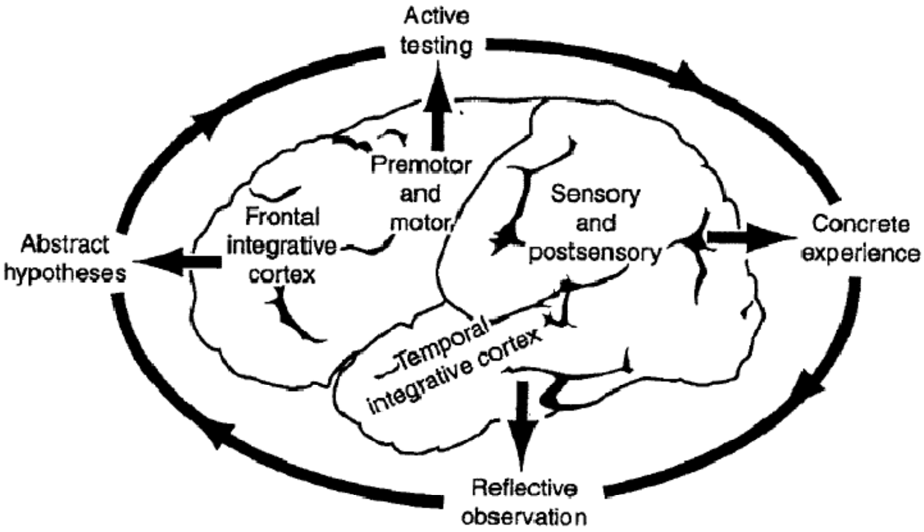


Figure 4.5. Cycle of learning from Zull 2002

This structure is similar to the Honey-Mumford Learning styles questionnaire, but very different from the conceptualization of approaches to learning put forward by Entwistle et al. It is important to note that as for Approaches to learning (Entwistle & Tait, 1990), *styles* should not be interpreted as trait or as fixed categorizations, which is also the stance taken by Kolb.

> "When it is used in the simple, straightforward, and open way intended, the LSI usually provides an interesting self-examination and discussion that recognizes the uniqueness, complexity and variability in individual approaches to learning. The danger lies in the reification of learning styles into fixed traits, such that learning styles become stereotypes used to pigeonhole individuals and their behavior." (Kolb, 1984, pp. 290-291)

The purpose of this specification is quite simple and was hinted at in the introduction of this chapter where we provided a definition of the terms and has the practical implication for teaching methods already highlighted in chapter 2.

Whenever a measure of learning styles is used in a practical intervention, being e-learning or another pedagogical adjustment, understanding the context remains essential to demonstrate the utility of the measure of style in order to enhance the interaction between the student and the material or the instructional method. In learning technology, it should be clear that the usefulness of styles is *descriptive* rather than *prescriptive,* and the metrics should be used to offer alternative ways of interacting with instruction, issues which is emerged in the last section about personalization and flexible delivery as well as inclusion.

4.4. A novel method for organizing the literature on styles

From the review papers we have examined so far, it is evident that idiosyncrasy and partiality are two core features of the various integrative attempts. As we have seen earlier Coffield et al. attempted a *systematic* review but were strongly criticised.
Only Desmedt and Valcke made an effort to undertake an impartial review based on objective indicators. However, their search was limited to the Social Sciences citation index and the result individuated only 1091 records from 1972 onward. The citation index reflects only a subset of the published literature, which is basically grounded on popularity and the cross- (or co-) citation only increases the bias of popularity of certain references/authors. Furthermore, the citation index used by Desmedt and Valcke limits the analysis to authors names, reducing the analysis to a pre-set dual categorization between cognitive and learning styles which can only re-confirm this broad distinction. Their effort is useful to position and categorise some authors as important within the conceptual spheres, but as we noted earlier, such distinction is not so straightforward and it is necessary to reconsider the narrow differentiation between terms, especially if the terms are contextualised in their domains of application.

Following a similar method, but exploiting the wider accessibility of search engines and databases available to date, in the next section we try to provide a broader and more systematic perspective of the literature concerned with styles, instruction and e-learning. This is based on the semantic and topical organization of the sources and modern visualization techniques. The aim is to provide a more impartial selection of the sources and replicable results. Even though we will go into more details about the overarching methodology of the thesis in chapter 6, the intention is to introduce *data mining*, *clustering* and information visualization (InfoVis) techniques as an application to a familiar problem; similar methods will be extensively applied to the dataset gathered for this thesis, therefore the exercise of organising the literature will be useful to highlight the basic procedures. Hand's words on data mining, in a paper published ten years ago, when the area of research was just starting off, are quite insightful:

> "Data mining is the science of finding unexpected, valuable, or interesting structures in large data sets. It is an interdisciplinary activity, taking ideas and methods from statistics, machine learning, database technology, and other areas." (Hand, 2000, p. 442)

It is feasible to comment that whilst statistics attempts to reduce the complexity of data using pre-specified models (with their assumptions and constraints), data mining is focusing on the algorithms (methods and techniques) to individuate interesting patterns in the data. Because of the exploratory nature of data mining, this technique is more suited to uncover interesting and useful patterns in the literature database than providing a model, and, as we will see in later chapters, to explore the relations between behaviours and individual differences.

4.4.1. Methods for the literature search

At the core of any data mining procedure is the preparation of a database in a format that the computer software can read and process. The first step was to identify and collate all possible papers available from the body of published literature related to the following keywords: “cognitive styles”, “learning styles”, “thinking styles” and “e-learning”. The latter was associated with a Boolean *OR* operator with “instruction”, “instructional technology” “education” and “personalization”.

The search was performed using five major databases for all available dates[7]: PsychInfo on OVID, ISI Web of knowledge, PubMed, Elsevier ScienceDirect, and the IEEE Xplorer. Data was collected either using a direct export of full records from the web interface of the respective search engines or fetched directly using EndNote™ X2. All records were collated into a single EndNote database.

The search generated a total of 45108 entries: results were automatically pruned of duplicates reducing the dataset to 43691 entries. Then, the database was cleaned to remove any further non-relevant record. ‘Non-relevant’ means that all articles with a connection to clinical and/or medical conditions in the samples used by each study were pruned. The following keywords (in the combination of title, abstract and keywords fields) were used to exclude papers as non-relevant: clinical, therapy, psychotherapy, preclinical, prenatal, menopause, cancer, autistic, autism, asperger, tumour, biochemical, chemical, crystal, endoscopic, alcohol, depress*, anorex*, bulim*, eating, disease, toxic, schizophre*, pathology, amino

[7] Ranges vary depending on the availability of journals online from 1698 to 2009, however only 25 entries dated before 1900.

acid, anaesthesia, ADHD, blind, deaf, dislex*, counsel*, patient, health[8]. Finally all entries with incomplete data (i.e. author(s), date or abstract) were removed from the database.

database source	no of records
IEEE Xplorer	605
ISI	8713
PsychInfo	13590
PubMed	2219
ScienceDirect	902
total	**26029**

Table 4.7. The distribution of references from the various databases after the pruning process.

At this point 26029 records were considered suitable from the database, but a further manual pruning of duplicates was necessary to remove entries, which has been missed by the automatic procedure due to small typographical differences in the records. The final dataset contained 23388 usable records, which is the largest number of entries in any review or meta-analysis performed on cognitive and learning styles to date.

4.4.2. Using clustering techniques to mine the literature

To organise the large number of records, the dataset was processed using OmniViz™ (Pospisil, Iyer, Adelstein, & Kassis, 2006; White, Cruz, Cameron, & Drabinski, 2008). This is a software tool written in Java which is normally used to support the semantic analysis and visualization of large datasets in chemistry and genetics. The software uses a range of clustering algorithms based on the semantic similarity (topicality) of list of keywords built from predefined fields. For this task, the list of keywords was created from the combination of the title, keywords and abstract of each entry. The software performs a cluster analysis with a pre-selected algorithm on the key terms listed in the pre-processing stage and identifies commonalities between records based on deviations.

Data can be plotted using a matrix-style heat map (which represents the correlations between clusters and key terms), a theme map which also adds the strength (significance or p value of the relations) and a galaxy-style configuration (which is a topicality map of the clusters).

Clustering was performed using a *k-means algorithm*: this aims to partition *n* observations (each reference) into *k* clusters (organised by major terms) in which each observation belongs

[8] The asterisk indicates all variations of the terms.

to the cluster with the nearest mean (similarity index based on keywords). The optimal number of clusters is automatically defined by the system to provide a suitable distribution of clusters based on the sample size and which also differentiates between different sources, however the *adequate* number can be manipulated iteratively according to an in-depth review of the clusters content and keywords.

An initial solution was generated with 152 clusters, (see Figure 4.6), but by inspecting the content of the clusters (Figure 4.6, middle) it is easy to see overlaps between the preliminary groups. An alternative solution was forced with 22 clusters: the choice was made based on the high correlations emerging from the heat map (red areas) which provides a rough estimate of the points of interests. The new solution was then evaluated and key terms promoted or demoted according to relevance: for example terms such as “nurse” or “energy” which are featured in a fairly high number of entries were demoted to the list of minor terms, but terms such as “CSI” or “approaches to learning” were promoted as more important as well as associated with their synonyms and spelling variations (see figure 4.7 for an example of the evaluation of two different solutions). After any change in the terms clustering is re-computed and data plotted again. The resulting solution (Figure 4.8) is a satisfactory and informative visualization of the literature on styles.

The core difference between the integrative reviews examined earlier and this approach is in the scope, breadth and depth of the analysis. In fact, the visualization of the literature allows us to put into evidence some intuitive patterns as well as unexpected ones. For example, in Fig 4.7 and 4.8 it is possible to pinpoint very accurately the correlations and the strength of the relations between clusters and the major terms in the pre-processed vocabulary. From the galaxy plots (proximity of terms) it is easy to confirm the overall intuitive split between cognitive and learning styles as well as the practical association of learning skills and strategies with education, e-learning and learning technology.

This, however, does not imply that the distinction is so clear-cut: in fact the visual inspection of the galaxy plot (Figure 4.9) and the respective frequencies shows that the use of the various terms is widely spread and the distribution of the conceptual labels defined in tables 4.1 and 4.2 is confirming the divide emerged from the Delphi study by Peterson et al. (2009). The concept of “cognitive style” seems to be more easily found in papers which examine aspects of cognition, abilities, perception, academic performance and specifically “individual differences" rather than "personality". The term "learning style" is more easily found associated with education, student, personality and e-learning.

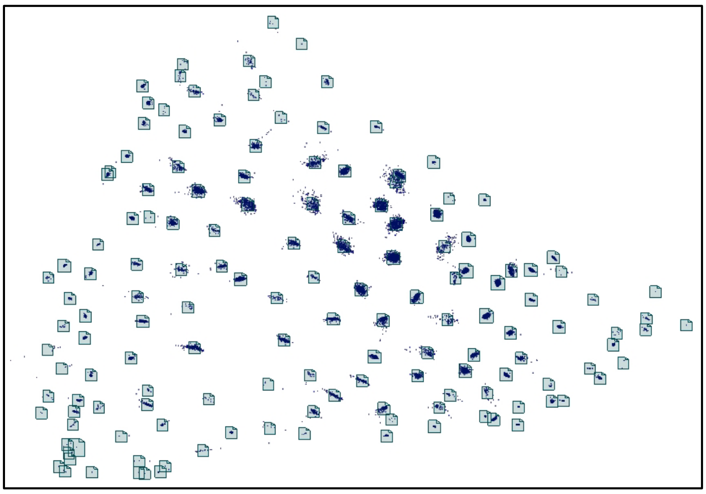
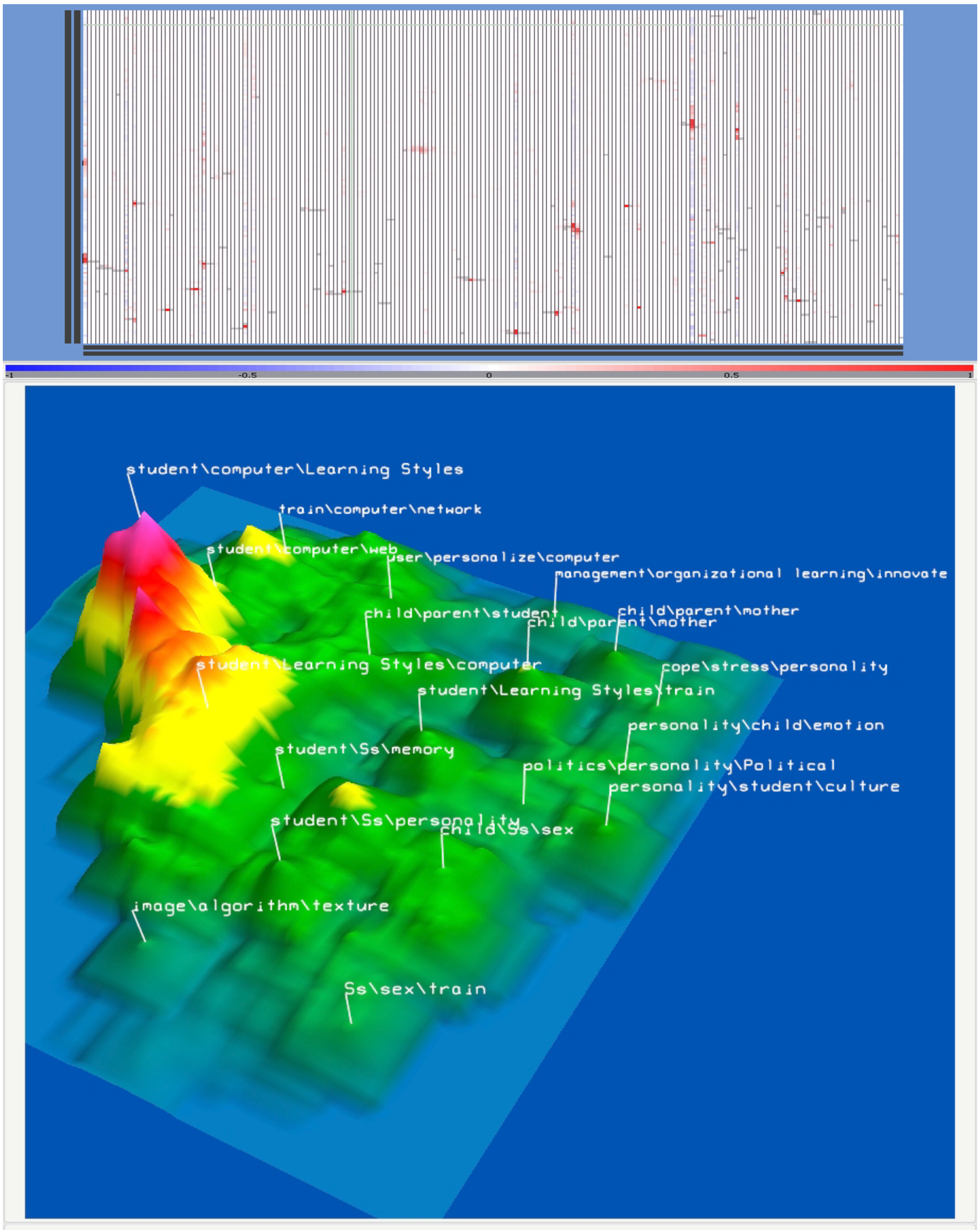


Figure 4.6. Three modes to visualise an initial clustering solution with 152 groups.

On the left the topic distributions (similar topics are closer to each other). Below, in the heat map of the key terms (X axis) and clusters (Y axis). In red are the highest correlations which allow to identify about 25 interesting 'spots' used to refine the clustering. At the bottom a 3d representation of the galaxy with the strength of key terms (topicality) represented as peaks.



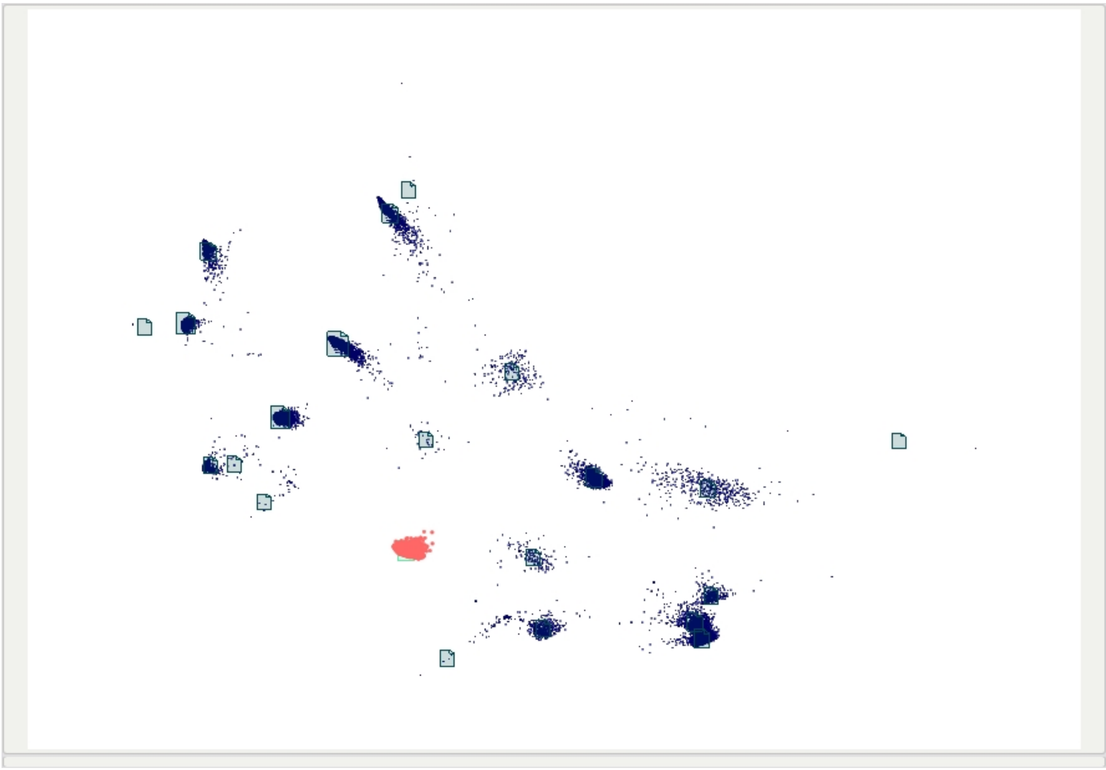
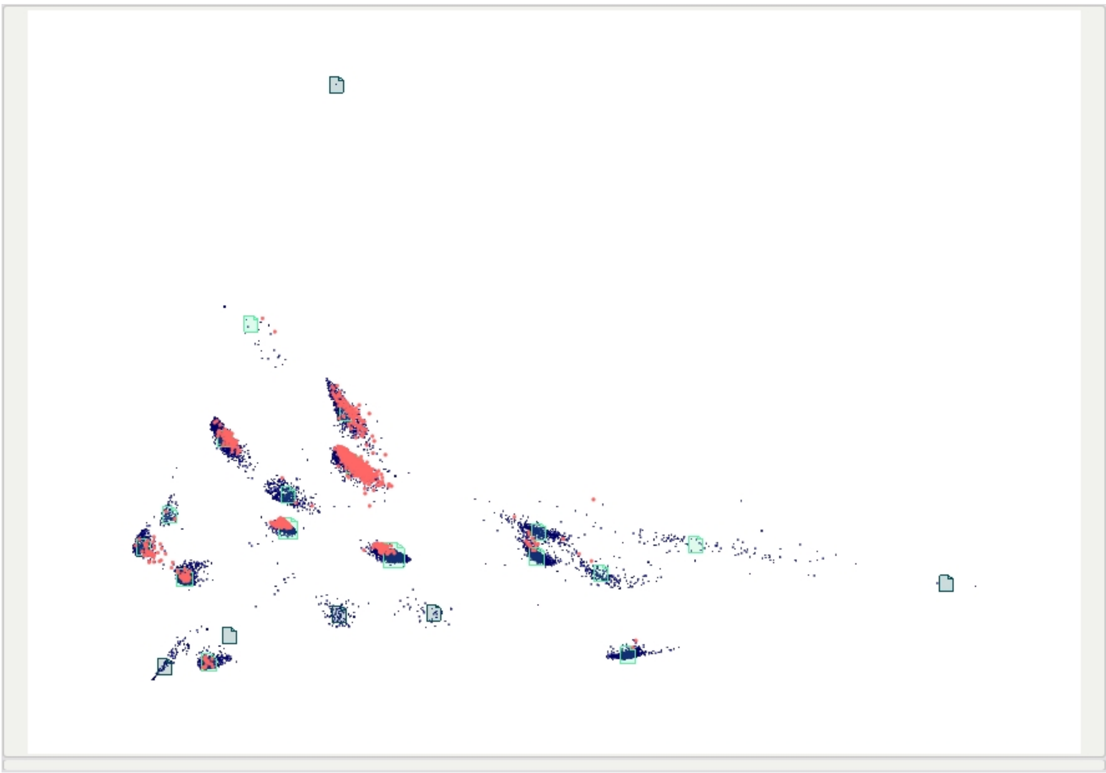


Figure 4.7. A direct comparison for two 22-clusters solutions after revisiting the importance of key terms: at the top the final solution (also characterised in the next page)

The middle group is highlighted in orange in a fairly compact cluster. At the bottom, the corresponding records in a previous iteration in which the terms "student", "learn" and "educate" are separated across a number of clusters (also in orange)



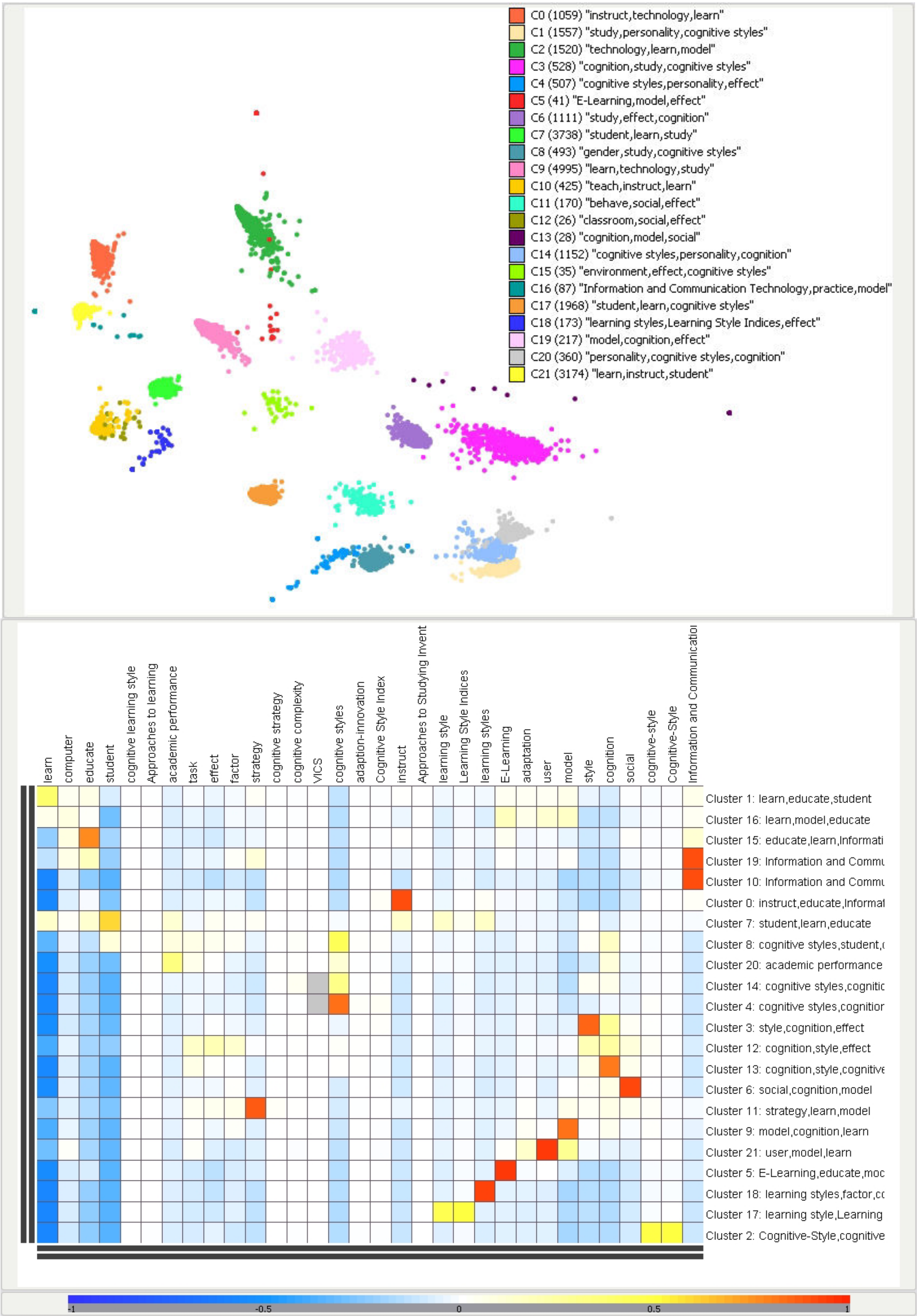


Figure 4.8. A more detailed overview of the final solution with a 22-clusters solution. Topicality and heat map with related key terms is shown in both graphs

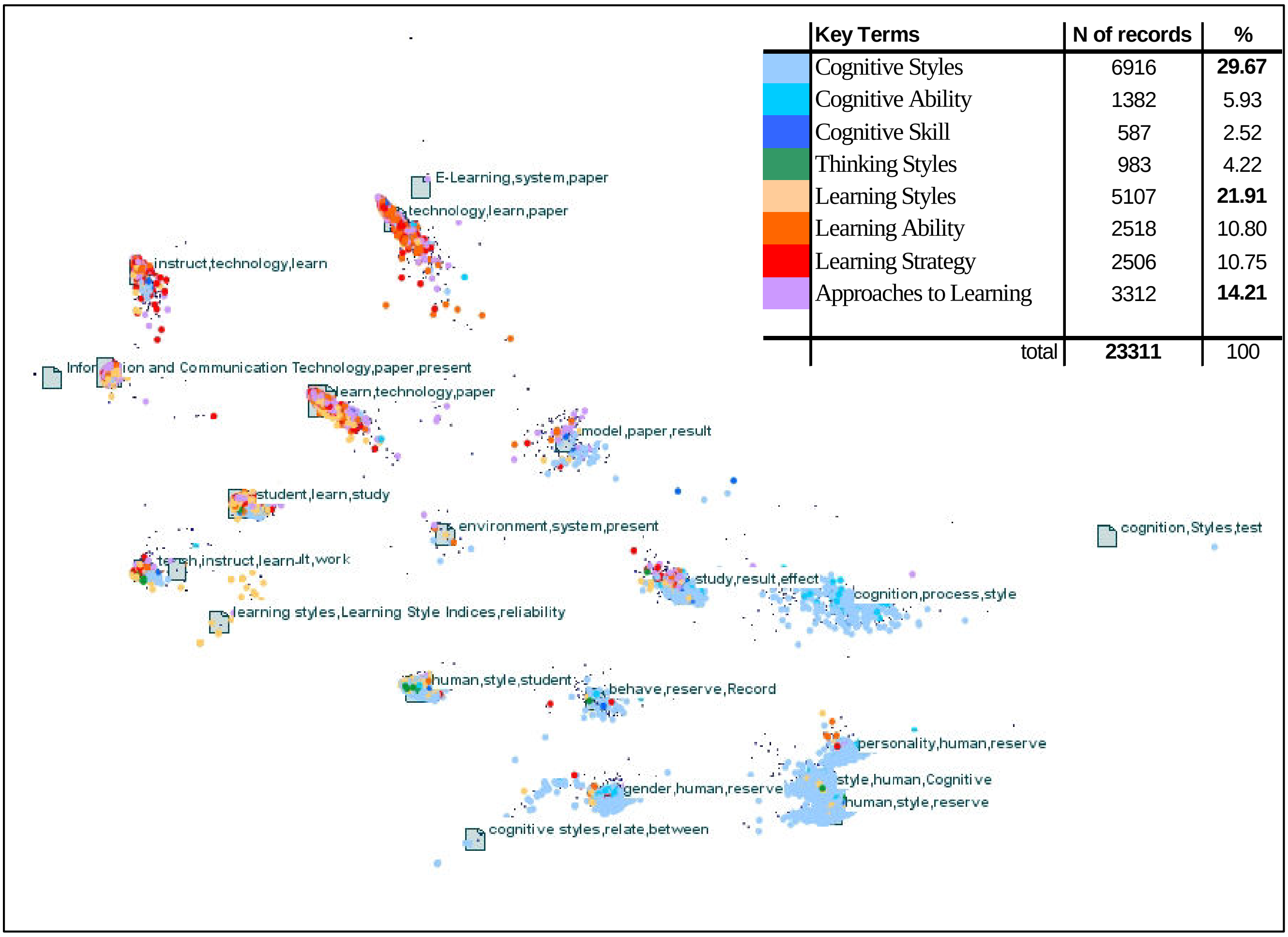


Figure 4.9. Frequencies and topicality distribution of the key style labels summarised in Table 4.2 and their associated clusters (with major terms for each cluster). The pages icons show the centroids of the clusters.

The core purpose justifying the work done to prepare this database is the flexibility afforded by the ability to query terms, concepts and topics, which is quite difficult to render statically on paper.

In particular, the database was used to identify more effectively all existing published work on individual differences and e-learning (1460 papers in the database), on data-mining related to styles (140 papers) and uncover patterns of relations, (i.e. highlight the fact that most papers talking about e-learning are associated with learning strategies or learning styles). The visualization of the galaxies of individual authors (Figure 4.10) also allows to identify similarities and overlapping topics between authors which in Desmedt & Valcke's representation look far apart, but in reality explored conceptually similar grounds.

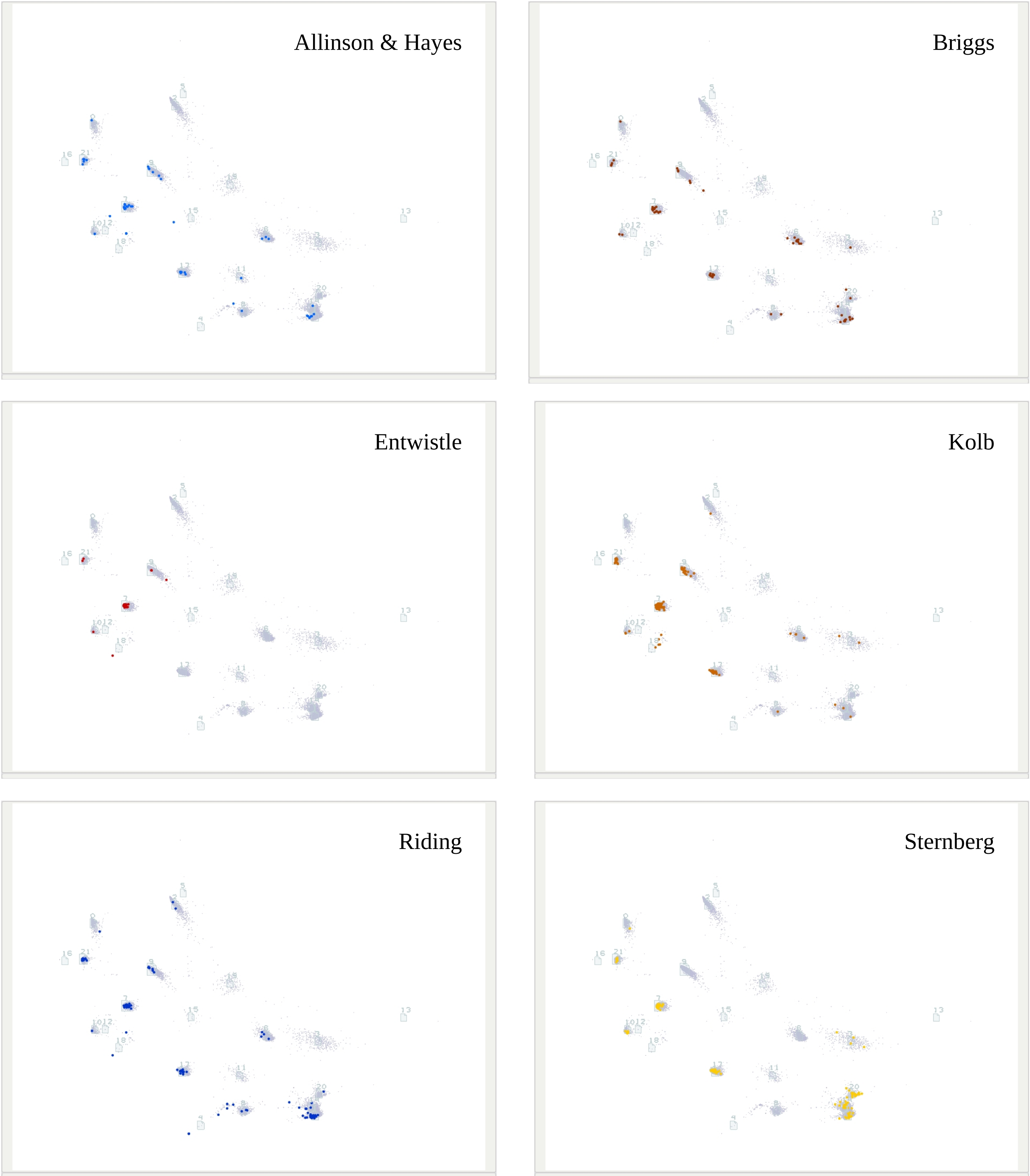


Figure 4.10. Distribution of the publications associated with the various authors/theories. Use the matrix at page 135, fig 4.8 to individuate topics related to each cluster.

Note that in this page we also included reference to other popular models which will be examined in more details in the following chapters

4.5. The utility of styles for the practitioner

There are three core aspects that summarise the utility of styles in the context of this thesis and provide a clearer foundation to direct the research agenda: 1) the relation between ability and task complexity, 2) the relation between the strength of styles and 3) their match with instructional/teaching methods and the complex interaction between ability, personality and styles.

The relations between the first two are exemplified in the graphs in figure 4.11which specify in more detail the directionality of the relations between the paths suggested by Furnham at the start of this chapter. The importance of styles becomes apparent in cases in which lower abilities are characterising the individual student and the tasks are more demanding and when specific patterns of styles are clashing with the instructional methods, making it more difficult for the students to grasp the content.

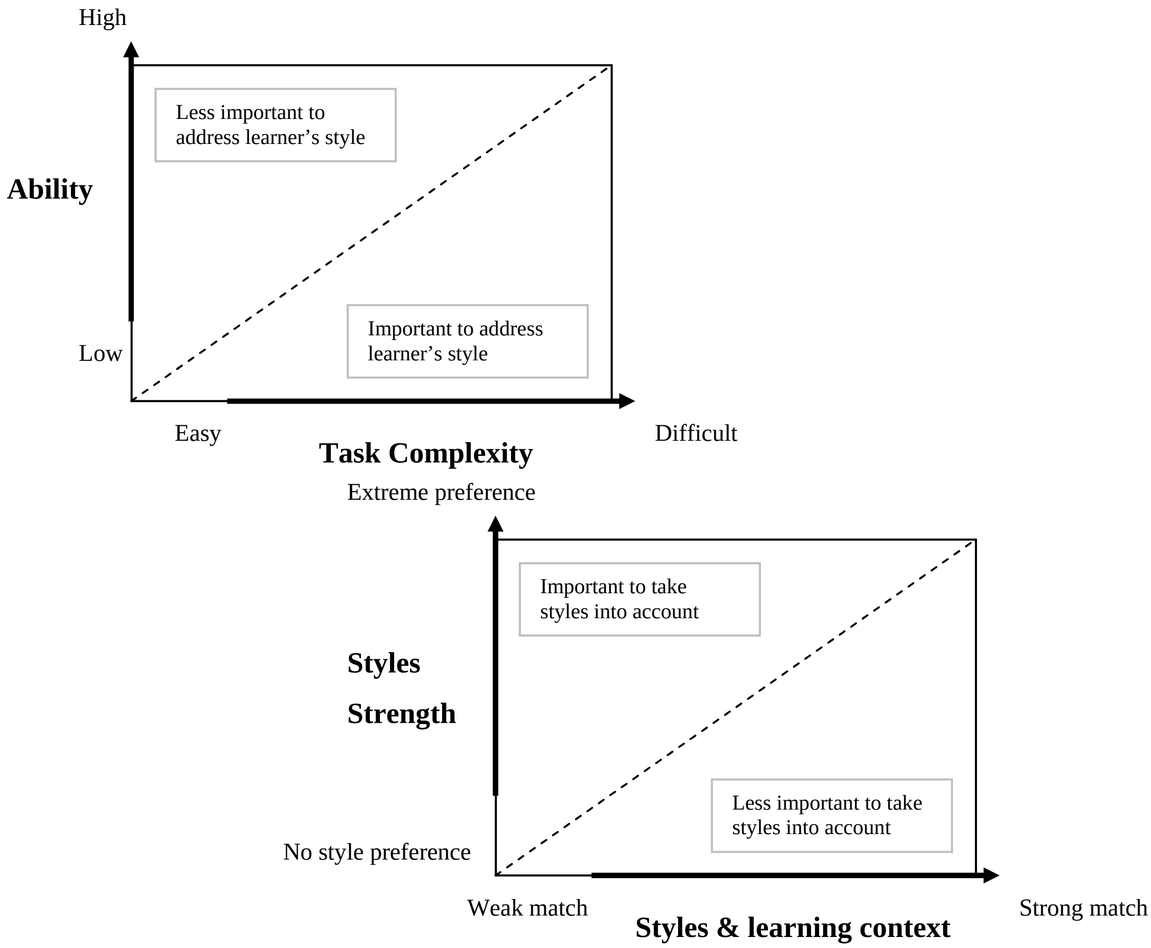


Figure 4.11. Possible relations between Ability and task complexity and styles strengths with learning context.

Top**, The importance of styles in the relation between ability and task complexity, Adapted from Riding & Rayner 1998.** Bottom**, The importance of styles and their match with the learning context, adapted from (Rosenfeld & Rosenfeld, 2008).**

The exact nature of the third relation could be explored only in a hypothetical experimental lab in which the other two variables could be controlled or, as we will try to do in the next few chapters, by mining available data to create explanatory models of the interrelations occurring in the samples available.

In fact, from the discussion so far, two aspects remain still open: the strength of the relation between ability and styles and the one between personality and styles. The core hypothesis regarding the first aspect is expected to be similar to the relation between IQ and personality in the last chapter. People with greater ability will probably display a greater range of styles as much as they will be equipped with a greater range of strategies to cope with life.

For the second aspect the relation is more complicated and because personality has not been investigated directly in this thesis, we will refer to it as the ‘personality-styles black box’.

For example, it is difficult to predict the direction of the relation between extroversion and the deep approach to learning. Intuitively, one might suggest that an individual with high extroversion may prefer social interaction to solitary study time; the individual has less time to spend on studying and the adoption of a surface approach is more likely. However, if this individual is socialising and revising with a group of bright people, their depth and breadth of understanding might well surpass any time spent in isolated studying. By using some proxy measures, such as punctuality in submitting coursework and attendance as proxies for Conscientiousness, we will reconsider this relation at a later stage.

4.6. Problems with styles measurements and setting a research agenda

After revising the definitions and theory of cognitive and learning styles it is important to bear in mind a number of intrinsic problems with the use of styles measures to assess individual differences:

- There is no consensual or coherent overarching theory with potential overlap between some measures; it should be noted that overall there is a disregard of effect size (and statistical precision) in comparing published results;

- validity and reliability of measures are inconsistent, the effect is a psychometric weakness in models and measures of styles;
- no clearly established evidence (but growing in e-learning as we will see in the next section) in reported research of positive effects related to the application and practice of learning styles;
- “commercial conflicts of interest under-mining reliability in the confirmation of proof for empirical research and in some extreme cases this is described as a messianic drive for field domination” (Rayner 2007, p 25): this is problematic as theories and explanations of the learning process are not providing a clear return for resource investment in the educational setting.

To overcome some of these problems and issues Cools (Cools, Evans, & Redmond, 2009) suggested a matrix merging the fundamental issues in Riding & Curry: this was meant to provide a systematic framework to develop a research agenda driving the styles community (Figure 4.12).

Curry Riding	**Theory**	**Measurement**	**Practical relevance**
Identity fundamental dimensions	**Knowledge networks of cognitive styles scholars**	**Research with various cognitive styles measures simultaneously**	**Attention for knowledge dissemination to practitioners**
Develop simple, valid & direct measures	**Scrutinise existing cognitive styles measure**	**Develop an overarching instrument**	**Research in environmental settings**
Situate cognitive styles	**Develop overarching theory**	**Longitudinal designs**	**Provide relevant, contextualised and concrete practical implications**
Link with other observable behaviour	**Interdisciplinary research teams**	**Multi-method and multi-source approach**	**Joint networks of scholars and practitioners**

Figure 4.12. A matrix to summarise issues in styles research from Cools (2008, ELSIN presentation)

As we will detail in the following chapters, the research conducted in this thesis places itself well in the middle of this landscape and covers research question in a number of these intersecting boxes. In particular, with regard to the theory, the thesis will look at the validity and relations of a number of existing measures of styles, tapping directly in the issues about measurement: it sets out to use a longitudinal design, using multiple sources of data and specifically multiple measures of styles, but, more importantly, the research is trying to be ecologically valid, collecting data from a real instructional setting, rather than an experimental setting, maximising the potential applicability of its findings.

4.7. Individual characteristics and learning technology

As indicated in the mining of the literature, learning technology in education introduces another layer of complexity in the interaction between instruction, styles and academic performance. In fact, because of the nature of the medium and the potential capability of e-learning, it should be easier to produce educational material which allows students to use the tools offered and the material to suit their styles best, and there are some cases in which such implementations had seemed to be successful. Two issues emerge with this statement: students' metacognition of their abilities and preferences is normally very limited, therefore it is unclear how such a better awareness of styles could help them direct their activity; in fact, as we illustrated from Miller's work, pushing students to become more self-aware might have a negative effect. Automated systems, instead, could actively use the information to lead students, but again they might be a hindrance due to the limited capabilities. Even though machine learning has made incredible advances in the last 30 years, automated systems are successful only in contexts in which the problem is well defined or the path for learning is predictable in some way.
The interaction with a human learner still remains a very complex affair and the combination of variables involved too unpredictable to fit in specific patterns.

From the last chapter and some of the issues touched in this chapter, we have seen that individual characteristics are closely related to the ability to learn and students' propensity to take learning opportunities when these are offered. We have also identified a number of these characteristics as very important from the literature on teaching and instruction: intellectual ability, personality, styles, work commitment and motivation are all related to a successful and effective learning experience. In a similar way we can draw from the literature in e-learning that there are characteristics which are also relevant and useful in the application and use of learning technology. One fundamental fact is that even though most people conduct their lives immersed in the technology, certain individual characteristics make some people use and work with technology more than others. Literature in human-computer interaction and human factors as well as in usability research, started to pay more attention to the user characteristics which are influencing interaction with technology. Witkin’s field dependence has been quite popular because its obvious link with visuo-spatial abilities (Dufresne & Turcotte, 1997; McDonald & Stevenson, 1998), but already in the 1980s the relation between information processing strategies attracted the attention of engineers (Robertson, 1985). We ought to consider in more detail how the literature is organised to determine which other features are important and have been demonstrated to be useful.

In the literature explored on e-learning and individual differences, it is possible to identify three main strands of articles: e-learning to improve or enhance instruction, student-centred learning and student-centred personalization.

One of the first aims of e-learning has always been to find way to improve (or make more efficient) the way in which instruction is delivered. In HE there is a pressure to make teaching more efficient to allow staff to dedicate time to research and attract funding. In the industry training must be effective and efficient with the maximum possible return for the cost. Training (or re-training) of employee is one of the most important expenditures in the running costs of a company, therefore it is natural to see e-learning as a natural tool to improve the process.

If one shifts the attention to the student, as we’ve seen in the chapter 2, it is important to adjust the methods of instruction to the students. Taking a student-centred approach to instruction, there are two distinct aspects, based on the different philosophies of instruction taken: on one hand the instructor has the responsibility to provide a wide enough spectrum of option to allow each student to perform and her best. On the other hand, it is the responsibility of the student to find which tool suit them best and take the opportunities offered by the instructors to personalise their learning.
These are partially compatible with one another, but the stress put on each aspect emerges very fragmented from the current literature depending if the main author is a lecturer, and educator, a system developer, an e-learning expert or an administrative person with institutional priorities.

Another dimension is related to four types of implementation strategies: static (offline evaluation), iterative (mixed, instructor led), flexible (mixed, system-controlled) and dynamic (online evaluation, adaptive system).
A static strategy derives straight from traditional educational evaluations: an instructor makes a change to a course and then evaluates the outcomes at the end of the year with the intention of revising the course for the following year (Toohey, 1999). A variety of evaluation methods and data sources could be used ranging from the simple evaluation of the final marks to course feedback forms and, in case of e-learning, summary data from the usage of the system. The system is usually not changed and remains stable in a quasi-experimental condition.

An iterative strategy is similar to the static in the sense that data sources are evaluated offline, but the system is modified ad-hoc based on a continuous feedback loop similar to the one characterising software development cycles.

A flexible system is also a hybrid of offline evaluation methods, but it is different from the iterative because a pre-planning phase establish the goals and suitable paths to achieve learning. It could be modified iteratively, but the e-learning system works independently, partially directing students' activity based on pre-define heuristics which are driven by specific variables. Examples of such systems are interactive tutoring systems.
A dynamic strategy is the one that relies more than any other on a student profile (and it is arguable which variables do count) and evaluation is done online based on preset heuristics.

It is easy to see where styles fit in and become essential for each strategy from the characterization of students, which allows grouping in the sample to detailed and rich profiles in the case of adaptive systems.
In the next section we review some examples of the practical use of style measures in the implementation of e-learning systems.

4.7.1. Individual characteristics and the relevance of styles for the evaluation of learning technology

As noted in chapters 1 and 2, there is a widespread belief that e-learning could improve both delivery and effectiveness of instruction, especially in higher education, in which mass-education is becoming more impersonal, to better prepare students for self-directed and lifelong learning. In most cases presented in education, the results advocate positive effects in promoting learners’ engagement to a varying degree of success (Lea, Stephenson, & Troy, 2003; Naidu, 2003a, 2003b; O'Neill, Singh, & O'Donoghue, 2004). However there are many that strongly endorse the 'no significant effect' of the use of technology in support of learning (Russell 2001) and others, after the hype surrounding e-learning initiatives are now reconsidering the issues of augmentation vs. disruption (Grainne Conole et al., 2008), Heilesen & Josephsen, 2008).

Upton & Adams (2006) argued that whilst there seems to be some evidence that *in general* learning technology is beneficial, not enough work has been done to explore which students benefit most. This statement should be revisited as a great deal of work done in computer science and engineering remains largely unknown to psychologists and educationalist. Already back in 1979, Zmud classified personal characteristics affecting interaction with technology into three groups: demographic (age, sex and training), personality-based (attitude, motivation), and related to cognitive style (the way a person analyzes and evaluates data). He found a relation between these variables and technology use and it is no surprise that 30 years later these categories remain fundamental and a lot of progress has been made.

Partially because of the predominantly visual nature of computer displays, one of the most prolific areas of investigation is the one exploring the *cognitive processes* involved in the interaction with web-based material. Rouet (1996) dedicated an entire book to the research on hypertext and cognition, and more recently, Chen & Macredie (Chen & Macredie, 2004, 2005) reviewed a number of research articles investigating cognitive styles and hypermedia navigation. Makkonen (1998, 2003), Shohreh & Garland (2000) and Jonassen & Reeves (1996) point out that hypertext/hypermedia can be considered as a facilitator when learners build their own knowledge. Kozma (1991) noted a parallel between hypertext technology and mental models that form associations (links) among various ideas by constructing meaning from these relations. Rouet et al. (1996) argued that hypertext can be considered a powerful learning tool, as there is an analogy between the structure of hypertext and the structure of contents in human memory[9]. Similarly, Kommers and colleagues (Fernandez, Kommers, & Asensio, 2004; Kommers & Lanzing, 1998) praised the ability of hypertext/hypermedia network structure to facilitate access to various documents without constraining the user.

---

[9] There is a resemblance to the concept of *memex* introduced by Bush in 1945 in his article on the Atlantic.

Others (Sheard 2002, Graff 2005), however, indicate that the non-linear structure of the web might easily cause disorientation, a issue already explained by Pirolli & Card (Pirolli, 2007; Pirolli & Card, 1999) in the early days of the internet. More recently, using a more user-centred focus, Graff (2003, 2005, 2006) conducted lab-based research which showed how cognitive styles related to browsing strategies as well as website structure. Cook (2005) reviewing the relations between web-based learning and cognitive and learning styles, also argued that styles should be carefully considered by educators when designing web-based learning, not just to address usability issues. In his review of the medical and health professions trainings, he identified clear expected patterns between the wholistic-analytic construct with aptitude-treatment interactions.

Other factors such as personal commitment and effort are also very important when learning through technology, as there may be constraints that cause negative learning, especially considering the original sources investigating distance-learning (Alderman & Fletcher, 2005; Hiltz, 1994; Hiltz & Turoff, 2002). The learner's commitment to study must be greater when learning through technology than in a traditional environment (Alavi, Yoo, & Vogel, 1997). The issue of self-pacing and complete control over one's own experience is flagged as both an advantage and a disadvantage: in fact the lack of self-directedness of some students can lead to a greater dissatisfaction with the learning experience. There is no doubt that the learner's attitude and motivation are vital for a positive virtual learning outcome (Hiltz, 1994; Williams, 2003), even if Cashion & Palmieri (2002, 2009) are more of the opinion that "the secret is the teacher". A learner's learning process can be a determining factor in the effectiveness of technology-based programs (Kizzier & Pollar, 1992). Alavi & Leidner (2001) maintain that technology can efficiently support a learner's cognitive activities, such as information selection, coding and understanding. However, these are all individual-oriented views: the nature of interaction online provides both a human-machine interaction, but also, and more exciting, human-to-human (both peers and instructors) dimension which can augment the learning experience.

Internet and its tools (such as discussion groups and electronic mail) may facilitate cooperative learning, allowing people to work and discuss together, thus developing their cognitive processes and constructing their own knowledge. Webster & Hackley (Webster & Hackley, 1997; Webster, 2001) argue that interaction is a critical factor when learning takes place through technology. When there is no such interaction, participants can easily get distracted and focus on other activities, which may result in negative learning. On the other hand, it is important to foster interaction among participants, so that the classroom can be perceived as a virtual learning community where an exchange of experiences takes place, rather than as a private session between tutor and learner (Almeda & Rose, 2000).

4.7.2. Individual characteristics and adaptivity

The success of e-learning, and teaching in a blended environment or with automated systems is essentially dependent on suitable models of the learners (de Koning, Bredeweg, Breuker, & Wielinga, 2000), or representations of the users/learners which can direct or enhance the interaction with the system and which is implicitly formed by human teachers and/or negotiated within the student/teacher interaction.

Without the mediation of a human teacher, the concept of adaptivity afforded by learning technology is the most appealing aspect for customization and personalization. Brusilowsky (2001) recently reviewed adaptive hypermedia as an area of research at the crossroad between hypermedia and user-modelling and showed a number of successful implementations. Since the middle of the 90s Brusilowsky regularly reviewed the progress in the field and introduced order in understanding adaptivity in hypermedia applications (Brusilovsky, 1996, 2001; Brusilovsky & Millán, 2007). One question, which recurs in these reviews, concerns what a system should be adapting to. Kobsa (Kobsa, 2001; Kobsa, Koenemann, & Pohl, 2001), in response, suggested three core aspects: user data, usage data and environment data. Introducing operational definitions here will ensure conceptual clarity: customization is based on a static definition of a model which can –in principle- provide an individualised experience for the learner/user. This however, is limited compared to personalization, which is based on the system's ability to produce a dynamic adaptation which ultimately emulates real instructors (Naidu, 2003). Before 1996, only user characteristics, such as goals/tasks, knowledge and preferences were typically implemented. However, with the boom of the internet between 1997 and 2000 publications about adaptive systems increased exponentially demonstrating the need to produce better adaptive experiences to suit the more demanding users. (Mobasher et al., 2002).

Adaptivity, however, does not necessary mean a better experience; the underlying expectation that the system can serve the right information at the right time for a user is deceptively simple. So far, user models have been built using two techniques: user-provided information (e.g. questionnaires or declaration of preferences) and interpretation of user behaviour. However, the user data aspect is usually limited. Even if the system could always provide the correct information, this would implicitly remove the learner from the constructive process of learning in which the interaction with the instructor is dynamically shaped by the learner's personal needs as well as the pedagogical needs. For example ignoring the learners' interests in selecting contents would stifle curiosity and drastically limit the possibility of exploration and discovery, which often characterise the effortless activity of learning. The contrast between using a system to aid instruction and the use of a system as an external validation of achievement is evident here: maximising the potential of the learner is not the same as ensuring proficiency with the subject matter.

In parallel with the idea of adaptivity and in contrast to Bruner's notion that there is an "effective sequence of the material" (Olson & Bruner, 1974), Laurillard (2002) suggested that there are many different pathways to learning; "the complexity of coming to know" as it has been termed, poses interesting questions and it is legitimate to speculate whether it follows from this that there are different yet equally valid pathways in the way that learning technologies are used by students. If this is the case, uncovering what part is played by students' characteristics (i.e. Bruner's attention to the learner's predisposition) and how much is constrained by the typology of the system is an essential question, which goes beyond usability issues.

In case of a VLE, which are now commonplace in HE, the systems are usually serving static content, but a certain level of personalization is afforded in the way material is accessed. More importantly, how students behave in the system can be monitored in an unobtrusive way providing very useful information about how online resources are used. Ramsden (1992) considered integrated monitoring (or tracking) of students' activities online an essential component of the "minimum standards" required for online learning provision. Hardy & colleagues (Hardy, Bates, Antonioletti, & Seed, 2005; Hardy et al., 2006) explored students' behaviour patterns using online resources in blended courses. This research revealed common patterns between students: for example first year psychology and physics students used online material in similar ways: cramming preparation during the couple of weeks before the exams, which showed in a massive increase in usage during this time. However, to overcome the limitations of a simple description it is necessary to make specific hypotheses about the relations between relevant variables. Relating usage and its effects on academic performance is a good start, but we believe that it is necessary to explore further the student model emerging from richer user data to draw meaningful evaluations.

For example, to measure the effectiveness of web-based material, user modelling provided valuable insights for customization. The early research conducted by Pirolli & Card (1998) which identified very specific behaviours when using online or web-based material has evolved in modern Web usage Mining (WUM) widely used in e-commerce, but rarely applied to education. The core aspect of this research is that web site structure and usability were the focus, and the effectiveness of websites was evaluated based on usage data (Berendt, Mobasher, Spiliopoulou, & Wiltshire, 2001; Mobasher et al., 2002). Thanks to the rapid advances in machine learning, however, the behavioural data collected can now be used to produce predictive models of users' preferences and styles of interaction for personalization purposes.

Romero and Ventura (2007) went as far as stating that the application of knowledge extraction techniques, not only provides rich information about the learners, but could be viewed as a formative evaluation technique for the educational system.
In practice these student models are based on a combination of cognitive styles heuristics such as Witkin's field dependence or Kolb's learning styles, demonstrating in practice how system interaction can be designed with relatively simple rules of engagement. For example Uruchuturu (Uruchrutu, MacKinnon, & Rist, 2005)attempted to customize the interface and sequence presentation of content in a web-based system according to the wholistic-analytic and verbal-imagery dimension to provide a better experience for students. Triantafillou et al. (Triantafillou, Pomportsis, & Demetriadis, 2003) used Witkins's construct (field dependence/independence) to drive the adaptivity of their educational system. Sheard et al. (Markham et al., 2003; Sheard & Markham, 2005) also used styles (Kolb) to partially inform the automation of the presentation of material and learning progression in their implementation of a VLE. Results were promising in these studies showing that concordant presentation, preference and styles provided better post-instruction results, but even though such examples are very useful, there is no attempt to provide a coherent framework to implement widely their design choices.

		ASSIST	VICS-WA	MSG	CSI
	Type of measure	Preference	Preference (RTs)	Preference	Preference
	Type of instrument	Self-reported questionnaire	Computerised task	Self-reported questionnaire	Self-reported questionnaire
	Degreee of ext validity	Some	Much	Some	Debated
	Reliability (test-retest)	Some	Some	Much	Some/stable (factor structure issues)
	Amount of research	Some	Much	Much	Some
	Biological base	Unknown	Some	Some	Some
	Other styles	Partial	Yes	Debated	Debated
independent of	**Intelligence**	Unknown	No	Debated	Some
	Personality	Partial	Yes	Debated	No
	Curry	Outer layer	Inner layer	Middle layer	unsure
	Cognitive perspective	concept formation	perceptual	Program	across 2-4
Relations with other models	**Riding**	wholistic/analytic		wholistic/analytic	wholistic/analytic
	MSG	complex (i.e. type1=deep)	complex (i.e. analytic=type2)		complex
	Coffield	Learning approaches, strategies…	Cognitive structure	Learning approaches, strategies…	Flexibly stable learning preferences
	Kolb				

Table 4.8. Basic features and relations between the measures of styles selected for this research.

4.8. Picking the right measures

As we've seen in the last section there is no easy recipe to justify the selection of a specific measure of individual differences: very few attempted to systematically investigate a unitary framework using styles, and nobody, to our knowledge, implemented a system taking into account the richness of *different* styles measures in association with e-learning. This is no surprise given that in chapter 3, one of the key issues emerging from the literature in differential psychology was the difficulty of keeping profiles rich enough to allow one to make practical design or instructional decisions, but also narrow enough to avoid overcomplicating the characterisation.

In the various iteration of research of this project we selected four measures of styles: Approaches to learning (Entwistle), Thinking styles (Sternberg, TSI and MSG are used interchangeably hereafter), VICS-WA (Riding, Peterson) and CSI (Allinson & Hayes).

These will be detailed in chapter 7, however table 4.8 provides some key features supporting the selection and reference to table 4.2 is useful to clarify the terminology.

The selection of appropriate styles measures, was informed by five basic criteria:

- Well known, measures of styles with a strong theoretical background;
- Metrics which have been used in the literature in relation to both AP or e-learning;
- Styles spanning across the spectrum (from proximal to distal measures and from relatively stable to malleable)
- Styles tested using different methods (i.e. self-reported inventory to cognitive-like experiments)
- Metrics providing a 'profile' report which could be interesting for psychology students to find out more about themselves.

From the theoretical perspective, if we relate the features of the four instruments listed above with the integrative models explored in this chapter, the choices should be more obvious.

Starting from the critical review by Coffield et al., which in some respect is similar to Curry's, this allows the *placement* of these four models on the *continuum of flexibility* (Fig. 4.4, p. 125) with ASSIST as the most malleable and changeable and the VICS-WA as the most fixed and unchangeable. Interestingly the selection of these particular tools was independent from what Coffield et al. elected as the 'most influential' models. The CSI and

MSG are roughly in the middle of the continuum, with the CSI leaning toward a more fixed side due Sadler-Smith's suggestion that there might be neural bases for the analytic style, and the MSG also roughly positioned in the right half of the continuum.

If we use Sternberg's triarchic model, this is the one that offers some insight in the relations between intelligence, personality and styles and the positioning of the various measure in the three core types of intellectual styles.

Moving to the more cognitive-oriented approaches, we have reviewed Miller and Nosal models which provided testable ways of looking at how the instruments selected are relating to each other and cognitive abilities. The fact that we included models which rely on both cognitive-type tasks (i.e. the VICS-WA) and self-reported inventories is not fortuitous and should provide a wider spectrum for evaluating the tools used as well as their relations with other sources of data.

By applying a data mining approach to the literature database, we can also position the models in the proximity map presented in Figure 4.10. This representation provides a clear spatial arrangement of each tool/theory, but also identifies a number of typical issues and keywords associated with each model. The representation helps to support Desmedt & Valcke's overall organization, with a clear differentiation between cognitive and learning styles, but the finer granularity of the plot enables us to see the *topicality* of each instrument associated with the topic clouds. The comments already made in this chapter and the figures at page 31-32, are self explanatory.

The four instruments proposed here provide a balanced approach to identify distinct measures of styles which are different in nature, but can be easily placed within the theoretical frameworks reviewed.

4.9. Chapter summary

In this chapter we have covered a lot of ground. From the general definitions of styles, we examined a number of integrative approaches to make sense of the vast array of available measures, tools and interpretations. A novel method of exploring and organising the literature was proposed which uses data mining techniques. By applying the method to a relatively simple problem we demonstrated the benefits of the technique to uncover patterns otherwise unavailable or difficult to grasp.

The relevance of individual differences for learning technology was also explored and we identified three core applications: customization of the learning experience, personalization of the learning paths and enhancement of learning aided by computers. Some applications were also presented in which styles were successfully used to automate or facilitate interaction.

Finally, the case for selecting four particular measures of styles has been presented and justified, contextualising the choices both in the theoretical background and the practical experiences.

This chapter provides the foundations to understand the selection and application of individual characteristics to this research projects and hints toward a systematic research agenda for both styles research and application of the theory of individual differences to e-learning. In the next few chapters the data collected will allow to test the hypotheses which will be put forward in the next chapter.

Chapter 5. The setting, data sources, methods of investigation and tools used

In the first four chapters, we attempted to define the nature of the complex patchwork behind the understanding of individual differences in student learning, teaching and the use of e-learning. Each field of research is quite vast on its own, but stressing the importance of student/user profiling seemed to be promising to appraise the variation in students' performance at university with the ultimate goal of improving the quality of both learning and instruction.

In this chapter we start with a brief presentation of the context in which the research was conducted, including the system used and providing relevant details of the courses.

Then, revisiting the literature reviewed, and with a clear picture of the research setting, we proposed some key research questions and hypotheses driving the experimental work. Because the data characterising individual profiles is at the heart of all the approaches considered, we explain in detail the overarching methodology adopted in collecting and processing data. A number of tools were used during the various iterations of the project to collect information providing a rich characterisation of the student/user.
The varied nature of the sources makes it necessary to detail the *techniques* used to collect and manipulate data as well as the *types* of data gathered to make sense of the large amount of data collected: both are essential to understand the limitations, the possible levels of abstraction and to interpret the findings presented in the following chapters.

These measures can be classified into four categories:

- academic performance (AP),
- self-reported psychometric indicators (measures of styles)
- online activity (observational data of online behaviours),
- survey/feedback opinions (end of year surveys).

Although we strived to obtain consistent, relevant and ecologically valid data, which are some strengths of the data gathered, it is important to note that the incremental nature of this longitudinal study led to one of the core limitations of the project: this is the difficulty of obtaining *complete* datasets. In considering the limitations of this research, we give a detailed overview of the samples used and the processing applied to each subset.

5.1. Putting the research in context: the setting

To better contextualise the research conducted, it is necessary to characterise the courses in which e-learning has been used and explain how systematic changes were introduced (both from the point of view of the course organization and system changes). In contrast to experimental settings, in which fictitious tasks are created to test the effectiveness of presentation of web-based material, typical in many experiments in human-computer interaction, students enrolled in the psychology courses had the opportunity to access the material in their own time suiting their individual way of learning.

Students were simply recommended to *explore* the material provided online throughout the courses, but they made an explicit choice of using what was on offer in an ecologically valid setting. Their activity online produced a stream of behaviours, which can be analysed and interpreted. Their usage supplied us with an experimental lab with a large database from which individual and aggregate characterizations were extracted.

5.1.1. Framing e-learning within the University

Back in 2003, the Principal announced the creation of an e-learning fund intended to support the integration of e-learning in all the University’s activities. About £1.2m was invested during in the following three years to fund original projects. An e-learning steering group was also formed and a number of key aims for the e-learning funding were agreed:

> "- Achieving widespread, appropriate use of e-learning in all Schools and at all degree levels as a normal part of learning and teaching
> - University of Edinburgh has an international reputation ('centre of excellence') for institution-wide implementation of high-quality e-learning
> - University of Edinburgh has an international reputation for leading edge developments in learning and teaching through the use of e-learning ('centre of innovation') "(AA.VV. 2003)

This ambitious agenda was complemented with a commitment for innovation and since 2003 a Virtual learning environment (VLE) platform was made available for all courses. As shown in table 5.1, the initial uptake was low, with only a small group of early adopters taking on the challenge of experimenting with the medium. Two years later the number of courses actively using the VLE was still small, but a drastic increase was registered in the academic year 2005-06 partially due to the full integration of the VLE with the University portal system (MyEd), and partially because of the big improvement with the upgrade to WebCT Vista 6.

	2003	2004	2005	2006	2007	2008
CHSS				**1420**	**3051**	**3781**
Business School				160	382	436
School of Arts, Culture and Environment				143	298	410
School of Divinity				102	210	237
School of Economics				42	57	74
School of Health in Social Science				48	91	152
School of History, Classics and Archaeology				179	404	459
School of Law				7	32	27
School of Literatures, Languages and Cultures				312	683	800
School of Philosophy, Psychology and Language Sciences				154	361	436
School of Social and Political Studies				185	391	535
The Moray House School of Education				88	142	215
CSEE				**550**	**1242**	**1441**
School of Biological Sciences				170	303	326
School of Chemistry				16	48	88
School of Engineering and Electronics				158	333	358
School of GeoSciences				160	416	435
School of Informatics				9	14	18
School of Mathematics				2	24	73
School of Physics and Astronomy				33	100	151
CMVM				**196**	**369**	**457**
Medical Teaching Organisation				33	71	122
Royal (Dick) School of Veterinary Studies				5	20	43
School of Biomedical Sciences				149	266	282
School of Clinical Sciences and Community Health				1	1	1
School of Molecular and Clinical Medicine				8	11	9
Grand Total	**640**	**650**	**1016**	**2166**	**4662**	**5679**
Percent WebCT active courses	16.41	17.18	23.92	46.69	94.01	108.42

Table 5.1. The growth of WebCT courses over time.

Note that since 2006 a small number of 'meta-courses' became available explaining the 108% for 2008. Meta-courses are WebCT courses not present in the student enrolment system which are used to organise content. For example the course Physics 1a and 1b are two separate semester courses with different course codes. In WebCT a third meta-course could be created to 'join' the two entities. Because of the way WebCT was deployed it is not possible to discriminate these entities.

In table 5.2 it is possible to see the corresponding increase in the users' uptake of the VLE at institutional level (top panel) and for the psychology courses used as source for online usage in this thesis. It is evident that in the case of the psychology courses most students accessed the online material whilst the corresponding average for the other courses was much lower (albeit rapidly increasing over time).

Many more courses took advantage of the facilities available to enhance teaching ranging from simply posting lectures on the web to more complicated interactive training courses or custom-made VLEs (this is the example of the School of Medicine which organised their own independent IT and e-learning support system).

Acad. Year	Tot enrolled students	tot WebCT users	WebCT active students	Percent WebCT active users	WebCT active staff	Percent of active users
2003	23205	n/a	5719	24.65	n/a	n/a
2004	23125	n/a	7313	31.62	n/a	n/a
2005	23715	n/a	12272	51.75	n/a	n/a
2006	24225	17801	16262	67.13	1539	8.65
2007	23555	25140	21348	90.63	3792	15.08
2008	24525	27828	23946	97.64	3882	13.95

Acad. Year	Students on the PSY1 Course	Percent PSY1 WebCT active	Students on the PSY2 Course	Percent PSY2 WebCT active
2003	320	n/a	142	n/a
2004	284	62.68	192	n/a
2005	293	96.59	208	93.75
2006	253	96.05	159	99.37
2007	287	100.00	139	87.05
2008	307	99.35	122	91.80

Table 5.2. The growth of active WebCT users (both students and staff) over the time span.

One of the key aspects moderating this enthusiasm was the fact that the learning curve for staff engaging with the new technology was underestimated, making it hard for staff to juggle the already busy schedule of teaching and research commitments and therefore limiting the amount of effort they could dedicate in this new enterprise. The level of support offered by Information Services was limited, relying on the *'ease of use'* of the system and expecting learning to happen in short and intensive training sessions. Even the most technically oriented staff faced a number of problems and the necessity of redirecting their time into the production of the online material didn't suit many. In this respect, new staff had the advantage over seasoned lecturers when working anew with the system and they did not have to re-work the existing teaching material used for a number of years in a different system.

This is possibly the biggest limitation of the uptake of new technology and innovation in teaching practices already highlighted in the literature (D'Silva & Reeder, 2005; Rogers, 2000)

5.1.2. The Psychology Courses

Both psychology courses in years 1 and 2 are fairly large, with an average of 250-300 students each. Taught in a research intensive department, ranking in the top ten of the last RAE assessments, this also means that teaching time for each staff member is shared across the full curriculum to allow them to be active in the research activities. Traditionally these two foundation courses have been taught by a course team of about 10-12 lecturers who deliver blocks of 6-7 lectures each in their own area of expertise.

The clear advantage is that the quality of research is also reflected in the teaching and that students have an opportunity to meet active researchers, and experts in their respective fields, early in their careers.

Y1	Y2
• Large class (about 300)	• Quite large class (up to 200)
• Diverse background and curricula	• Diverse background and curricula
• Few contact hours (3 hrs lecture per week & 1 tutorial every fortnight)	• Few contact hours (3 hrs lecture per week & 1 weekly 3-hrs practical)
• Sparse assessment (3 essays)	• Heavy practical load with stats, SPSS & research experimentation
• Large teaching team (~13 lecturers and ~18 tutors)	• Sparse assessment (4 experimental reports)
	• Large teaching team (~13 lecturers and ~12 tutors)

Table 5.3. Similarities and difference between the psychology 1 and psychology 2 courses.

It should be noted that 'few contact hours' is compared to other EU countries listed in the Bologna convention. This is a common standard in the UK, and the workload of the Y2 is closer and more typical of some science courses (in which practical labs are necessary).

As well as lectures, students also participate in research activities (and receive some credits for it in the first year) and produce their own research project as early as the second year, which is a unique feature of our course. Students take part in fortnightly tutorials in first year and weekly practical sessions in their second year, both of which are facilitated or directed by postgraduate students in the School. Table 5.3 is useful to summarise the differences and similarities between the two courses, including assessment details.

The downside of this approach is its fragmentation, which often emerges in students' feedback about these courses. The fact that students meet a new lecturer every couple of weeks makes their experience fresh and interesting, however it also encourages a certain compartmentalization of the course content and this is reflected especially in the relations between the lecture series and the tutorial/practical components running in parallel courses.

5.1.3. E-learning in psychology

Formal support of e-learning started in 2004-05 for the year 1 course. The specific aim of promoting the introduction of the VLE was to create a support system able to integrate the wealth of material made available to students, and create a sense of community, which was supposed to extend beyond the lecture theatre.
The core problem to be addressed was to create a supportive environment for the learning experience in a course which was mainly based on a one-to-many with a traditional face-to-face way of teaching.

The large number of students in a very heterogeneous cohort, did not help tailor teaching to the learner, nor accommodating possible differences in styles or learning strategies adopted by students. The addition of online material was intended to bridge the traditional methods of delivery with a more student-centred approach to teaching as well as facilitate the collation of material from a variety of sources into a single 'space'.

		2003/04	2004/05	2005/06	2006/07	2007/08	2008/09
	VLE details **IS integration**	WebCT CE4 external	WebCT CE4 partial data	WebCT CE4 MyEd	WebCT Vista 6 MyEd	WebCT Vista 6 MyEd	WebCT Vista 6 MyEd
Psychology 1	**other Tools**		epacks:research tasks, Thompson	epacks research tasks, Thompson, polls system, Turnitin	epacks: Thompson, Turnitin	epacks: Pearson. Turnitin	epacks: Pearson. Turnitin
	timeline						
Psychology 2	**other Tools**			Turnitin, Students projects intra-web	Wikimedia, UCCASS, Turnitin	Confluence wiki, Turnitin, UCCASS, LimeSurvey	Confluence wiki, Turnitin, LimeSurvey
	timeline						

Figure 5.1. The timeline of the e-learning integration with the courses with reference to the additional tools not included in the standard VLE. The grey line is indicating a partial support or a preparatory stage of development.

Provision of e-learning for the psychology 2 courses started in the 2005-06. This built on the previous year experience with psychology 1 students who had became accustomed to using the online resource and expected something similar in year 2. Moreover the type of instruction, highly based on weekly tutorial/lab provided different challenges for the implementation and delivery of course material.

The result of these two implementations was a blended learning environment in which traditional methods of teaching were enhanced by the delivery of online content, which became the unified point of reference for students and teachers alike. From feedback surveys since 2005, students consistently reported that they liked the system and happily embraced the innovation (Hardy et al., 2006; Vigentini & McGonigle, 2005).
Since 2003 I have been responsible for the task of maintaining and developing this resource whilst liaising with the course organiser to deliver an enhanced experience for students. In this position I was able to leverage on the system to explore how students interacted with the system and how individual differences affected their online behaviours.

5.1.4. Course features and pedagogical goals

Although the mode of delivery has remained constant over the last 6 years, small modifications were required in the forms of assessment due to the introduction of the two semesters' structure. In particular, back in 2004, two end-of-semester exams replaced a class exam for the lecture and one for the statistics in January and a formal registry exam in May. Another important change was the inclusion in 2006/07 of a statistic component in the second year exams, which was traditionally assessed as a separate class exam. This element of assessment together with a minimum B grade was necessary to get into Honours. Such requirements were lowered to a minimum C grade and a pass in the statistics component after this ceased to be an open book exam.

First year students were required to submit 3 essays during the year. In second year the requirements included the submission of 3 experimental reports and 1 students' led project report. The final mark for each course was computed including the coursework (and participation in year 1) and the two end-of-semester exams.

		skill area	Intellectual Skills:			Communication skills:			Organizational Skills		Interpersonal Skills		Research Skills			Numeracy & Statistics			Computer Literacy				
		skill tasks	critical and analytical thinking	synthesizing & integrating	problem-solving skills	scientific writing	summarising and presenting scientific data	oral communication talks & presentation	taking initiative & working independently	time-management (handling pressure and multiple deadlines)	working with or motivating others	flexibility/adaptability (not every one is the same in group work!)	Experimentation as participants and researcher	create own project as investigators (final projects)	develop own ideas and define project details	statistical skills	data handling	research methods in practice	word processing	fine-tune information search abilities (WWW & databases)	use Spreadsheet & stats software	engaging with web 2.0 tools (wiki)	develop online questionnaires (limesurvey)
learning tasks & activities	teaching & learning methods	lectures	V													V	V						
		tutorials	V			V					V					V	V						
		practicals	V		V		V				V	V	V			V	V	V	V	V	V	V	V
		independent work			V				V	V				V	V	V	V	V	V	V	V	V	V
		reading material	V			V				V				V	V		V			V		V	
		extra VLE support							V	V					V	V				V		V	
		group work			V	V	V	V	V	V	V	V			V	V	V	V					
	summative assessment methods	reports	V	V		V	V		V	V						V	V	V	V	V	V		
		exam essays	V	V		V	V																
		exam stats			V											V	V	V					
	teacher lead	research participation							V	V	V	V	V					V				V	V
		blitz test			V											V	V	V					
		wiki poster		V		V	V				V	V		V			V	V	V	V	V	V	
		presentations		V		V	V	V			V	V		V			V	V	V	V	V		
	two-ways	research design	V	V					V	V			V	V	V	V	V	V	V	V	V		
		wiki use			V				V		V	V		V	V			V	V	V	V	V	
		project work	V		V		V	V	V	V	V	V		V	V	V	V	V	V	V	V	V	V
	student driven	find relevant literature							V					V	V			V		V			
		evaluate literature	V											V	V			V	V				
		data analysis			V											V	V	V			V		
		data synthesis	V	V										V		V	V	V	V		V		
		reporting & presentation		V		V	V								V	V	V	V	V	V	V	V	
		conducting research	V		V				V	V	V	V	V	V			V		V	V	V	V	V
		stats practice			V				V							V	V				V		
		task management	V		V				V	V	V	V		V	V			V				V	
			formative assessment with feedback provided																				

Table 5.4. A summary of the relations between course goals, assessment methods and skills in the courses considered

The core pedagogical goals of the courses are summarised in table 5.4 in which we attempt to organise the learning activities promoted by the courses according to their function, and link them to the skills associated with the learning outcomes.

This is not intended to be exhaustive, but it should be helpful in relating instructional goals, assessment and students' activities in a single frame of reference. The table is also useful to show that the ICT tools were integral in the philosophy of teaching rather than standing out as simple adds-on to the traditional course teaching.

5.2. An overview of the VLE and tools used in the courses

In the latest version, the online material is organised in a similar way in both courses, but the type of items available and the format of teaching largely drives how students access it and the additional tools offered.

5.2.1. Psychology 1

The provision of e-learning offered students access to lecture, tutorial and self-test materials and a single focal point for important communications and course organisational notices. It provided a wide range of teaching and research material in support of essay preparation, augmented by MCQ and self-tests, including both "home grown" (research tasks) and commercially available packages (e-packs), which aided preparation for the formal examinations and provided an interactive component to the course. The commercial e-packs are associated with the main textbook.

5.2.2. Psychology 2

For administrative information and lecture resources the material is similar to that devised for Psychology1. However, the nature of face-to-face practical class instruction in Psychology2 is much more intense than the tutorial format in Psychology1, with significantly more interactive tutor-student instruction. Nevertheless the VLE provided a dynamic virtual space for delivering e-learning instruction, not just as a repository for material. Weekly tasks, evaluated (but not formally assessed), are a mean of extending student-tutor interaction beyond the classroom and, especially towards the end of the course, the interactive space is extended to the whole class as students are requested to design and run their own experiments. We used a wiki as a support tool, allowing students to share and present all the material produced: from the definition of a project proposal, to the delivery of a final presentation to exhibit their work and findings, the wiki was the solution (Cowan, Vigentini, & Jack, 2009; Vigentini, 2008).

All lecture material is normally released in a phased way as it becomes available from the lecturers. However, admin information, support readings, self-test and exercises are available from the start.

5.2.3. An overview of the content in WebCT

The core of the provision of e-learning in both psychology courses was made of: 1) lectures notes, 2) a readings repository and 3) tutorial/practical material.

Core readings were all provided online via a reading repository except for some widely available book chapters, which can be accessed in the library. The coordination with the Library system over the years greatly improved over the past 3 years, however we tended to keep the documents for the core reading in the system, allowing us to track them internally.

The tutorial and practical material is a blend of resources. For the psychology 1 course it mainly consists of essay writing guidelines, short guides on referencing, researching databases and additional tutorial readings. For the psychology 2 course a number of short guides on referencing, report writing and detailed information about the experiments are made available throughout the course. The practical handbook was also converted for the students to an online version, for reason of time this is simply a transposition of the text into web pages, but it would be desirable to adapt it in a more interactive format.

WebCT built-in tools: utility tools.

The announcement tool was very useful to post messages to the entire class or to specific sub-groups, creating redundancy in the communication system. This used in coordination with the calendar tool, which has been used to remind students about lectures and tutorial times, constitute what we later classify as *utilities*.

Discussion Board & chat rooms: social tools

Both tools have a potential to promote interaction between peers and encourage communications with teaching. Chat rooms were trialled in 2003, but limited usage and students' feedback prompted us to deactivate the tool from the courses.

In the new version of WebCT™ we activated the 'Who's online' tool, which allows students to view if anyone is currently logged in the system and send each other quick messages.
In the original version of the psychology 1 course we created a private discussion board for every tutorial group and a number of topics related to general course issues. Because the tutorial-size discussion topics were virtually unused, we opted for general class-size forums in following years. It should be noted that in 2005-06 we used an external discussion board system linked to WebCT™, which caused a biased representation for this tool in the overall distribution of activities for the 2005/06 courses.

e-packs

E-packs are a set of tools produced by the publishers to complement a specific textbook. Two different versions were used over the years. One type was associated with the Atkinson & Hildgard textbook (the course textbook in 2004-06, Introduction to psychology 3rd ed. - Thompson). The content was licensed to the Department and in our case the installation was set up as an extra WebCT course instance available to the students from their course selection menu. A second version was the self-contained proprietary implementation offered with the Martin & Carson textbook (used in 2006-08, Psychology - Pearson). In contrast to the previous system this was hosted on their proprietary VLE based on the Blackboard platform.

The hosting difference had two important implications: 1) students had to log in a different system to gain access to the resources, and 2) tracking and monitoring of activities were highly limited.

The advantage of the solution offered by Pearson was pedagogical: students could create individualised study plans, which are constructed after the student has taken pre-tests on the material. After completing their study of the online material and activities, students could take post-tests, which provided detailed feedback on their weaknesses.

Students' presentations & the Wiki system

As part of the assessment in year 2 students had to run their own research project in the second semester. A variety of support systems have been tried over time to maximise the effort and allow students to share ideas and work together. For the first time in 2006-07 a WIKI system was introduced to support psychology 2 final projects. The advantages of this system have been introduced elsewhere (Cowan et al. 2008), however it is important to note that like the external discussion board mentioned earlier, this meant that activity for the year

2 in the second semester mainly shifted outside the tracking of the VLE. As we will see, this has not affected the overall usage, but certainly caused a big gap (i.e. we have limited data) in the real online usage for these students in the external system

Self-assessment and stats exercises

Self-assessment in the form of multiple choice questions and activities has been present since the beginning. The largest bank of tests is available via the publisher's e-packs, however a number of self test are also available for the psychology 1 course in the statistics exercises section which was implemented in 2005. Both of these were used widely, certainly justifying the investment.

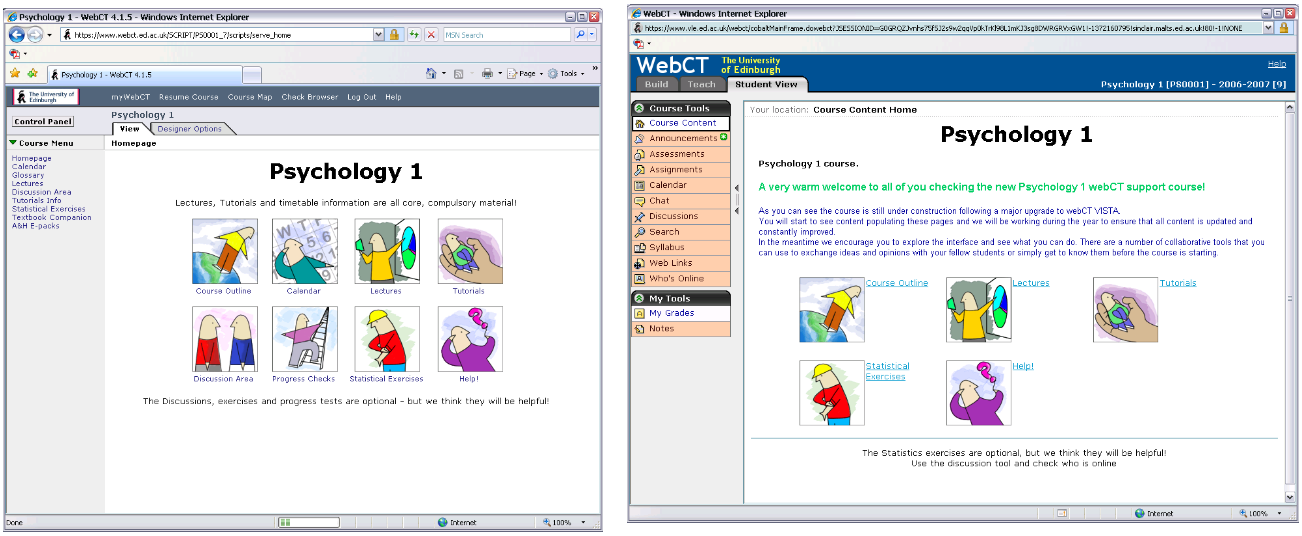


Figure 5.2. The main page of the first and last implementation of the psychology 1 course.
It should be noticed that in the first version direct links to specific pages were included in the left side menu (left). In the latest version direct links to tools were removed as these are forced in the 'Course tools menu'

The stats exercises in Psychology 1 allowed students to progress at their own pace through 7 units: an introduction to research methods, two units about distributions, correlation, chi-square, Mann-Whitney U and t-test.
The psychology 2 lacks the same level of interactivity, but statistics calculation exercises were offered in preparation for the exams.

5.2.4. Structural and semantic topology of the courses

Unlike a course handbook with a table of contents, which can guide students to the right section, online resources are non-linear. Recent research conducted by Nicholas and colleagues (Nicholas, Huntington, Williams & Dobrowolski, 2004), also CIBER project, December 2009), showed that in normal use of the web, users tend to hop between sites after two or three clicks. Because the content in the VLE is quite different students are expected to spend more time *browsing* the content necessitating a simple and clear structure to guide students' activity.

The material in both the psychology 1 and 2 courses is organised *semantically*, grouping resources into categories (see 5.1). The physical location of files into directories always mirrors the organization of content pages; the content was easily maintained and often prepared and released in a phased way.

The structure of the e-packs is quite different, closely followed the chapter organization of the textbook rather than the logical relations in the content.

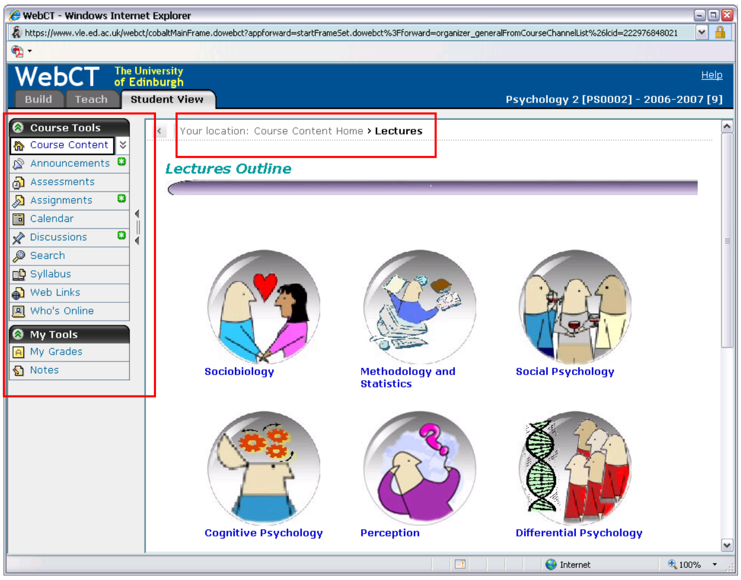


Figure 5.3. Navigation elements in WebCT 6 (Vista)

Highlighted the history pane in which the current location and the previous clicks are listed at the top. This feature was available in version 4 as well.

On the left side the Course Tools menu.

It is important to note that WebCT allows the designer to use a fixed number of tools and differentiates between 'organizer pages' (called folders in the latest edition), 'content modules pages' (renamed learning modules) and 'content pages'.

In practice, this means that whilst single content pages can be created outside the system as simple html (or other formats) files, content modules and organizer pages are composed from the various resources within WebCT and could simply contain links to files of external links. The distinction between these types of resources is important to identify the structural topology of the courses, which in turn has consequences for the tracking of individual pages.

Another crucial navigational feature is the organization of the links and the accessibility of the browser ‘normal’ toolbars: back and forward buttons which normally allow the user to move bi-directionally from a single page are not available. A history panel is offered as a substitute, but for the new user this is quite cumbersome.

Whilst the interface of the courses is fundamental for the users to navigate efficiently through the material available, the structural organization is essential to provide a navigation map useful to understand browsing patterns of the users.

In most web sites the map of the site is not explicitly available to the user, but the interface should hint at the structural organization and the difference between user expectancy and the site’s structure is one of the key issues addressed by usability studies to improve a web site. In the case of WebCT the constraints of the environment (i.e. the navigation tools like the history and tools menu shown in fig 5.2) force the user to behave in a different way than ‘normal’ browsing and therefore the semantic organization becomes essential to facilitate navigation.

5.3. Constructing feasible hypotheses: a synthesis

In the last section we contextualised the role of the VLE both in terms of the University settings and the pedagogical aims for the psychology courses.
With a better understanding of the context and the tools used, it is now possible to specify the most relevant questions, emerging in this context in relation to the review of the literature in the first four chapters.

After making a case for the evaluation of e-learning in higher education in chapter 1, the focus shifted onto a holistic conceptualization of learning and the conclusion that, in theory, ICT could dramatically improve students’ experiences of learning by using *personalization* and *adaptation*. At the end of chapter 2 it was hinted that students’ profiles could be a very powerful tool to direct the implementation and introduction of e-learning. In chapter 3, however, we looked at the existing literature in differential psychology, which studies the relations between intelligence, personality factors and academic performance. There is fairly strong evidence that academic performance is a useful proxy for intelligence -in the sense that there is a strong correlation between the two- and the predictive power of IQ is quite high, but such predictive power decreases at higher levels of instruction, with University grades being predicted well, but less so than in primary school children.

The relations between personality and AP, instead, are more complicated. For instance, there is some evidence that Conscientiousness and Openness might have an impact on AP, but the correlation is usually quite low. We also argued that the complexity of interrelations between IQ and personality is demonstrated by the evidence that the higher the IQ the more likely a person has a greater differentiation of personality traits. These make the consideration of one's profile too complex to be useful in practice and we suggested that *measures of styles* might be a more effective alternative.

In fact, styles proved to be valuable in a number of implementations of e-learning and in teaching narrowing the focus on specific *types*. We looked at the literature on styles in chapter 4 to make sense of this field of research and to decide whether appropriate measures were available. It turned out to be more challenging than expected as the field is riddled by a number of criticisms and shortcomings. Nevertheless, a number of integrative reviews were essential to steer the selection and to provide evidence of the utility of such measures.

Overall, the literature review seems to raise three core questions which require closer investigation:

1. What are the relations between the differential, behavioural and performance/achievement measures?
2. Is it possible to create *useful* models or students' profiles that can be applied to support teaching/learning as well as aiding automated instruction?
3. Does e-learning matter for students' learning?

To address these questions, a number of more specific hypotheses were formulated.

H1. Prior grades & future performance.

Considering the literature reviewed in chapters 2 and 3, it seems that AP is a fairly stable indicator: students with good grades generally perform well at higher levels of education. Given that entrance into university is quite competitive and entry requirements are high, is prior performance a good indicator of success in the psychology courses and the degree? The hypothesis is that there is a positive link between prior performance and university grades, however, as potential performance (like intelligence) is not infinite, some will do better than others *despite* excellent grades at lower levels of education. Even though it is arguable that league tables of schools should be taken into account to assess the *quality* of the grades, because exams are standardised, it is assumed that a grade is a valid and objective measure. Furthermore, it is not expected that there is a strong differential value of success for types of As (i.e. students more versed in maths or biology will not necessarily perform better than students who have As in subjects like literature or history, but it could be the case that students prepared in certain subjects might actually avoid other options at university).

H2. Styles and performance

There is some evidence that individual differences might account for some of the variance in AP, especially for those students who don't perform as well as expected from prior performance: measures of styles might help to identify possible differences. Using four different measures of styles (approaches to learning, two types of cognitive styles and intellectual styles) we will look at the emerging patterns and try to uncover what is most relevant. It is also expected that some measures will be more effective than others in predicting AP and we list here some more specific hypotheses, without excluding that other patterns will emerge from the exploratory nature of the data analysis:

1. A deep or strategic approach to learning will be positively related to better performance, whilst a surface approach will be associated with poorer performance; (ASSIST)
2. Because of the nature of the subject and the cohort, a more verbal style of processing will be related to better performance in essays and a more analytic style better performance in MCQ-type and report assessment; (VICS-WA)
3. There is a strong relation between certain styles and certain types of assessment and these have been found consistently in the literature (see table 4.6); (MSG)
4. Due to the nature of the assessment, students scoring higher on the intuition scale will have more trouble with practical experimental reports; (CSI)

H3. E-learning and performance

One of the most sought after results from the literature in e-learning is that AP is actually improved after the introduction of e-learning. We will assess this via both behavioural data and feedback surveys:

1. There will be a small but significant difference in performance for those who use e-learning *appropriately* (concept that will need a tighter definition emerging from the analysis of usage patterns);
2. Students using e-learning will be generally more satisfied by their experience.

This hypothesis is also the first step toward a better understanding of the third research question.

H4. Styles in e-learning

The most interesting and novel aspect of this thesis is the investigation of individual differences of usage of e-learning: the relations between styles measures and patterns of use is the core issue, however it is also the one which is the most difficult to pinpoint because of the scarcity of previous studies with a similar focus. For this reason, to investigate this question and exploit to the maximum the data obtained, we will use data mining techniques. We will address the specific methodological issues later in this chapter.

In general it is expected that patterns similar to the relations between styles and AP will emerge. In fact it seems obvious that students who are generally better motivated or inclined, will tend to maximise the full range of resources offered to them and hence perform better at the end. However, the most interesting aspect of this analysis will be to drill down and find out how those who do not perform as well as expected are affected.

H5. E-learning styles

Partially following from the previous hypothesis, if we consider that there are numerous and equally effective ways of coming to learn, another way of looking at the *appropriate* use of e-learning is to individuate patterns of usage that are more efficient than others. In computing science it is implied that after an algorithm to perform a specific task is implemented, optimality metrics will be provided to clarify the value of the improvements. As in the case in which one tries to reach point B from A on a map and discovers that there is a 'shortest route', in a similar way there could be a 'better' way of using the online material: the hypothesis here is that emerging patterns explored using data mining techniques will demonstrate in a bottom-up sort of way, specifically from the differentiation of performance at the end of the course, that some behavioural patterns will be closely related to measures of styles (i.e. with a positive relation between a 'chaotic' style of access and surface approach to study). Similar ideas were present already in Witkin's work in which stylistic preferences partially contributed in performance of specific tasks.

The second and third research questions are more complicated to address and will require a synthesis from the exploratory data analysis emerging from H4 and H5, and specifically leading to the last hypothesis.

H6. Useful patterns?

The final core hypothesis of this research will lead to investigation of the existence of useful patterns; using machine learning algorithms we will try to model performance based on the variety of parameters collected, and demonstrate that not all measures have an equally strong predicting value and will propose in the last two chapters which ones do have relevance. Such analysis will also contribute in answering the third research question.

5.4. The nature of the research conducted

5.4.1. Time span and cross-section of the implementation framework

The most important and original aspects of the research conducted for this thesis are the longitudinal nature of the data collected and the in-depth investigation which attempted to collate information from a wide set of sources for the same participants.
Overall we followed 2022 *starting* students[10] (1398 females, 69%, and 624 males, 31%, average age 19.6).

The nature of the longitudinal exploration was partly dictated by the cyclical nature of academic teaching, but the advantage of the organization of academic terms gave the opportunity to consider the research questions iteratively and refine the scope, before the next cycle started. Even though the study is quite large, the major drawbacks of this approach are the length of time required to collect the data, the high chance of obtaining incomplete datasets and the difficulty of obtaining adequate samples sizes to conduct appropriate analysis at each iteration point.

Despite these limitations a large database was constructed and the table (5.5) provides an overview of the data collected. The panels show a longitudinal as well as historical cross-section of the sample used, the e-learning implementation levels and the details of the academic performance measures collected at each iteration of research, which allows us to reflect on the impact on students' learning.

---

[10] The sample mentioned here comprises all students starting from the academic year 2002-03, however 2476 *active* students were enrolled, including those who started before 2002, but were in successive years within the period 2002-09, and this figure will be used for some statistics (i.e. graduation data).

e-learning Support Level				
	Y1	Y2	Y3	Y4
2002-03	NS			
2003-04	NS	NS		
2004-05	MS	NS	NS	
2005-06	FS	MS	NS	NS
2006-07	FS	FS	MS	NS
2007-08	FS	FS	MS	MS
2008-09	FS	FS	MS	MS

Academic performance				
	Y1	Y2	Y3	Y4
2002-03	EoY			
2003-04	detailed	EoY		
2004-05	detailed	detailed	EoY	
2005-06	detailed	detailed	EoY	EoY
2006-07	detailed	detailed	EoY	EoY
2007-08	detailed	detailed	EoY	EoY
2008-09	detailed	detailed	EoY	EoY

Table 5.5. Summary of the record of academic performance and corresponding e-learning support available.

Left: Level of e-learning support over the period of implementation. (NS= no support, MS=minimal support, FS= full support). Right: availability of measures of academic performance in the various cohorts/years (EoY: end of the year results for each module taken, detailed means that coursework assignments are also available)

Given the varied scope of the analysis, the database will be *sliced* to provide the adequate data to address specific questions. Throughout the thesis, we will use the term *class* when referring to a group of students taking a course at a specific level in a single year (i.e. the psy1 class in year 2006/07). We will use the term *cohort* when referring to a group of students who start in a specified year (i.e. the cohort of 2003 means students that started in 2003 and expected to graduate in 2006/07).

In general when using the term cohort it is implied that not all students starting in year 1 will reach year 3 or year 4 of the psychology degree. Nevertheless grades (including those in modules other than psychology) for all students taking the Psychology 1 or 2 course have been extracted for comparison.

An historical cross-section can be examined taking one class-level at a time: this gives an overview of the effects that each implementation had on the performance. It is particularly important especially for the cohorts starting from 2004/05 onward to note that this was the point at which e-learning was fully supported. In total, six longitudinal samples were followed (shaded with different colours in table 5.4).

The cohorts 2002 and 2003 are interesting as *controls*, in which no e-learning support was provided at all. Table 5.4 left, specifically indicates whether there was no e-learning support (NS) was present, minimal e-learning support (MS) or full support (FS). For each cycle it is possible to compare at least one instance of each degree of support (NS,MS,FS), but in years 1 and 2 we can also compare multiple instances with full support and track the evolution of these courses, which provides an in-depth picture of the impact of e-learning on both teaching methods and learning outcomes.

5.4.2. Data types, data collection and data mining

After drawing attention to the overall structure of the data collected, a brief diversion is necessary to address two important issues: one is related to the *types* of data collected, the other is about data aggregation and integration.

Psychology has been often criticised as the "science of the invisible" (Kennedy, 2009). The theoretical assumptions in psychology usually rely on the fact that processes can be *inferred* from the results of such processing. Observational data, on the other hand, are based entirely on metrics of behaviours which attempt to make observations systematic and *accurate*, but also rely on a number of assumptions to interpret expressed behaviours. The reliance on response times (RTs) in experimental psychology has been driving the field for the past century. However, as we explored in chapters 3 and 4, RTs are not the only way of studying psychological phenomena. Back in 1957 Cronbach expressed the hope that the two parallel streams of scientific psychology, *experimental* and *correlational*, would converge at some point in the future. In the previous chapters, we noted that it often seems that the two approaches are still going their own ways.

Strict experimentation and statistical hypotheses testing are the methods commonly applied to give psychological research a scientific value. However, with the advent of more powerful computing and cheaper storage, researchers have massively increased their ability to collect information. The problem of what to do with the amount of data collected has become a popular one in a variety of disciplines and computer science has developed some methods to explore massive datasets –commonly referred to as data mining or data modelling. In this thesis, data mining is adopted as a third methodological alternative to experimental and correlational research.

The breadth of this field of research is evident from the variety of possible definitions of data mining. For example, we considered in chapter 4 the definition of data mining given by Hand. The Encyclopaedia Britannica Online provides a similar alternative:

> "In computer science, the process of discovering interesting and useful patterns and relationships in large volumes of data."

However, it is quite interesting to note the narrow focus to its application in computing science. On the other hand, the Wikctionary (the first free, wiki-based, online encyclopaedia which is updated more frequently and by a wider variety of editors), uses the following definition, which gives a much wider scope:

> "A technique for searching large-scale databases for patterns; used mainly to find previously unknown correlations between variables that may be commercially useful." (accessed on 18/09/09).

One of the key features of both correlational and data mining approaches is the fact that the nature of the data used is mainly observational or from survey data. These two sources are familiar to researchers in psychology, and are typically valued because unlike experimental data, they offer better ecological validity, however they suffer from the lack of rigid and controlled ways of manipulating a few variables at a time typical of the experimental method.

In the research presented in this thesis, there are three core sources consisting of both observational and experimental types which will be explored in detail in the next section: *performance data, survey data (styles)* and *usage data*.
Performance data is very similar to the experimental paradigms in which a test of maximal performance is performed by participants and the degree of accuracy, proficiency or response times are used as indicators of ability.
Survey data includes both opinions and self-reported psychological metrics. Observational data provides a rich picture of students/users behaviours online. Contrary to some observational data, which might be subject to a subjective recording of the activity, logs of online activity give an accurate and objective record of the streams of actions performed online, over time and in multiple sessions, in which the user engages naturally with the medium. This minimises the role of prior assumptions regarding the expression of certain behaviours.

The purposes of different types of data allow encompassing both a *descriptive* and a *prescriptive* function for the goals of this research. The correlational and data mining perspectives offer a descriptive overview. The performance and experimental data, however, with the datasets sampled for modelling, provides a useful way of testing specific hypotheses more suitable for experimental designs. In fact, the processes and techniques used in computer science to *learn from data,* less familiar to researchers in psychology, prescribe that some samples should be used to *train* the machine learning algorithm, what other samples from the dataset should be set aside to verify the *goodness* of the algorithms and test the hypotheses in a quasi-experimental fashion (Witten & Frank, 2005).

Blending these methods of analysis provides a unique insight in the data collected and an original approach to study the relations between behaviour, personality and AP.

One of the most important shortcomings of such an approach, however, which in some respect is familiar to researchers in experimental psychology and medicine, is the necessity of reducing the complexity of data using some form of aggregation. The way in which data is aggregated and integrated requires particular attention as it might attract potential criticisms. For example, it is possible to introduce an intrinsic bias in the data by arbitrarily aggregating or separating groups. The effects observed in the study of a single variable in two groups might be reversed when groups are combined: this is known as the Simpson's paradox (or the Yule-Simpson effect) which needs to be briefly considered at this stage.

5.4.3. The Simpson's paradox: a 'textbook' example

Let's use a simple example referring to a conceptually relevant scenario: we want to consider the relationship between the length of time a user spends on a web page (i.e. short/long) and a qualitative attribution of the page value (i.e. essential/optional) with the distributions in the table below (5.6).
The categorical variable (Short/Long) is already derived from a raw numerical 'dwell time' variable; for clarity, this could be obtained from a median split of the distribution.

	Essential	**Optional**
Long	**10 (50%)**	**8 (40%)**
Short	**10 (50%)**	**12 (60%)**
tot	**20**	**20**

Table 5.6. Simple distribution of time spent on a page based on the type of page.

Considering the group data, a further aggregation is performed to count the frequencies. These two step contribute to simplify greatly the results and according to this table we could take two different interpretations:

- people are equally distributed in the length of time on essential pages;
- more people do short visits to the optional pages and overall, shorter visits are prevalent for optional content.

If we formulate these as experimental hypotheses, we would then be able to provide an appropriate statistical test. In the case of an experimental design, the interpretation is straightforward, but it is entirely possible that in a second experiment, which considers a third variable, we discover that the original grouping is actually biasing the findings showing a reversed effect.

To exemplify the importance of the reversal effect, we add and control the data of the example above according to a third variable (i.e. good/poor students as defined by prior grades) resulting in a distribution as in table below (5.7).

	Essential		**Optional**	
	Good	**Poor**	**Good**	**Poor**
Long	1 (25%)	9 (56.3%)	5 (33.3%)	3 (60%)
Short	3 (75%)	7 (43.7%)	10 (66.6%)	2 (40%)
tot	**4**	16	**15**	5

Table 5.7. Similar distribution of visits (see table 5.5), but with a third variable.

This table highlights a very important, but controversial point about the interpretation of the aggregation as it looks like ‘good’ students do shorter visits than bad students to both essential and optional pages, but this trend is hidden in the table 5.5 containing aggregated information.

In a purely experimental design, the presence/absence of the third variable is part of the design. However, the example provided here is essential in demonstrating that with a massive dataset and a number of possible variables contributing at different levels, the size of the groups cannot account for all possible combinations of the interactions and the extent of manipulation could affect the findings and the following generalizations.

Therefore, even *exploratory* data analysis, typical of data mining techniques, requires strong questions and falsifiable experimental hypotheses rather than a random or ‘melting-pot’ inclusion of variables in order to avoid the type of paradox presented here as an example.

5.5. Data sources

As indicated earlier, we collected a variety of data in each academic year. Each individual metric is explained in more detail in this section; however it will be useful for the reader to gain a broader understanding of the collection procedures by looking at a visual map of data collection approaches.

Figure 5.4 characterises the structure of the academic year in semesters and weeks. In Year 1 an introductory lecture on research participation, and basic study skills was provided in week4 after which the research participation experiments are opened to students. A number of studies are made available to students who can choose to participate for course credits. The measures of styles inventories (ASSIST, CSI, TSI) and the experimental sessions (VICS-WA) were offered as possible studies to take part in and the majority of students (roughly 60% of participants) completed the questionnaires between week 5 and 7. The rest of participants completed the inventories and the experimental sessions in the gaps between after the exams and the winter holidays. The end-of-year survey (week 10/11 of semester 2) was an anonymous form which students were able to complete in the last lecture of the course.

In year 2 we maintained a similar organization re-presenting the inventories as part of the experimental tasks given in class in week 6/7 and week 11 of the first semester.

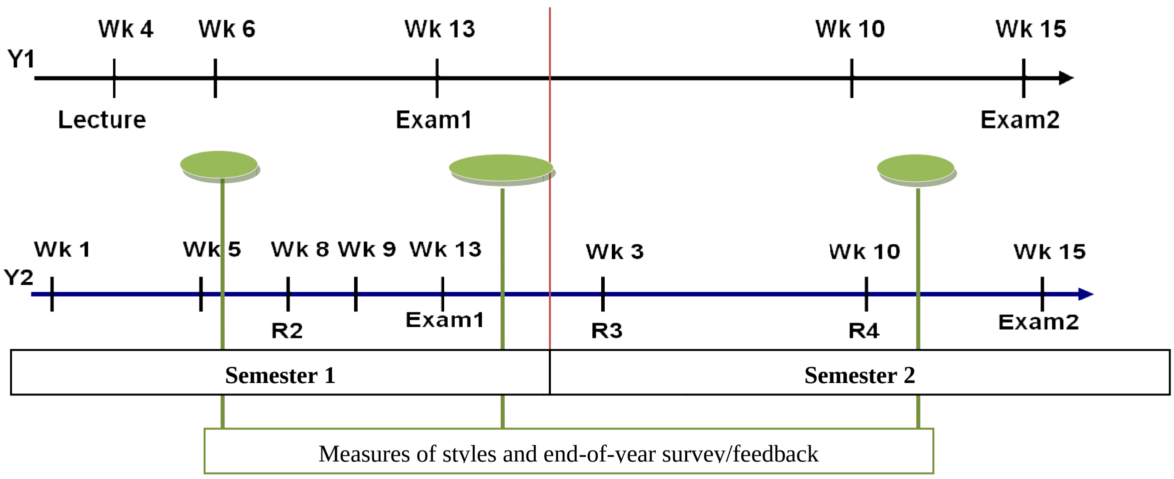


Figure 5.4. Data collection map over an academic year with details of the data collection points for the metrics of styles and assessments. In year 1 assessed essays are due in mid term and in year 2 Reports as in indicated by R2-R4

5.5.1. Prior performance

Entry grades were extracted for all students who enrolled for the Psychology 1 or the Psychology 2 courses (the latter to account for direct entry in the second year) from the centralised student management system (WISARD). As indicated in chapter 5, both courses are components part of the psychology degrees and essential for progression into the Psychology Honours degrees, but they are also open to other students as outside courses. WISARD is maintained by the University Registry and is the main repository for individual student information from admission grades to course marks and degree results.

Admission grades and subjects are normally accessible to the Directors of studies, who are assigned to each student and have a pastoral role in helping students to select courses and advise them throughout their degree. As we will consider later in the next chapter, admission grades were not available for about 20% of the students considered (see table 6.1).

Extracting this information for students taking psychology is essential to determine a possible effect on AP at university based on prior performance. In chapter 3 we noted that previous success is a good indicator of academic achievement at university; given the high grades required for admission on the psychology course at this University, there is an obvious potential bias determined by a high level of performance before entry.

5.5.2. Academic performance

Academic performance is generally believed to be the most objective measure for determining learning outcomes. It is regarded as a valid indicator of how well students do and can be considered at the same guise as a test of maximal performance (i.e. the outcome is the best possible result obtain for a test/assessment). We also suggested that this is a biased measure, because it is strongly dependent on the nature of the assessment.
For these reasons, in the attempt to set apart the effects of certain types of assessment, we recorded not only the final grades in the courses, but also a number of intermediate points including coursework, exam grades and, where available, self-directed or formative tests. Chapter 5 provides a more detailed overview of the assessment types in the various courses, but as a memento, coursework essays (in year 1) and reports (year 2) both contribute to a percentage of the final grade; self tests and practice submissions were also assessed, but didn’t count for students’ final marks.

5.5.3. Survey data

End of the year survey data was collected as part of the feedback process at the end of the year. These surveys contained questions regarding the quality of the courses, but we also included questions related to the e–learning provision assessing four different areas: Usability (related to the system), Motivation (for taking the course and using the resources), Confidence (in using computers, the internet and the resources), and Satisfaction (in using the system and in relation to expected outcomes). As the survey provided only group data, this source was only use as informative background (see Hardy et al. 2006 and Appendix 3 for more details).

The inventories of styles are detailed in the next section; it should be noted that styles metrics were collected between weeks 6 and 9 of the first semester each year considered and students received course credits for taking part in year 1, or they were embedded in the practical course tasks in year 2.

5.5.4. Measures of styles

We have already explained the rationale for choosing four measures of styles; here we provide more specific details of the tools used, frame their expected validity and reliability based on previous research, and identify any modification (if any) implemented, which emerged from specific criticisms in the literature.

Approaches to learning: ASSIST

The way in which students learn and study has attracted considerable interest in the past half century, especially because of the rapid changes in education highlighted in chapters 1 and 2. The ideas of strategies or approaches to learning and studying, can be traced back to the seminal work conducted by Marton & Saljo in 1976. They defined two broad but different strategies to tackle a learning task, which they termed *surface* and *deep* approach. Over the years, Entwistle and colleagues, taking a phenomenographic approach, developed the concept further. In collaboration with Ramsden, Tait and McCune he refined a questionnaire which identifies students' preferences for learning and provides further evidence for the deep and surface construct. In their work, however, they also added a *strategic* approach to take into account for a preference and modality that largely resembled the deep approach for some aspects, but could also take features of a surface approach in determined circumstances.

It is fundamental to understand that, in their view, approaches are malleable: contrary to other researchers, scholars embracing this approach believe that learning is contextually based and bottom-up (Ramsden 1992). For this reason, *approaches* could be different if considered in different disciplines and often even between courses in the same subject area.
Another important aspect is that learning outcomes are directly influenced by students' orientations to learning. This is shaped by previous experience of education and their perception of task requirements.

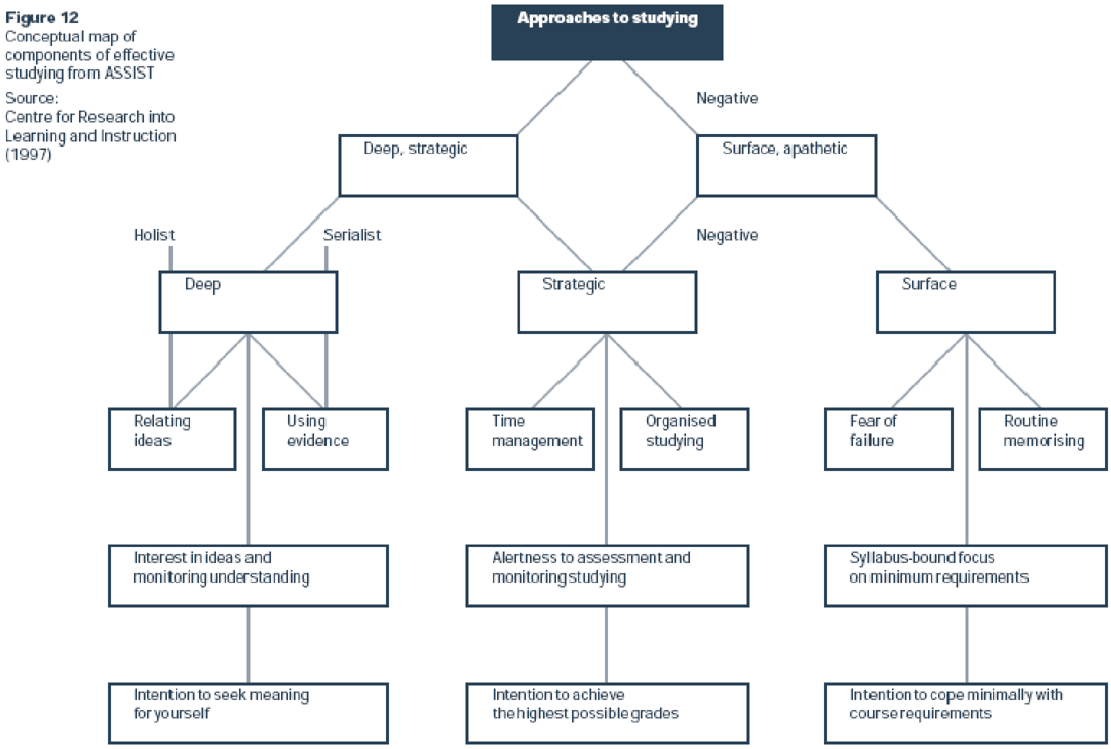


Figure 5.5. Conceptual map of the components of the ASSIST (from Coffield et al. 2004)

In the tradition of students' approaches to learning we can find a number of instruments developed around the world: Bigg's (1987) Study processes questionnaire (Australia), Schmeck (Schmeck & Grove, 1979). Inventory of Learning styles (US), Vermunts' (Vermunt, 1995, 1996) Inventory of learning styles (Nederlands) and Entwistle (1981, 1982) Approaches to studying inventory (UK). The latter has been used widely in higher education and has been going through a number of revisions over the years, making it effective and easy to use.

In this study we used a very recent modification of the ASI, the ASSIST inventory from the ETL project (Entwistle & McCune, 2004, Hounsell et al. 2005), which identifies the three dimensions (deep, surface, strategic) and 13 subscales detailed in the figure (5.4). The reliability and factor structure reported by McCune et al. (table 5.7) shows the conceptual structure of the constructs and dimensions.

Factor loadings and Cronbach alpha coefficients for ASSIST sub-scales				
(N = 1284)	factor I	factor II	factor III	alpha
Approaches to Studying				
Deep Approach	0.84			0.84
Seeking meaning	0.67			0.57
Relating ideas	0.79			0.59
Use of evidence	0.75			0.53
Interest in ideas	0.65			0.76
Surface Apathetic Approach				0.8
Lack of understanding		0.77		0.54
Lack of purpose		0.37		0.68
Syllabus boundness		0.35	-0.3	0.62
Fear of failure		0.68		0.76
Strategic Approach				0.87
Organised studying			0.77	0.57
Time management			0.86	0.76
Monitoring effectiveness	0.42		0.48	0.55
Achievement motivation			0.78	0.69
Preferences for learning environments				
Deep (Encouraging understanding)	0.55			0.62
Surface (Transmitting information)		0.38		0.69
Self-rating of academic progress		-0.31	0.47	
Correlations between factors				
Factor I (Deep)	1			
Factor II (Surface Apathetic)	-0.2	1		
Factor III (Strategic)	0.35	0.22	1	

Table 5.2. Factor loadings and alphas of the ASSIST. From McCune & Entwistle 2000.

There are two interesting elements: on one hand there is a partial conceptual overlap of the well known style types (holist/serialist). On the other, there is a clear identification of value-laden constructs, in which the surface-apathetic approach is negatively affecting learning. The ASSIST is interesting for another reason: about 10 years ago, McCune used the same instrument with a similar group of psychology 1 and psychology 2 students at this university, creating a very useful benchmark for comparison.

Measures

From table 5.7, the factor analysis of the original ASSIST questionnaire provides evidence of the three core approaches (surface/apathetic, deep and strategic) and within each subscale which allows to individuate more specific aspects of each approach. Two facts are notable:

firstly, the factors are not completely independent of each other and secondly small correlations exist between the approaches.

Furthermore, even if in this study McCune reported a large sample of 1284 students in the UK, she also indicated that the mean scores for each dimension are centroids and *should not* be considered as *typical* or *normative*. In the next chapter, however, we will compare our sample with this study and argue for a certain degree of consistency and stability of the constructs.

This position is not dissimilar from the theoretical point of view supported by Entwistle: he explicitly drew from an influential study by Perry (1970) to argue that students' understanding and progress of the approaches to learning are strongly dependent on their *conceptions of learning*. The approach used by a student is the *preferred way* of tackling a learning task and the student *chooses* the way to deal with a specific learning task in light of the *perceived demands*. It is no surprise that Entwistle also indicated that the type of assessment strongly affects the approach used by a student and that a strategic approach (which also includes the use of the surface approach for certain learning tasks) is required at university level, where summative assessment has such a central role in the curriculum. The terms italicised in the last sentences have a particular prominence in the argument, which will emerge from the data analysis, specifically because we will take into consideration different modalities of assessment and sample students with different motivations and expectations from the courses.

Relation with other instruments

According to Coffield et al. (2004) there are over 100 studies which address the empirical and theoretical value of the ASI and its successors. These can be broadly categorised into four types related to: the theoretical and conceptual development of the tools, the refinements and reliability analysis, the implications for pedagogy and the theoretical implications and relations to other tools. Some of these relations were identified in chapter 4.

Applications with learning technology

Smith et al. (Smith, Colbourn, Cunliffe-Charlesworth, & Lever, 2006) reported the use of the ASSIST as part of the *CLaSS* (Cognitive Learning Strategies for Students) project. This was intended to provide an online facility to improve personal development planning (PDP) in first year psychology students across a number of UK universities. Rather than using the measures of styles to direct the interaction with the system, the results of the questionnaires (ASSIST, LSQ and VARK) were used for personal reflection and as the basis for face-to-face focus groups.

Such an application of the ASSIST is typical of the strand of educational research in which self-awareness and personal development are the focus and the ASSIST is one of the instruments used to gain a better understanding. Scores are not important per se, but used as a starting point for self-reflection. Unlike psychometric measures which are considered objective, approaches to learning are *subjective* preferences and as such difficult to compare. It is arguable that the feature of the metrics provided by this tool are meaningless taken out of context and without a personal reflection, but the validity and reliability of the scores seems to lead to a stronger interpretation.

Limitations and validity.

The verdict on validity and reliability from Coffield et al. (2004) is that the internal and external evaluations of the ASI and ASSIST suggest a satisfactory reliability and internal consistency, but the authors observed that many of the subscales are less reliable. Furthermore test-retest reliability is rarely shown in the published literature: this makes claims about the predictive validity of the tool, as well as the identification and understanding of changes in the conception of learning over time, weak in many respects.

In the context of this thesis we will try to tackle some of these issues: for example, by using the comparison between Y1 and Y2 we will be able to provide some evidence of the test-retest validity as well as the possible change of approaches in different courses (within the same discipline). We will also be able to compare the motives and conceptions of learning in students taking the same courses but with different scopes (i.e. psychology and non-psychology students). The direct comparison with other measures of styles will also provide some evidence for the conceptual overlap of styles constructs.
Finally, we will be able to compare the overall patterns with McCune's sample investigating approaches to learning 10 years ago within the same courses.

Intellectual styles: the theory of mental self-government

The Mental Self-Government (MSG) and the Thinking Styles Inventory (TSI) are based on Sternberg's (1988) theory and measure 13 different styles of thinking. The mental self-government theory was Sternberg's attempt to unify the three main categories of theories and types of styles: cognition centred, personality centred and activity centred.

According to this model, mental self-government is used metaphorically to provide a coat hanger to fit in styles which resembles the societal structure of governments. Sternberg stated that the various scales of the TSI measure styles generated by the functions and mental processes which utilise them. There are 13 styles that fall into 5 core dimensions: *functions* (legislative, executive and judicial), *forms* (monarchic, oligarchic, hierarchical and anarchic), *levels* (global and local), *scopes* (internal and external) and *leanings* (conservative and progressive). Within this theoretical framework, thinking style is defined as a profile of styles describing a person's preferred ways of thinking in specific contexts. The table below (5.8) offers a more detailed overview of the features of each scale.

Dimension	Thinking style	Key characteristics
Function	Legislative	Work on tasks that require creative strategies; Choose one's own activities.
	Executive	Work on tasks with clear instructions and structures; Implement tasks with established guidelines.
	Judicial	Work on tasks that allow for one's evaluation; Evaluate and judge the performance of other people.
Form	Hierarchical	Distribute attention to several tasks that are prioritized according to one's valuing of the tasks.
	Monarchic	Work on tasks that allow complete focus on one thing at a time.
	Oligarchic	Work on multiple tasks in the service of multiple objectives, without setting priorities.
	Anarchic	Work on tasks that would allow flexibility as to what, where, when, and how one works.
Level	Global	Pay more attention to the overall picture of an issue and to abstract ideas.
	Local	Work on tasks that require working with concrete details.
Scope	Internal	Work on tasks that allow one to work as an independent unit.
	External	Work on tasks that allow for collaborative ventures with other people.
Leaning	Liberal	Work on tasks that involve novelty and ambiguity.
	Conservative	Work on tasks that allow one to adhere to the existing rules and procedures in performing tasks.

Table 5.3. Dimensions and characteristics of the Mental self-government theory. Reproduced from Sternberg (1988).

Measures

The original inventory had 104 statements (8 for each scale). Participants have to indicate their agreement with the statements based on a 7-points scale.

Even though this is already a fairly long questionnaire, as noted by Nielsen & colleagues (Nielsen, 2006; Nielsen, Kreiner, & Styles, 2007), the inventory completely lacks a *democratic* form of government. This is quite strange as democracy is one of the classic forms of government, already present in ancient Greece, and at the core of the organization of modern western societies. According to Nielsen the classic definition of democratic values is perceived as a positive form of government, but the concept has become more ambiguous because of the varied implementations observable in western societies. This makes it more difficult to pinpoint the concept precisely. Nonetheless, following DeVellis (2003) guide to scale development, they extended the inventory and added a further 8 items to account for a *democratic* scale in the forms of government, in the Danish version of the inventory.

Items for the democratic style scale from Nielsen 2001	**Revised wording used in the thesis**
DEMO1) When I have many things to do, I prefer to do the ones that are important to myself as well as to others. DEMO2) In talking or writing down ideas, I prefer to involve both my own ideas and the ideas of others. DEMO3) When deciding how to go about a task, I prefer to do it through dialogue with others. DEMO4) I prefer to switch from one task to another, if the new task is more important to someone else. DEMO5) When starting on a new task or job, I prefer to know its possible significance to other people. DEMO6) I like to work with problems that are important to myself as well others. DEMO7) I can easily prioritize the needs and goals of others above my own. DEMO8) When deciding which of several task to do, I weigh my own needs against the needs of others.	DEMO1) When I have many things to do, I prefer to do the ones that are important to myself as well as to others. DEMO2) In talking or writing down ideas, I prefer to **draw in** both my own ideas and the ideas of others. DEMO3) When deciding how to go about a task, I prefer to do it through dialogue with others. DEMO4) **Sometimes** I prefer to switch from one task to another if the new task is more important to someone else. DEMO5) When starting on a new task or job, I **like to understand** its possible significance to other people. DEMO6) I like to **engage** with problems that are important to myself as well as others. DEMO7) I can easily prioritize the needs and goals of others above my own. DEMO8) When deciding which of several task to do, I weigh my own needs against the needs of others.

Table 5.4. Items of the ‘democratic style’ scale from Nielsen and revised wording.

Nielsen et al. (2006) then translated the items back into English. However, in this research, we slightly revised the wording and included them following a similar structure to the original inventory (see table 5.9).

In the original pen and paper questionnaire, items were presented in 7 pages with 14-16 statements each. The new statements were inserted in the pages regularising the number of statements in each page to 16 items per page. (1,1,2,0,2,1,1 items added for each page in the sequence). The final version of the revised inventory has 112 statements.

Relation with other instruments

The TSI has been used extensively and validated with different cultures (Sternberg, 1999; Sternberg, 1999; Zhang & Sternberg, 2005). It was also used with both students and teachers and related with a number of tools. In particular, the recent work by Zhang, conducted in multiple research projects across different cultures contributed a wealth of knowledge on the relationships between the construct of intellectual styles and other styles measures as well as personality traits and academic performance.
The table below (6.7) provides an overview of the relations between the TSI scales and other styles constructs. In her regression analysis, Zhang (Zhang, 2004; Zhang, 1999) provided evidence that the hierarchical, judicial and monarchic dimensions are good predictors of academic performance in a variety of disciplines, but unfortunately, there is no specific mention of psychology in her list.

Applications with learning technology

Despite the wide use in educational settings and the wealth of knowledge accrued about the relevance of intellectual styles, the TSI has not been used in an applied e-learning implementation and has not been used with the ASI (only Bigg’s SPQ).
Given the nature of e-learning (i.e. self-paced resources and potential interactivity with the system, peers and tutors) the typology of styles identified by the TSI offer the potential to draw interesting relations between styles and our observational data, especially when considering the clear relations between learning activities and styles already mentioned (see table 4.6).

Limitations and validity.

The mental self-government theory is quite appealing for a number of reasons. As part of a wider integrative model it offers useful connections with research in a variety of areas spanning from academic performance, teaching, personnel selection and personality research. The practical relevance of the model is apparent and Sternberg made a compelling argument, detailed in chapter 3, about the importance of intelligence not measured by IQ or measures of maximal performance.

Furthermore the TSI has been found to be a reliable tool by Sternberg and his associates, with a lot of evidence from multi-cultural studies (Cronbach alpha value ranging between .5 to .8 in all studies) . However external validation has been more cautious.

For example Porter (2006) reported that the length of the questionnaire used is one of the things that most affect responses with first year students. Even though Coffield et al. (2004) dismiss the value of the TSI as “An unnecessary addition to the proliferation of learning styles models”, the variety of comparisons with models in both personality and styles research allows placing the TSI clearly against its theoretical background. The threefold model of intellectual styles emerging from this research is quite explicit in identifying three core aspects of the theory:

- Intellectual styles represent states as they can be socialised, and are modifiable/teachable;
- styles are variables across tasks/situations and change over the life span, but are generally stable;
- there exists a varying degree of overlap between the TSI and other construct, therefore a clear position about the dependence/independence of the TSI is not taken arguing that there are important similarities.

In fact, Coffield et al. are quick in saying that the reliability scores are much lower than those reported by the authors from two studies: (Demetriou & Kazi, 2001; Porter, 2003), on a Cypriot sample (which required a translation of the items, possibly introducing a spurious effect on validity). These is not enough to support the strong claim made by Coffield et al.

	Style type	Type I	Type II	Type III
	[a]Learning approach	Deep	Surface	Achieving
	[b]Career personality type	Artistic	Conventional	Realistic, Investigative, Social, Enterprising
	[c]Mode of thinking	Holistic	Analytic	Integrative
	[d]Personality type	Intuitive, Perceiving	Sensing, Judging	Thinking, Feeling, Introversion, Extraversion
Style construct	[e]Mind style	Concrete random	Concrete sequential	Abstract random, Abstract sequential
	[f] Decision-making style	Innovation	Adaptation	
	[g]Conceptual tempo	Reflectivity	Impulsivity	
	[h]Structure of intellect	Divergent thinking	Convergent thinking	
	[i]Perceptual style	Field independent	Field dependent	
	[j]Thinking style	Legislative, Judicial, Global, Hierarchical, Judicial	Executive, Local, Conservative Monarchic,	Oligarchic, Anarchic, Internal, External

Note. Theoretical foundations: [a]Biggs's theory of student learning, [b]Holland's theory of career personality types, [c]Torrance's construct of brain dominance, [d]Jung's theory of personality types, [e]Gregorc's model of mind styles, [f] Kirton's model of decision-making styles, [g]Kagan's model of reflectivity-impulsivity conceptual tempo, [h]Guilford's model of structure of intellect, [i]Witkin's construct of field-dependence/independence, [j]Sternberg's theory of mental self-government.

Table 5.5. Relations between Intellectual styles and other styles constructs. Reproduced from Zhang & Sternberg 2005, p 38

Cognitive styles: the Cognitive styles index, a self-reported measure

The Cognitive Styles Index (CSI), developed by Allinson & Hayes (Allinson & Hayes, 1996; Hayes & Allinson, 1998) is another self-reported instrument. Originally it was developed as an alternative, simple method to assess the unitary construct of intuition-analysis. In their view, this unitary framework offered an overarching dimension underpinning a number of styles constructs. Such overlaps were already observed in other integrative frameworks (see chapter 4 for details). However, even from early research, such uni-factorial structure of the construct has been openly criticised. In particular, work conducted by Sadler-Smith and Hodgkinson (Hodgkinson & Sadler-Smith, 2003; Hodgkinson et al., 2003; Hodgkinson, Sadler-Smith, Sinclair, & Ashkanasy, 2009) brought evidence that intuition and analysis should be considered as orthogonal factors rather than the opposite of the same dimension, but even in the most recent publications the two variants (unitary and dual metric) are often reported together. Hodgkinson & Sadler-Smith (2009) specifically revisited the dimensions of intuition and analysis and explicitly related the CSI with Epstein's Rational-experimental inventory (Epstein, 1998; Epstein, Pacini, Denes-Raj, & Heier, 1996). Sadler-Smith has been a strong supporter of a dual view of analysis and intuition, partially informed by Epstein's views and partly by recent neuroscientific evidence identifying separate neuronal pathways dealing with these two types of processing (Lieberman 2007). However, in this later study, after conducting an analysis at both item and scale levels, Hodgkinson & Sadler-Smith were led to the discovery of a three-factors model, which doesn't make the argument any clearer, but certainly provide evidence against the unifactorial model originally advocated by Allinson & Hayes.

Measures.

The CSI was developed from a pool of 129 items from the domain of intuition and analysis. In its latest version, there are 38 items (17 intuition items and 21 analysis items) which are scored by means of a trichotomous scale (true, uncertain, false). This is a quite unusual method for psychometric tools which are often scored with the more familiar Likert scale and it is a possible criticism in the evaluation of the psychometric properties of the instrument and the validity of the scales.

Relation with other instruments.

The CSI has been used extensively in large-scale organizational studies, particularly related to cross-national and cultural differences, gender differences, personality scales, entrepreneurial behaviour and leadership as well as learning styles. The predominance of research in a field other than education makes it particularly appealing as it provides yet another perspective to the examination of the value of styles in practice. There is also a study by Sadler-Smith (Sadler-Smith, Spicer and Tsang 2000) in which the CSI was related with the Riding's CSA, but no significant correlations were found.

Applications with learning technology.

Despite the wide use of the CSI in both academia and the industry, only one study has been found in which the CSI was used in relation to e-learning and instructional technology. This is the work conducted by Graff (Graff, Davies, & McNorton, 2004) who explored the interactions between the cognitive style, a computer attitude scale and internet use. Although this study did not go into the details of how the internet is used, it is interesting because it highlighted differences in gender and cultures, and certainly affects the way in which attitudes toward computer-based learning is promoted and implemented. We will consider these observations in the evaluation of the use of the CSI in our courses.

Limitations and validity.

Despite the open argument about the structure of the intuition-analysis dimension detailed by Sadler Smith and colleagues, the CSI has proved to be a reliable instrument (both in terms of internal consistency and test-retest reliability). The variety of studies relating the CSI with other dimensions provided strong evidence for its construct validity.

The core limitation and crux of this instrument is the careful consideration of the value of the metric: as indicated earlier the two-factors model advocated seems to be a more convincing framework, but in our analysis we will also explore the unifactorial version of this model to verify both validity and reliability in more details. Coffield et al. (2004) gave a positive assessment of the CSI (despite the fundamental issue of metrics highlighted above): "Overall, the CSI has the best evidence for reliability and validity of the 13 models studied". This is probably because the CSI has been always used in parallel to other tools providing a better insight of its relation with personality measures, job/management metrics and in education. However, the relation between learning styles, academic achievement and the CSI require further investigation, which we intend to provide in this thesis.

Cognitive styles: VICS-WA, a task-based measure

A different measure of styles which is exploring the “visual vs. verbal” and “analytic vs. wholistic” dimension is the VICS-WA. It is different because rather than using self-reported measures via questionnaire-based items, the VICS-WA uses response times in a modified version of the CSA (Riding & Cheema 1991); the two dimensions are measured using median reaction time ratios. Even if Peterson et al. (Peterson, Deary, & Austin, 2003a, 2005) demonstrated that the VICS-WA had a much higher reliability than the CSA, when directly compared with the CSA, Peterson (Peterson & Deary, 2006; Peterson et al., 2003a, 2003b) didn’t find a correlation between the two measures. This makes the VICS-WA more like a cognitive task than a measure of styles used by others in this research and therefore provides a useful way to relate styles to ability.

Measures

The VICS-WA is a computer program used to present a number of items on the screen and the participant is called to make a decision about the items. The choice reaction times computation of a visual/verbal ratio and a wholistic/analytic ratio

The VICS-WA has four main sections: a verbal task (116 trials), Imagery task (116 trials), wholistic task (40 trials) and analytic task (40 trials). Participants are presented with pairs of stimuli in which they have to make a choice between class (natural or man-made), size or have discriminate figures in an embedded figure task.

VICS

The VICS (Verbal Imagery Cognitive Styles) test is based on the Cognitive Styles Analysis (CSA) by Ryding & Rainer (1998) and was developed by Peterson and colleagues (Peterson, Deary & Austin 2003) to measure the preference for verbal or visual information.

The VICS has two sections: in the first section, pairs of items are displayed and the participant has to make a choice about the category of the two objects presented (items are in either the verbal form –i.e. the word rabbit, in visual form –i.e. picture of a rabbit or mixed). The categories used are simple distinctions between natural, man-made or mixed (if the objects are of different types) and the choice is made by pressing the key 1 to 3 of the keyboard.

In the second section participants, using the same method of presentation and response, are asked to assess if the size of the object on the left is smaller, bigger or the same size as the

one on the right. Note that the size of the picture is the same for all, but there are small differences in how the items are filling the space (see examples in Appendix 3).
The number of stimuli used in the VICS test and their component break down is given in the table below. The number of stimuli is shown in brackets.

Verbal Imagery Cognitive Style (VICS) test - (232 stimuli)											
Verbal Task (116)						Imagery Task (116)					
Words (58)			Pictures (58)			Words (58)			Pictures (58)		
N (26)	M (26)	Mx (6)	N (26)	M (26)	Mx (6)	B (26)	S (26)	E (6)	B (26)	S (26)	E (6)

Table 5.6. The structure and number of trials of the VICS test.
Key: N = Natural, M = Man-made, Mx = Mixed, B = Bigger, S = Smaller, E = Equal

Extended CSA-WA

The Extended CSA-WA test measures preferences for structuring information in a wholistic versus an analytic way. This test was based on the original Cognitive Styles Analysis (CSA) by Ryding and Rainer, but in the current version the test was modified to increase the test-retest validity of the task (Peterson et al. 2003).

Extended CSA-WA (80 stimuli)			
Wholistic Task (40)		Analytic Task (40)	
Original CSA Wholistic Items (20)	New Wholistic Items (20)	Original CSA Analytic Items (20)	New Analytic Items (20)

Table 5.7. The structure and number of presentations of stimuli for the CSA-WA test.

The task is delivered using the same program of the VICS with a similar methods of presentation and response. In the first section the participant is asked to decide if the two geometrical shapes presented are the same or different by pressing the keyboard (yes /no). In the second section of the task they are asked to evaluate if the shape presented on the left is contained in the picture on the right; a yes/no response is again activated by key presses.

The number of stimuli (shown in brackets) used in the Extended CSA-WA and their component break down is given in table 5.12.

Relation with other instruments

As this instrument is fairly recent and the testing of each participants requires a substantial amount of time (30-45 minutes), there are only a few papers published which used the VICS-WA. Apart for the research conducted here and the comparison between the VICS-WA and the CSA (Riding) in Peterson's work, we are not aware of other published material which compares the resulting styles with other instruments.

Applications with learning technology

An interesting application is the one presented by Uruchrutu et al. (2005) who investigated whether the types emerging from the use of the VICS-WA affected the way in which students on an MSc course in electronics interacted with an e-learning system.

Their findings were not as clear-cut as predicted and they showed that even though some individual differences affected the way learners react to and perform under different interface conditions, no simple effects were found in the relationships between styles and interaction. One of the reasons might be that they used a very small sample of students all from a specific background (i.e. engineers): in fact all the participants could be found in the analytic/visualiser quadrant. Even though the tool itself has seen limited applications, the constructs underlying the tool have been explored in a number of research studies. For example, Pillay (1998) looked at the effect of the constructs measured with Riding's CSA on computer-based instruction and found that even though there was no significant difference between performance with mismatching styles of presentation of material and preferred cognitive styles, the Wholist/Verbaliser group performed better than all the others.

Graff (2003, 2005) also investigated the relations between the Riding CSA and the hierarchical structure of websites, particularly looking at browsing strategies. He found some differences in how verbalisers and imagers with verbalisers visiting more pages in the hierarchical structure and imagers visiting more pages in the relational structure.
In particular, they found that the recall of information was affected by the interaction of the architecture and styles, but no significant differences were observed when learning was

assessed. Such finding is interesting because it seems to imply that in an experimental condition, at a more superficial level, the way in which information is organised and presented affects the interaction for people with different styles, but does not affect complex learning. Given the important relevance for web design and e-learning implementation, as well as the extent of analysis provided in this thesis, we will try to investigate if this is the case in a more ecologically valid context, and study potential differences in the online behaviour for people with different styles.

More recently, Miller (2005), using the Kolb Learning styles inventory found that students identified in the *concrete sequential* group learned significantly less than other students. This seems to suggest that students with analytic/visualise might emerge as less effective in learning online.

Limitations and validity.

A full discussion about the validity of the VICS-WA can be found elsewhere (Peterson et al. 2003, 2005, 2006) The fact that this test seems to be more reliable than the CSA makes it an ideal candidate for this thesis, because it is intrinsically different to any of the other measures of style and provides a wider set of variables to identify students' profiles. Coffield et al. (2004) praised the simplicity and potential value of the model, but, like Peterson, they heavily criticise the validity of the CSA as a suitable instrument to measure the constructs in the model.

An important counter-argument is that like most experimental psychology, the tasks are based on a core assumption about what reaction times can tell us about mental processing that the self-reported measure cannot do. The criticism that the range of difficulties offered by the test is rather limited is countered by Peterson's extension of the test. Because it purports to assess preferences, it might be possible that with the form of a test of maximal performance, given the narrow sample of students of similar ages and intelligence, this test might not be able to provide a satisfactory discrimination between different styles of processing, but it does offers a useful alternative to self-reported measures.

5.6. From online usage to web usage mining (WUM)

The web is a massive, widely distributed hyper-medium in continuous growth. The nature of its structure is untamed and standards are continuously updated to improve the user experience by offering methods which seamlessly allow integration of wide variety of information. This is the case also for a smaller internal network such as the one supporting the Virtual Learning Environment within a University domain which requires constant attention and frequent updates to scale the service. In chapter 2, we introduced the value of usage metrics to evaluate e-learning. In this section we provide a more detailed overview of the measurement of online usage and the extraction of behavioural information; a more technical treatment is offered in Appendix 1 which should be used to complement this chapter and the following analysis.

The range of online activities recorded within the VLE is the main source of behavioural data used in this thesis. Although this is only one aspect of students' learning activities, the literature reviewed indicated that this learning support tool is an important aspect of students' experience.

To better understand the value of the data collected, in this section we first look at how behavioural data from web usage has been used and represented in the literature, then we will consider the practical methods used to make sense of the sequences of behaviour applying data mining and visualization techniques.

5.6.1. Users' models from web usage

Three prominent models of web usage behaviour have emerged in the literature: Pirolli et al. "information foraging model", Kitajima's CoLiDeS (Comprehension-based Linked-model of Deliberate Search) and Miller and Remington's MESA (method for evaluating site architecture).

What is most interesting, however, is that in many respects, all these *user-centred* frameworks largely ignore the *characteristics* of the users, which is what makes them different from each other.

The relevance of web architecture: the MESA model

For example, usability and information architecture design had a central role in human-computer interaction research for a long time.

Since the advent of graphical user interfaces (GUI) on desktop computers, HCI researchers have contributed a great deal to make software applications more usable and efficient both from an ergonomic perspective and a productivity point of view.

Within this tradition during the 1980s and 1990s a great deal of evidence was gathered about the optimal way of structuring menu trees and navigation paths. In particular, the focus was on the number of items which should be displayed on a list and the depth or levels of sub-menus (Landauer & Dumais, 1997; Landauer & Nachbar, 1985). Lee and MacGregor (1985) indicated that the optimal number of selections per page ranged between 4 and 13, assuming a reasonable range of reading key-press times. One of the first studies focusing specifically on web navigation was the study by Larson & Czerwinski (1998). Their results were at odd with the menu studies as when a user was searching information in a website, their users took significantly longer times in a three-tiered structure with 8 links per page than in a comparable 16 and 32 links structure. Although Miller and Remington (2004) argued that the discrepancy was due to the quality of the labels used, these studies stress the importance of structure and organization mentioned above.

Research centred on architectural and structural features of websites, contributed to advance our understanding of web user interfaces and optimising them. Many are unimpressed by the 'minimalist' and 'hobbyist' look of HTML pages prevalent in the 1980s, but few realise that the modern dynamic features which make the visual effect rich and the user interaction efficient are the result of years of work on the user interfaces.

Information foraging

The research on architecture also showed a fundamental problem of the world wide web. The growth of pages exponentially multiplies the links, making it more difficult, rather than easier to find relevant information. The main driver for a theory of information foraging came to light as a reaction to the mainstream literature in user interaction and tutoring systems in the 1990s. What was termed as 'cognitive engineering models' lead to applications and implementations which, although fairly large and complex, were well defined, had limited goals and were specifically framed within a particular domain of knowledge.

In contrast, it seemed that people searching for information were largely constrained by the environment or the structure/architecture of the content, and only minimally influenced by the user's knowledge of the interface or system used to access the information (Pirolli 2007).

Inspired by behavioural ecology, and in particular by the optimal foraging theory by Stephen & Krebs (1986) as well as rational analysis proposed by Anderson (1981) Pirolli argued that because knowledge has become central in human activities, the concept of foraging for information is very plausible. Miller (1984) already suggested that humans could be viewed as *informavore*, “a species that hungers for information in order to gather and store it as a means for adapting to the world” (in Pirolli 2007). In this sense, because humans can obtain and use the right information increasing their gains, in evolutionary terms, their ability of accessing and finding valuable information increases fitness. When this idea is applied to web usage, Pirolli used the well-known ACT-R model (Anderson 1981)to provide a rational way of interpreting behaviours. In particular Pirolli and colleagues (Chi, Pirolli, Chen, & Pitkow, 2001; Chi, Pirolli, & Pitkow, 2000; Fu & Pirolli, 2007) defined the concept of *information scent* as the mathematical formulation to model user information needs and action on the web. The information scent is determined by the perception of the value and cost of the information with respect to the goal of the user. Even though the concept is appealing, Chi et al. (2001) make a big leap in arguing that using information scent it is possible to “analyse the needs of the users based on their surfing patterns” and “simulate usage of the Web based on user goals” (p 490). Pitkow (Catledge & Pitkow, 1995; Pitkow, 1999) successfully implemented an algorithm based on the mining of longest repeated sequences (LRS) to predict usage behaviour. This had obvious practical uses in recommender systems. The reader might be familiar with the Amazon website which dynamically provides interesting suggestion to logged in users based on their browsing and shopping histories. The original system used data mining on the raw logs for all users to find common patterns: this allowed to identify common sequences and was able to predict with good accuracy where a user would go next.

The most recent application of the concepts of scent and information foraging is the recent implementation of the SNIF-ACT model (Scent-based Navigation and Information Foraging in the ACT cognitive architecture - Fu & Pirolli 2007). As one may infer from the name, this is a mechanistic implementation of Pirolli’s cognitive model of user navigation strongly based on Anderson’s ACT-R. SNIF-ACT provides a rational account of how people use information scent cues in order to make decisions on what resources merit attention and leads to the accomplishment of their goals.

A text comprehension approach: CoLiDeS.

The two approaches to model behaviours in searching for information on a website used structural information and localised value attributed to information beacons based on a model of the user borrowed from foraging theory. Although this system might be useful to provide a mechanism to inform automated systems, it is unclear how the value of information is relevant to the individual for the task at hand without a coherent understanding of the content.

CoLiDeS, an acronym for *Co*mprehension-based *Li*nked model of *De*liberate *S*earch (Kitajima & Polson, 1995, 1997), was based on Kintsch's (1998) model of text comprehension, and at the core of the model is a realistic attention mechanism. Specifically cognitive in nature, CoLiDeS involves four essential cognitive processes: parsing, focusing on, comprehension and selection. The action planning processes involved in searching are guided by user's goals and ultimately determine the resources allocated in the four processes. This particularly interesting model tries to define a richer concept for relatedness than the scent used by the information foraging model. Three degrees of relatedness between items on the screen were suggested: a degree of similarity, the frequency with which the user has encountered an object, and whether the representation of the task goal has a literal match with the object.
Using cognitive walkthroughs, Kitajima and colleagues provided detailed scenarios in which a simple forward search (much like the pursuit of the scent trail in information foraging), fails by leading the user to an impasse.

Evaluation of the three models.

The MESA model evidently focused on the structure of the site: the user model is a mechanistic one, in which core perceptual and motor features are considered. The Information foraging models take a cognitive, information processing stance, and the integration with the ACT-R model is an exemplification. CoLiDeS offered a more in-depth, content-based approach to the search.

All three provide an excellent starting point to understand users' behaviours, but all put more emphasis on the structure and content rather than differences in the users, which limits the value of the models. It has been shown that even in simple tasks, and in ability tests presented in chapter 3, there is a variation in performance unaccounted by ability alone. The assumption that all users are equal and the tasks at hand are what cause differences is deceptive.

Based on the empirical data obtained from the use of online resources, we will reconsider the ideas presented in these models in chapter 9.

5.6.2. Behavioural sequences online

At the most basic level, whenever a user starts a request by either typing an internet address in their browser or by clicking on a link, these activities generate *transactions* between a *client* and a *server* that are often stored in web logs on the server. Such logs were originally intended for system administrator to monitor and debug possible issues related to the provision of the service. However, as also mentioned in chapter 2, these are now considered crucial in the model implementation of e-learning.

The sequences of actions performed by the user, or click-streams, provide a simple and unbiased record of users' behaviours. Servers, however, also store additional information associated with clicks (meta-data). This extra information is the most interesting feature considered in this thesis, but the one that makes handling and interpretation of the data stored a non-trivial problem, similar in magnitude to the one posed by large datasets which we referred to when we discussed data mining.

Data mining of web data featured originally in a paper in the mid nineties: Etzioni (1996) questioned if the web was at all suitable for data mining techniques. At the time, data mining was only beginning to be used extensively in database research and led some researchers to advocate a transformation of the web into a massive layered application, similar to databases, in which meta-information could be easily associated with content. This would have had the obvious advantage of providing a more efficient search and possibly speed up the communication and transactions. Etzioni, however, presented a number of prototype methodologies pioneering feasible examples of data mining for the web. He concluded that, in principle, these methods had a great potential to help people *search*, *navigate*, and *visualise* the content of the Web. Since then, the area of Web Mining has been a very fertile area of study mainly pushed by the exponential growth of e-commerce, which in the past 10 years allowed the convergence between a number of research communities such as in databases, information retrieval, machine learning, natural language processing, semantic structures and visualization. The strong relevance of this strand of research is apparent when considering the analysis and evaluation of online activity applied to education and online learning.

More recently, Romero and Ventura (2007) reviewed the application of data mining in the field of education over the decade 1995-2005: in their assessment they highlighted the relevance of this field of research as an upcoming area of investigation which requires more research. Presenting a number of cases and implementations, they were very clear about the contributions of the application of data mining techniques for e-learning, adaptive hypermedia, intelligent tutoring system and web recommendation agents. Web usage and tracking featured as the main source of data for the young field of educational data mining.

However, web usage and user tracking is perhaps the trickiest of all data examined thus far because of the lack of standardised data structure and methods of investigation, poor integration between mining tools and e-learning system and unfamiliarity of researchers in education with the specific data mining techniques suitable for the problems to be investigated.
In fact, even though the exact sequence of actions performed by a user/student is recorded by the system as is, the metadata attached to each action is determined by the system itself; this dictates the granularity of information (and analysis) which is available in the data structure. The practical implication is that the nature of the data obtained is changeable and the raw logs stored over the past 6 years required a considerable amount of manipulation to provide a coherent structure. Furthermore, as outlined in Hardy et al. (2005, 2006) and explained in more details in Appendix 1, the use of external tracking methods, attempted in 2005-06 (STEER project), does create an extra burden without achieving the desired outcomes.

Beside the technical difficulties posed by the task of accurately tracking user behaviours, in this thesis we had to create a coherent way of representing online usage with four different systems. Over time, two version of WebCT were used in the University-based VLE, a proprietary implementation of WebCT e-packs in connection with the Atkinson & Hildgard textbook (the course textbook in 2004-06, Introduction to psychology 3rd ed. - Thompson) and the self-contained proprietary implementation of the blackboard system offered with the Martin & Carson textbook (used in 2006-08, Psychology - Pearson). The main cause of the inconsistency for the university-based system (also reflected in the e-packs provided by Thompson) was an upgrade of the VLE system which drastically changed the way in which both the interaction and usage tracking was performed half way through this research. As mentioned earlier, the elected VLE for the courses is WebCT/Blackboard. Before autumn 2006 the installed version was WebCT CE 4. This was then upgraded to WebCT Vista 6, which brought a number of important changes and improvements in the way the VLE is serving content to the browser. WebCT CE allowed a very minimal tracking facility from the user-interface: it was possible only to identify general trends and some class statistics, but HTTP server logs could be extracted from the server as it was impossible to effectively create reports and drill-down to individual users' data from the default interface. In the later version standard server logs were not available as the database-driven backend recorded millions of unique transaction with automatically generated web pages. In simple terms this means that the system does not refer to a single page as such, but that the content unit served would be generated dynamically every time the reference (or link) was requested.

An alternative way of tracking users was attempted with the STEER project (Hardy et al. 2006), but the reliability of the data was questioned following a comparison with raw server logs. More details about the format of the logs is given in Appendix 1, but here it is essential to understand how log data was integrated into a more coherent set of *tools* and *actions*.

Beside the idiosyncratic differences of the various systems, it is easy to justify why such an abstraction is essential: between 2004 and 2009, each WebCT instance had between 500 and 3000 unique units of content available. Because of the changes made over time and the growth of the material offered, it would be impossible to compare directly the activity in different courses, or different instances of the same course at different points in time. The tables in the next pages provide a structural and semantic classification of possible activities abstracted from single content units and makes it possible to reduce the complexity of the behavioural analysis. Such elaboration allowed to make the observational data more systematic using a common denominator to objectively classify the log entries.

From the individual links (over 67000 from all courses between 2004 and 2009) we individuated 71 *actions* and 53 *tools*. Which can then be further reduced to 11 *abstractions* (representing intentions). The table in the next page (6.10) provides an overall picture of such categorization. The reduction to tools and actions was partially informed by the system structure and organization and partly by the way in which WebCT is serving content. Such categorization is semantic in nature rather than following a topological organization of the website graph (i.e. based on links between pages or the folder structure of the website).

The abstraction to intentions is derived uniquely from the nature of the content served. For example, if a student enters the site, and immediately navigates to the section containing lecture notes to find a specific file with lecture notes for day X and then logs out, there is little room for subjective interpretation about what their intention was.

The abstractions used at this stage follow a similar pattern to tools and actions with obvious categories based on content. Some could be condensed further: for example, the *self-directed learning* categories are separate. This was done intentionally to keep activities distinct. All involve an active pursuit of resources to support learning; however it is apparent that reading extra material for study purposes is very different from testing oneself by means of interactive activities such as quizzes and crossword puzzles.

The obvious advantage of such categorization is the simplification of the data structure, however, even by applying it to the data collected, each episode in which the user interacts with the system might be characterised by very different patterns and intentions.
Web usage mining is the methodology used to make sense of such differences and the particular techniques will be considered in the next section.

Core content	Information seeking	Learning activity	Social activities	Utility/tools
compiler	**Announcements**	**artifact**	**chat**	**browser check**
compile-lm	view-announcement	artifact-created	chat-room-entered	tool-activated
content-page	view-announcement-list	artifact-saved	UoE Toolbox	**Calculator**
content-page-viewed	**Calendar**	**assessment**	view_chatwb_log	tool-activated
home	view-calendar-entries	assessment-open-instructions	view-chatwb-lobby	**Calendar**
content-page-viewed	**Course information**	assessment-started	**discussion**	delete-calendar-entries
Learning objectives	content-page-viewed	assessment-submitted	assessment-deleted	private-entry-added
content-page-viewed	view-announcement-list	assessment-view-attempt	compiled-message-viewed	view-calendar
lecture material	**Course information (DRPS)**	assessment-view-list	discussion-home-viewed	**file-manager**
content-page-viewed	Access	assessment-view-timelog	message-deleted	artifact-created
organizer	**help**	**Assignments**	message-posted	**login**
organizer-viewed	content-page-viewed	artifact	message-read	login
practicals	**syllabus**	assignment-deleted	message-replied-to	logout
content-page-viewed	syllabus-viewed	assignment-read	topic-deleted	**Media-library-home-viewed**
Readings	**Tutors**	assignments-listed	topic-viewed	media-library-entry-viewed
content-page-viewed	content-page-viewed	**Assignments-listed**	**Group project**	**Notes**
tutorials		file-manager	artifact-created	notes-added
content-page-viewed		**dropbox_view**	**mail**	view-notes
		file-manager	delete-mail-message	**Password reminder**
	Information-seeking (Personal)	**file-manager**	folder-selected	Access
	my-grades	file-added	message-forwarded	**search**
	Assignment-reviewed	file-deleted	message-read	search-accessed
	view my-grades	file-manager_accessed	message-replied-to	search-performed
	view-grades	file-uploaded	message-sent	**UoE Toolbox**
	my-webct	**group presentation**	tool-activated	Access
	Assignment-reviewed	Assignment	**tracking**	**web-links**
	view-mywebct	**Submit work**	execute-report	url-viewed
		Assignment	wio-user_view	
		Tutors	**who-is-online**	
		do-marking	wio-user_view	

Self-directed Learning (Media)	Self-directed Learning (Quiz)	Self-directed Learning (Study)	Self-directed learning (WWW)
internet activities	**Crossword**	**Glossary**	**Connect to another web server**
url-viewed	assessment-started	Glossary	Access
Media-library	**MCQs**	**internet activities**	**Connect to UoE Wiki**
media-library-collection-viewed	assessment-started	notes	Access
media-library-entry-viewed		**stats exercises**	**Log into another web server**
Media-library-home-viewed		content-page-viewed	Access
scorm		**student-bookmarks**	**web-links**
scorm-module-deleted		bookmarks-listed	url-viewed
scorm-module-imported		**web-links**	Web-Links-home-viewed
scorm-module-viewed		url-viewed	

Table 5.8. Categories and sub-categories of activity: abstractions, tools and actions.

Acedmic year	2004/05	2005/06	2006/07	2007/08	2008/09
PSYCHOLOGY 1 Course	**PS001_7**	**PS001_8**	**PS0001_9**	**PS0001_10**	**PS0001_11**
Core content	61.03%	16.32%	47.90%	25.09%	19.49%
Information seeking	2.41%	0.18%	2.67%	2.36%	0.64%
Information-seeking (Personal)	0.20%	0.04%	0.13%	0.17%	0.16%
Learning activity	0.00%	0.53%	16.01%	2.97%	15.96%
Self-directed Learning (Media)	0.11%	0.00%	0.38%	0.87%	0.48%
Self-directed Learning (Quiz)	7.85%	1.43%	0.00%	0.00%	0.00%
Self-directed Learning (Study)	6.91%	0.84%	0.13%	0.09%	0.08%
Self-directed learning (WWW)	0.02%	0.01%	2.29%	1.92%	1.44%
Social activities	18.19%	0.61%	15.37%	46.42%	47.79%
Utility/tools	3.30%	80.05%	15.12%	20.10%	13.95%
Grand Total	100.00%	100.00%	100.00%	100.00%	100.00%
e-packs for PSYCHOLOGY 1	**zuPS001_7**	**zuPS001_8**	**zuPS0001_9**	***Pearson***	***Pearson***
Core content	40.11%	25.76%	8.53%		
Information seeking	0.04%	0.36%	0.15%		
Information-seeking (Personal)	2.95%	2.45%	0.15%		
Learning activity	0.00%	0.00%	86.47%		
Self-directed Learning (Media)	0.70%	0.52%	1.91%		
Self-directed Learning (Quiz)	48.10%	32.39%	0.00%		
Self-directed Learning (Study)	3.30%	3.50%	0.00%		
Self-directed learning (WWW)	0.50%	0.86%	0.00%		
Social activities	2.33%	3.48%	2.35%		
Utility/tools	1.98%	30.69%	0.44%		
Grand Total	100.00%	100.00%	100.00%		
PSYCHOLOGY 2 Course	**PS002_7**	**PS002_8**	**PS0002_9**	**PS0002_10**	**PS0002_12**
Core content	70.22%	10.54%	26.09%	39.12%	23.37%
Information seeking	0.00%	0.32%	1.98%	5.09%	0.96%
Information-seeking (Personal)	0.00%	11.17%	0.10%	0.35%	0.12%
Learning activity	0.00%	18.14%	25.57%	2.11%	22.29%
Self-directed Learning (Media)	0.00%	0.00%	0.00%	0.88%	0.36%
Self-directed Learning (Quiz)	7.26%	0.21%	0.00%	0.00%	0.00%
Self-directed Learning (Study)	0.00%	0.03%	0.10%	0.18%	0.00%
Self-directed learning (WWW)	0.00%	0.00%	1.14%	3.86%	2.65%
Social activities	0.00%	1.02%	34.93%	40.35%	45.18%
Utility/tools	22.52%	58.58%	10.08%	8.07%	5.06%
Grand Total	100.00%	100.00%	100.00%	100.00%	100.00%

Table 5.9. Prevalence of content available online (not usage) over the whole academic year in the different instances of the courses.

It should be noted that the Utility/tools category is skewing the distributions considerably in 2005-06 because all coursework was submitted via WebCT. This action is included into this category.

5.6.3. Web mining as data mining

Historically, mining the web had three different purposes mostly grounded in the tradition of usability research (Srivastava et al., 2000): mining the web *to find information* (search and organize the material are very much unresolved issues in the internet age), mining *to discover structural information* (i.e. the architecture and design of websites) and mining *to understand usage* (or how users interact with the medium). Even though we are more interested in the latter, advances in all these areas of research are useful to explain *how* a user is interacting with the web and understand *why* they are doing what they do.

Defining what *web mining* means is not as straightforward as it seems; this confusion has various causes ranging from inconsistent terminology in the different areas of research to simple disagreement about what the concept involves (Berendt, Hotho, & Stumme, 2002; Cooley, Mobasher, & Srivastava, 1997; Cooley, 2000; Mobasher et al., 2002; Spiliopoulou, 2000)

Before 1997 a number of papers were published to explore how hypermedia and hypertext were used. This had two main drives: a usability concern in which the non-linear organisation of a website was the most important feature compared with traditional text (a review of experimental research on hypertext and cognitive processes is available in Rouet et al. 1996). The second was a clear concern about how to index the growing amount of material and make it searchable. The PageRank algorithm developed by Page and colleagues (Page et al. 1998) is possibly the most well known system driving the Google search engine to these days. Data mining proved to be useful, but with the establishment of the WebKDD group web mining gained a status of its own. The reason for the schism from data mining research was that traditional data mining and machine learning predominantly deal with numbers, whilst data representation of internet documents is symbolic.

Kosala and Blockeel (Kosala & Blockeel, 2000) identified four specific sub-tasks involved in the web mining. These processes are similar to any other data mining task:

- finding resources: the process of searching and retrieving *intended* web documents;
- information selection and pre-processing: the automatic selection of specific information retrieved from the web;
- pattern discovery and generalization; automatically identify common patterns in individual or multiple websites;

- pattern analysis; concerned with the validation and interpretation of the patterns uncovered.

In more recent accounts (Chen & Liu, 2004) a distinction between web data mining, web structure mining and web usage mining has been consolidated.

	Web Mining			
	Web Content Mining		Web Structure Mining	Web Usage Mining
	IR View	DB View		
View of Data	Unstructured Semi structured	Semi structured Web site as DB	Links structure	Interactivity
Main Data	Text documents Hypertext documents	Hypertext documents	Links structure	Server logs Browser logs
Representation	Bag of words, n-grams Terms, phrases Concepts or ontology Relational	Edge-labeled graph (OEM) Relational	Graph	Relational table Graph
Method	TFIDF and variants Machine learning Statistical (including NLP)	Proprietary algorithms ILP (Modified) association rules	Proprietary algorithms	Machine Learning Statistical (Modified) association rules
Application Categories	Categorization Clustering Finding extraction rules Finding patterns in text User modeling	Finding frequent sub-structures Web site schema discovery	Categorization Clustering	Site construction, adaptation, and management Marketing User modeling

Table 5.10. Taxonomy of web mining (adapted from Kosala & Blockeel 2000). Information retrieval (IR) and database (DB) views of content mining are of more interest for designers/developers.

The table above (5.15) shows a possible taxonomy for web mining concepts/structures, with applied areas of study. Although this might look somewhat cryptic to the less-technical reader, we will address in more detail the techniques relevant to this thesis in a later section.

Baldi et al. (Baldi, Baralis, & Risso 2005) reviewed extensively the techniques applied to model both user behaviour and web structure. What is important to learn at this point is that from all the literature reviewed in IT and AI; it is evident that WUM is done with four concepts of *utility* in mind:

- improve design/architecture of websites (Mobasher et al. 2002);
- improve navigation from common usage path (Srivastava et al. 2000, Mobasher et al. 2002);
- inform navigation models to be used by learning algorithms and automated agents for web personalization (Koutri, Avouris, & Daskalaki, 2004);
- improve information retrieval for both user and agents (Witten, Bray, Mahoui, & Teahan, 1999);

For the purposes of this thesis, we will focus specifically on aspects of *web usage mining* in the table above (5.15) and *web structure mining* will be used to characterize and visualize the e-learning material. The latter is useful to provide utility metrics for navigation (i.e. a formal way to characterise what a user does in a visit to the site) and the former will be used to provide a better insight in content fruition (i.e. what does the user do and try to infer why). Understanding the patterns of usage over time, with multiple sessions and with different intentions for each session is a particularly challenging problem and will be the aim of the analysis in chapter 8. In this context, it is essential to highlight that treatment of the technical details will be kept to a minimum and (refer the Appendix 1).
The next section tackles only the fundamental aspects of the techniques used as the aim of the argument is to discuss how the applications of these methods provide interesting avenues to study behaviours rather than explore the technical aspects of the analysis.

Discovery of usage patterns from HTTP server logs

Three phases define the Web mining process and require some clarification; these are the major processes in manipulating the raw logs available to us:

- pre-processing;
- pattern discovery;
- pattern analysis.

The pre-processing phase

The pre-processing phase is the most important and most time-consuming step leading to the preparation of data for further processing. Five essential steps characterize this phase:

1. data cleaning;
2. user identification;
3. user session identification;
4. path completion (creating session files);
5. transaction identification (creating transaction files).

Cleaning up the raw logs is essential to make the logs information intelligible. Data on each transaction to the Web server is transcribed to very large comma-separated text files (the web logs). With each content page a number of associated files (i.e. picture and styles) are downloaded and transactions recorded in the log. However, only the actual content pages are of interest and the rest can be removed for the purpose of this analysis. Furthermore, only a few fields in the logs are relevant for discovering behavioural patterns (IP address, timestamp, user, the link and its referrer page).

In the log, each row contains either a user id, an IP address (an ID for the computer from which the request is originating) or both. Some heuristics need to be applied to ensure that both fields are complete: in normal scenarios a user id is not available and only the IP and timestamp are used to infer a user session. In our case all students are authenticated users of the system and their id is explicit.

Some rows of the log need to be filtered out: as well as those related to associated files which are not classed as content, cases of entries with the *HTTP status code*[11] equals to 404 (resource not found on the server) and similar. Such entries are very useful to system administrators to identify missing files or broken links, but have little bearing on the usage behaviours.

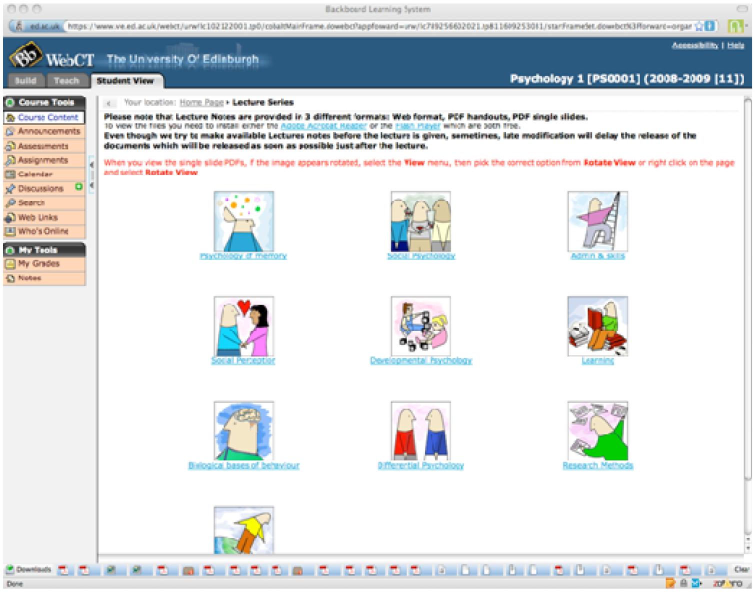


Figure 5.6. A typical WebCT organiser page.

This contains 3 frames: a top navigation pane with university logo, a left hand side menu with icons and a main page with more icons. Concurrent links in breadcrumb, navigation bar and content page are available. The request for this resource would log about 30 requests: the download of 25 images, 3 content html pages a style file and the scripting support file.

Once all users/IPs combinations are identified, it is necessary to work out two core resources: a user *session* file and a *transaction* file. As the cleaned log file ordered by identified users

[11] The Status code in the log is explained in Appendix 1: it refers to each transaction to identify if problems occurred.

and time of request could contain series of requests over long periods of time by the same users, it becomes important to *divide* the log entries of the same users in sessions, or *visits*. The session file contains a list of unique visits to the website. For what concerns visits, it is customary to use a 30 minutes *timeout* to identify different sessions. Timeout means that if users have dwell times longer than 30 minutes on a single page, it is possible that they have been doing something else in the meantime (but there is no way of confirming this), therefore sessions are cut at this point. When activities recommence it counts as a new session. Like when trimming reaction times, this step ensure that unusually long sequences are removed from the data as outliers.

The transaction file contains a list of links pairs and links sequences; this is a fairly complex procedure based on two core references: *path completion* and *transaction identification*. Path completion is easy when both the link and the referrer -the page containing the link to the second page- are present in the log. Unfortunately, for a number of reasons (see Appendix 1 for details) often the referrer needs to be inferred. When the site structure is simple, it is possible to determine the referrer from the previous line of the log, however the task becomes impossible when there are multiple links and frames (which are in themselves different documents) available in the web browser at any given time. In our specific case, this task was particularly difficult with logs produced by WebCT 4, but much easier with the later logs.

Pattern discovery.

The identification of transactions varies depending on the Web Usage Mining technique that one plans to use. The key purpose of this phase is to organize the data in such a way that structural patterns become evident.
Cooley et al. (1997) indicated that “the goal of transaction identification is to create meaningful clusters of references for each user”. There are three different methods commonly used which lead to different results:

- time window;
- reference length;
- maximal forward reference.

Time window, popularized by Cooley et al. (1997) is a procedure that divides the log of user entries into specific intervals using a predefined parameter.
The *reference length* and *maximal forward reference* methods allow the web miner algorithm to identify transactions, assigning meta-information regarding the navigational purposes of a

user. In its simplest form, Cooley et al. (1997, Cooley, 2000) argued that for an Internet user a web resource can be classified as either an *auxiliary* (or *navigation)* resource or a *content* resource. An *auxiliary resource* represents an intermediate step for the user in order to get to the *content* resource. Such meta-information is provided using either the topological structure of the website (when this is straightforward), inferred from the semantic analysis of the content, as in the case of user modelling (i.e. in CoLiDeS, Kitajima et al.) or, like in our case, each unique link can be tagged manually (see for example table 6.10), based on specific information about each page.

The *reference length* method is the time difference occurring between two transactions made by the same user on the same server (usually two adjacent rows in a user session file). This represents the estimated amount of time (dwell time) the user spends on a resource. When using the simple distinction between content and auxiliary resources it is common to infer the resource type from the time spent on a resource: if this is *long enough*, the resource is considered as content. Cooley et al. (1997) defined this parameter as an arbitrary cut off time suggesting different suitable thresholds for different types of websites. This is considered too vague here: instead, the meta-data afforded by the categorization presented in table 6.10 makes this metric redundant as the specification of each page is not automatically inferred by the miner.

More interesting for our data is the identification of the maximal forward reference, which requires the creation of full sequences from the sessions. As the users can navigate both backward and forward within the structure of the website, a *backward* reference is the revisit of previously visited resource and a *forward* reference is the visit of a new resource in the user session path. According to Chen et al. (Chen & Liu, 2008) when a backward references occur, a forward reference path terminates. The resulting forward reference path is what determines a *maximal forward reference.*
To clarify, let's suppose to have the following sequence from a user session (or visit):

{A, B, C, D, C, B, E, G, H, G, W, A, O, U, O, V}

From the application of the MF (Maximal Forward) algorithm, the resulting set of maximal forward reference transactions is made of:

{ABC**D**, ABEG**H**, ABEG**W**, AO**U**, AO**V**}.

The interpretation of users' behaviours become complex here as what the user does when s/he backtracks to C or B (at the start of the sequence) before going to E or G is difficult without knowing the goal of the user at any given time. A *maximal forward reference* could also be considered as the final *destinations* of each temporary navigation path, that is to say the *content* resource, whereas the previous ones in the sequence could be seen as *navigation* or *auxiliary* resources.
Based on the assumption that the last item of each sequence is content, another version of the MF algorithm would consider only the final item in each subset as relevant. In the example used above, the resulting cluster with content-only resources would be:

{**DHWUV**}

This method ignores completely the dwell times between resources and it is evident that it would not be suitable in Web sites with a dense content or high connectivity between resources which is the case for e-learning websites. Nonetheless we will look in more details to the sequences using both the dwell time and the meta-information used to tag each item in the clickstream to abstract the intentions of the user in a given session.

The repetition or revisitations of pages can however have other interpretations. For example, Sheard (Markham et al., 2003; Sheard et al., 2003) looked at revisitation patterns as a measure of disorientation in the web structure. Using a more stringent experimental context, within a psychological framework of research investigating serial behaviours, McGonigle and colleagues (McGonigle & Chalmers, 2007, 2008) considered repetitions or revisitation in a search sequence as an indicator of a less efficient executive control. Their research with both non-primate monkeys and very young children demonstrated that self-emergent strategies based on core features of the items set make the search not only a tractable problem, but very efficiently organises the search. We will use a similar metric to evaluate sessions for each user in chapter eight.

5.7.4. Data visualization

Another way of identifying patterns is to use a visual representation of the data. In the case of the literature review in chapter 4 we used semantic and heat maps to identify the similarity of each record in the data. In the case of web usage mining (WUM), without any doubt, this is a very difficult enterprise. Whilst visualization of single sessions is a trivial problem, and a number of techniques are available (see example below to see the path and time spent in a single session), the representation of multiple visits or patterns of visits over time constitutes a very difficult, largely unresolved problem. In chapter 8 we will explore some possible methods in the attempt to discover useful patterns through visualization, of which a couple of examples are reproduced in Figure 5.7. Each one of the two examples depicts the navigation path with nodes as pages and links as transitions numbered to indicate the order of actions.

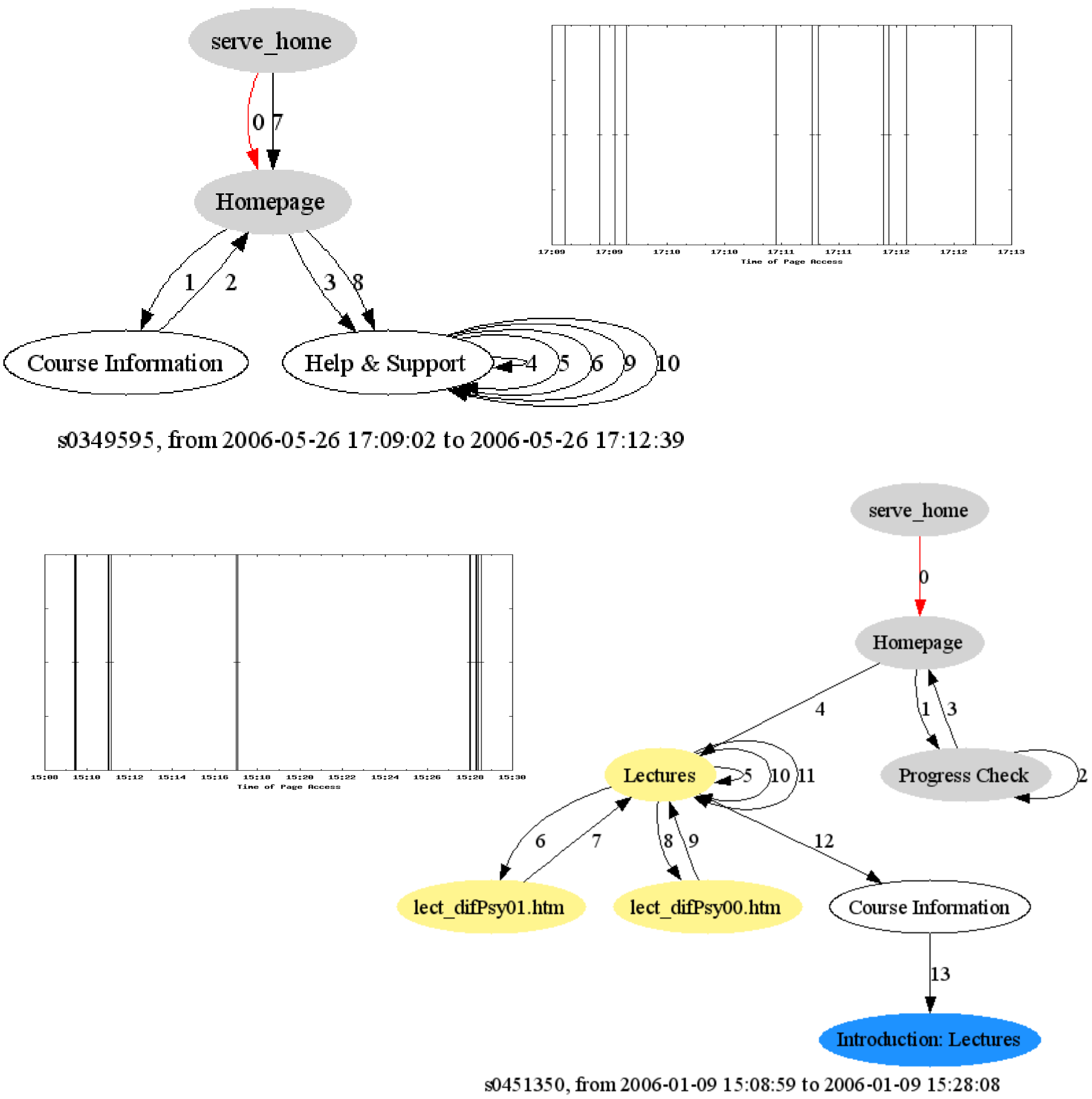


Figure 5.7. Two examples of a possible representation for the sequences of actions performed by a user during a visit. Two different sessions and users are presented.

The second type of graph represents the time sequence (about 3 minutes in case for the top example and 20 minutes for the one at the bottom) in which each click is a dark line with spaces between the lines showing the dwell times. The total number of clicks is 10 in the top case and 13 in the second, but the patterns of navigation and dwell times are very different in both depth of visitation and goals. The examples used here demonstrate the wide range of possible characterizations of sequences with similar length.

5.7.5. Pattern analysis

As we have already considered, patterns can emerge from the mathematical or algorithmic emergent features of data or the visual representation. The final step of the web usage mining process is the analysis of the patterns discovered. Witten and Frank (2005) enlist five possible families of methods of analysis:

1. Statistical modelling (usually from linear or regression models);
2. Decision trees/classification algorithms;
3. Association rules;
4. Clustering;
5. Instance learning and Bayesian networks.

Depending on the hypotheses tested, the data structure and the goal of the analysis, one or more different methods are suitable to solve specific problems. In particular the first three methods presuppose that some sort of prediction is made from a number of variables: this makes them more suitable to investigate the relations between usage and levels of AP.

The other two are more suited to explore hidden patterns in the data, and in many respects can be compared to data reduction techniques such as factor analysis which is very familiar psychology researchers. We will consider these techniques to explore online behaviours and their relations with other variables in more detail in chapter 8.
Both sessions and users behaviour will be explored using a combination of data mining techniques.

5.6. Chapter summary

In this chapter we outlined the setting of the research conducted and reviewed the overarching methodology to collect and analyse data.

We considered in turn the type of data, the metrics and tools and the procedures applied to organise and structure data. In particular, we individuated three core sources of data and categorised them as: academic performance, measures of styles and individual differences and behavioural data.

After giving more details about the styles measures, we briefly reviewed some basic concepts relevant for the application of data mining techniques on the behavioural data stored in web logs. This provides the basic understanding to explain the initial categorization of activity and the data mining procedures which will be applied in later chapters.

As each type of data has its own strengths and limitations, the synthesis of metrics proposed in this thesis is expected to produce a unique insight into the relations between individual differences in learning and the use of e-learning material.

In the next three chapters, we delve deeper in the data and study academic performance, the individual differences in performance (from a demographic point of view in chapter 6) and styles (chapter 7), then consider the relations between instruments. In chapter 8, we examine the patterns of online usage and relate them back to academic performance and stylistic differences.

Chapter 6. A characterisation of academic performance in the sample

After reviewing the theory, the context and presenting a number of hypotheses, the next three chapters are dedicated to the study of the three core elements in our database: academic performance, styles and usage.

The aim of this chapter is to provide an overview of the sample studied, detailing a number of features. Starting with the demographic characteristics of this group of students, we then consider their enrolment data which provides information about their degree paths, the composition of the cohorts and their overall performance, particularly for those who are already beyond their first year of study.

Where available we will look at prior performance and for the first time we will introduce a data mining technique applied to provide a better insight on how prior performance can be categorised and related to later grades at university. By the end of the presentation, it will be clear not only how students perform, but how their history of achievements might have influenced the paths they undertook.

6.1. Characterising the samples

To better understand the nature of the samples used in this thesis, it is necessary to provide more information about *who* the students are, what their abilities are and define what direction they take in their degree paths. One of the frequent criticisms of research conducted in psychology is that students are often used as *guinea pigs* for experimentation and the generalisation of the findings suffers from the limitations of the samples used. This section has two key purposes: an in-depth characterisation of the composition of the samples used, and providing evidence that even if the participants are psychology students, they are still representative of a much wider stratum of the population.

First year cohorts 2002-08																	
		N	%	sex	N	%	age	N	%	origin	N	%	title sought	N	%	degree subject	N	%
2002	**tot students**	270		males	94	34.8	17-18	104	38.5	UK	237	87.8	MA	184	68.1	Psychology	60	22.2
				females	176	65.2	19-20	135	50.0	EU	15	5.6	BA	0	0.0	Biology (Psy)	0	0.0
							21-25	20	7.4	World	18	6.7	BSC	80	29.6	Psychology with	50	18.5
							over 25	11	4.1				other	6	2.2	other HSS	72	26.7
	missing entry data	51	18.9													other Sciences	88	32.6
				tot		100	tot	270	100	tot	270	100	tot	270	100	tot	270	100
2003	**tot students**	320		males	87	27.2	17-18	137	42.8	UK	278	86.9	MA	219	68.4	Psychology	105	32.8
				females	233	72.8	19-20	141	44.1	EU	29	9.1	BA	0	0.0	Biology (Psy)	0	0.0
							21-25	24	7.5	World	13	4.1	BSC	96	30.0	Psychology with	66	20.6
							over 25	18	5.6				other	5	1.6	other HSS	61	19.1
	missing entry data	76	23.8													other Sciences	88	27.5
				tot		100	tot	320	100	tot	320	100	tot	320	100	tot	320	100
2004	**tot students**	292		males	79	27.1	17-18	140	47.9	UK	264	90.4	MA	216	74.0	Psychology	132	45.2
				females	213	72.9	19-20	119	40.8	EU	16	5.5	BA	1	0.0	Biology (Psy)	41	14.0
							21-25	18	6.2	World	12	4.1	BSC	75	25.7	Psychology with	55	18.8
							over 25	15	5.1				other	0	0.0	other HSS	31	10.6
	missing entry data	70	24													other Sciences	33	11.3
				tot		100	tot	292	100	tot	292	100	tot	292	99.7	tot	292	100
2005	**tot students**	293		males	109	37.2	17-18	139	47.4	UK	265	90.4	MA	204	69.6	Psychology	81	27.6
				females	184	62.8	19-20	115	39.2	EU	10	3.4	BA	4	1.4	Biology (Psy)	35	11.9
							21-25	17	5.8	World	18	6.1	BSC	83	28.3	Psychology with	52	17.7
							over 25	22	7.5				other	2	0.7	other HSS	67	22.9
	missing entry data	58	19.8													other Sciences	58	19.8
				tot		100	tot	293	100	tot	293	100	tot	293	100	tot	293	100
2006	**tot students**	253		males	84	33.2	17-18	106	41.9	UK	215	85.0	MA	169	66.8	Psychology	55	21.7
				females	169	66.8	19-20	114	45.1	EU	14	5.5	BA	7	2.8	Biology (Psy)	27	10.7
							21-25	24	9.5	World	24	9.5	BSC	73	28.9	Psychology with	42	16.6
							over 25	9	3.6				other	4	1.6	other HSS	73	28.9
	missing entry data	58	22.9													other Sciences	56	22.1
				tot		100	tot	253	100	tot	253	100	tot	253	100	tot	253	100
2007	**tot students**	287		males	93	32.4	17-18	134	46.7	UK	246	85.7	MA	194	67.6	Psychology	52	18.1
				females	194	67.6	19-20	130	45.3	EU	26	9.1	BA	4	1.4	Biology (Psy)	33	11.5
							21-25	12	4.2	World	15	5.2	BSC	82	28.6	Psychology with	38	13.2
							over 25	11	3.8				other	7	2.4	other HSS	98	34.1
	missing entry data	62	21.6													other Sciences	66	23.0
				tot		100	tot	287	100	tot	287	100	tot	287	100	tot	287	100
2008	**tot students**	307		males	78	25.4	17-18	162	52.8	UK	255	83.1	MA	222	72.3	Psychology	90	29.3
				females	229	74.6	19-20	123	40.1	EU	20	6.5	BA	2	0.7	Biology (Psy)	20	6.5
							21-25	11	3.6	World	32	10.4	BSC	66	21.5	Psychology with	55	17.9
							over 25	11	3.6				other	17	5.5	other HSS	75	24.4
	missing entry data	80	26.1													other Sciences	67	21.8
				tot		100	tot	307	100	tot	307	100	tot	307	100	tot	307	100
TOT	**tot students**	2022		males	624	30.9	17-18	922	45.6	UK	1760	87.0	MA	1408	69.6	Psychology	575	28.4
				females	1398	69.1	19-20	877	43.4	EU	130	6.4	BA	18	0.9	Biology (Psy)	156	7.7
							21-25	126	6.2	World	132	6.5	BSC	555	27.4	Psychology with…	358	17.7
							over 25	97	4.8				other	41	2.0	other HSS	477	23.6
	missing entry data	375	18.5													other Sciences	456	22.6
				tot		100	tot	2022	100.0	tot	2022	100	tot	2022	100	tot	2022	100

Table 6.1. Distribution of students taking psychology between 2002-2008.

Data organised by year, gender, origin and degree types at the time of registration; overall missing data for entry grades is also included.

6.1.1. Students' demographic composition

Earlier in the last chapter (table 5.4) we identified 6 different cohorts of students in 22 classes over a 7 year period. The dataset is quite large and a composition of the different cohorts at the start of each year is provided in table 6.1. This gives an overview of the gender distributions, age groups, origin of the students and shows the heterogeneity of the degree programmes sought by our students. As expected in psychology courses, there is a majority of female students with a ratio of about 2:1 (average females 69%) which is slightly below the figure in the 2004 BPS annual report of about 78%.

Age groups are also consistent with the general trend in other UK universities: at the start of their degree programme, the great majority of students are aged between 19-20 years old with only a small number of 'mature students' aged of 25 and over.
The average age of students starting the degree has also changed very little over the years; apart for the slight difference in the ages of males enrolled in 2002 and 2006 (differences were statistically not significant).

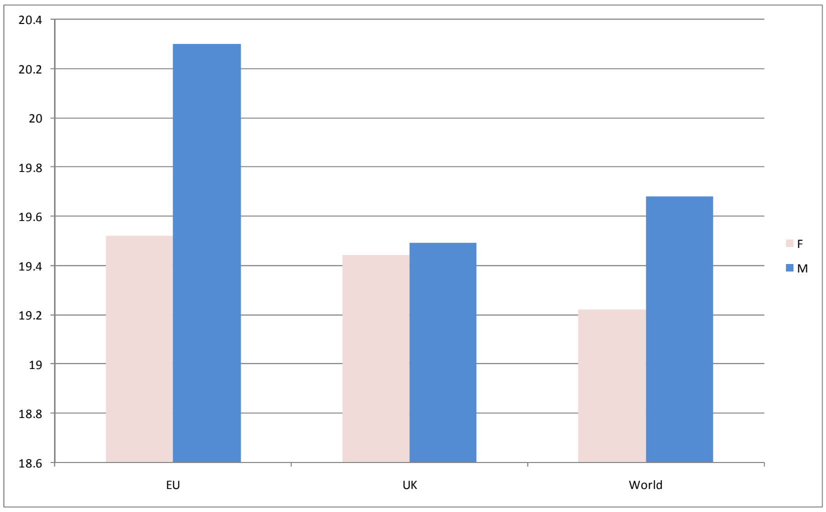


Figure 6.1. Average age at the start of the degree by gender and origin.

Although a specific ethnic analysis was not possible from our data, similar trends to those published in the same BPS report were substantiated in the course at Edinburgh. In fact, the country of origin was available for every student. Figures 6.1 and 6.2 depict the combinations of gender, country of origin and starting year, but it should be noted that neither factor (gender or origin) has a significant statistical bearing on the average age at start.

Because of the nature of the degree programmes at Edinburgh, where a lot of students taking psychology as an outside subject, we have provided figures for students who intend to study psychology as their main subject from year one and those who takes psychology in year 1 and year 2 as an 'outside' subject.

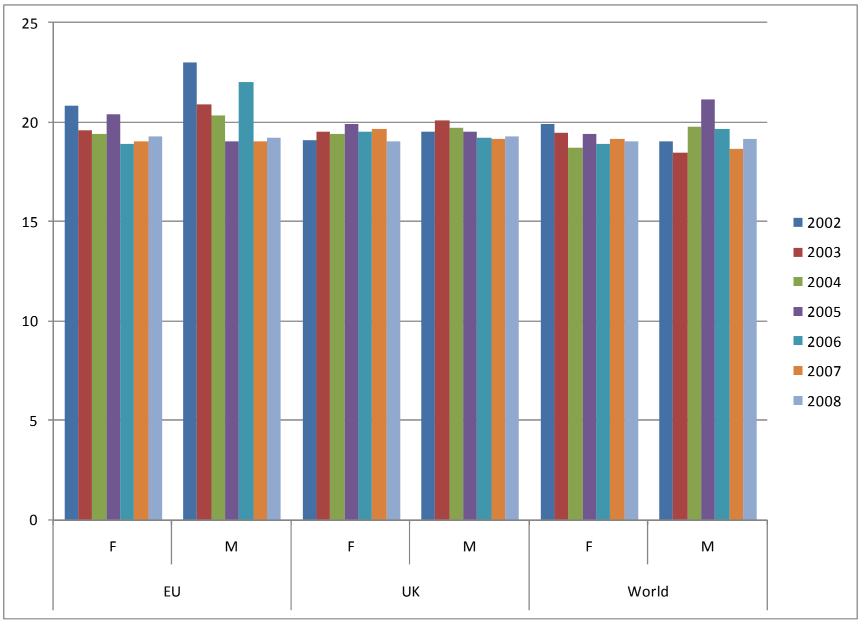


Figure 6.2. Average age at the start of the degree by gender and origin over the years.

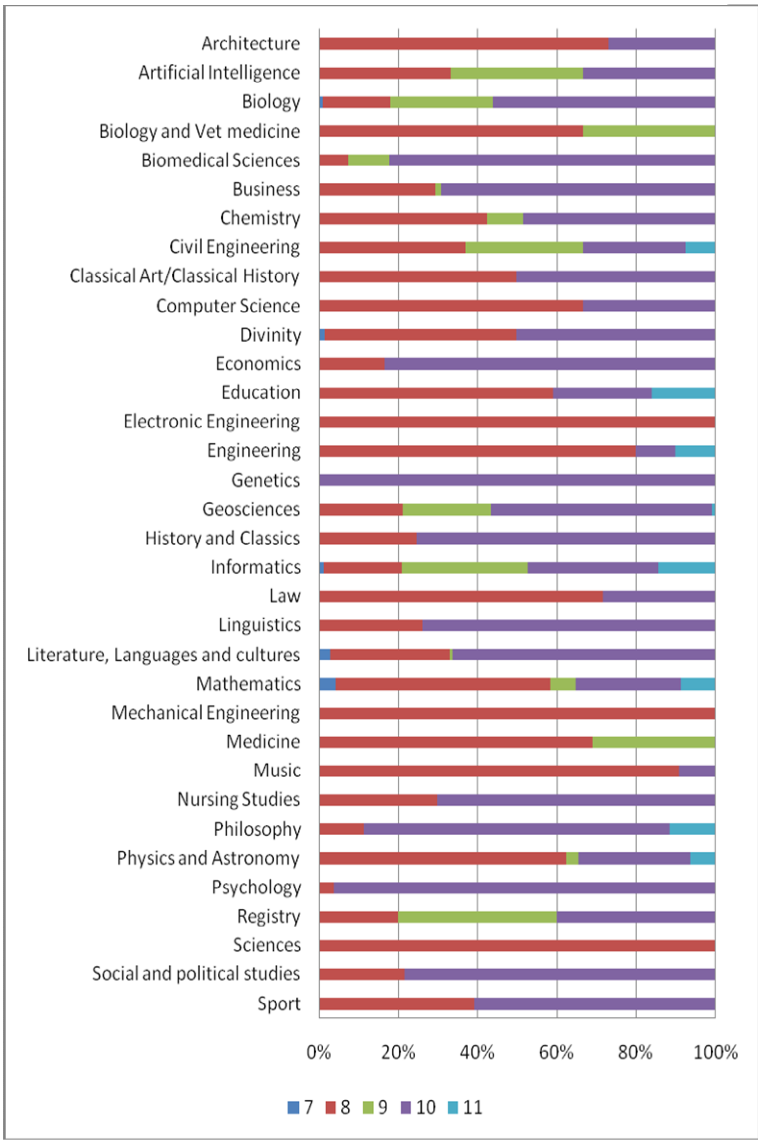
The summary table (6.1) shows that a considerable percentage of students take psychology as an outside subject (about 46%). It will be interesting to examine the various groups in parallel for both performance and usage as the motivation for taking the course is considerably different and the outcomes might also be quite different. In this respect, it is possible that students choosing psychology as their main subject area tend to perform better and will improve during the years, refining their skills and expertise in psychology, whilst students taking it as an outside subject show different motivations and performance levels. It is also possible that some might change their degree subject to psychology after their first experience. Both questions will be addressed in the later sections.

6.1.2. An overview of the courses taken by psychology students

As already noted, the students taking the psychology courses in year 1 and 2 are an extremely heterogeneous group. To provide a more detailed picture it is useful to identify not only what courses students take as part of their degrees, but more importantly, to give some details of the selection available to them. In theory, at pre-honours level, students could choose any course offered within the University at a given level. Obviously, the choice is constrained by timetable clashes, and careful advice discussed with director of studies who normally dissuade students from particular ‘risky’ combinations, however, the choice is still very wide and full details can be found in the publicly available reference (http://drps.ed.ac.uk). Nevertheless, a narrower view emerges from analysis of the courses that students have been taking in the sample between 2002 and 2009.

College	% subject area	of which:
HSS	71.87	Psychology MAs (33.8)
MVM	0.06	
SCE	28.07	Psychology BSc (8.97)

Table 6.2. Distribution of students taking psychology from the three main areas (Human & Social sciences- HSS, Medicine and Vet. Medicine –MVM, and Sciences and Engineering –SCE). In brackets, the percentage of the class registered for either the MA or BSc degrees in psychology. Note that the very low number of students from MVM take an intercalated honours year.



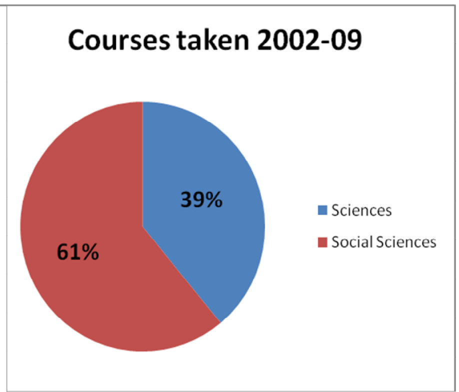


Figure 6.3. Distribution of the composition of all courses by area and SCQF level, taken by students who read psychology at any point during their degrees between 2002-09.

Figure 6.3 provides a detailed breakdown of the types of courses divided by disciplines and relative levels taken by students who at some point in their degree read psychology. In the sample considered (N=2467 active students) all students took at least the psychology 1 or psychology 2 course. The most interesting aspect of this distribution is that in psychology both courses at pre-honours level (year 1 and 2) are classed at level 8 of the SCQF; the variety of options increases drastically at honours level, but all the modules are offered at level 9 of the SCQF.

The second interesting observation is the broad spectrum of courses taken through their degrees, which reflect the heterogeneity of the cohorts in the psychology 1 course.

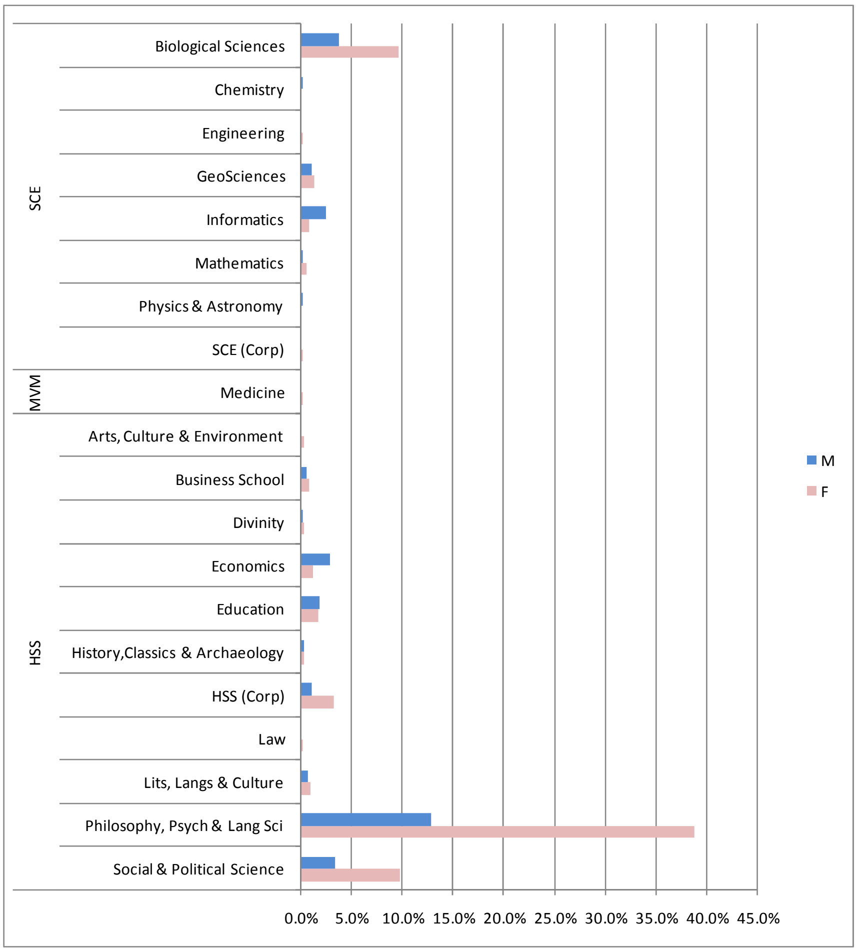


Figure 6.4. Prevalence of graduating students by gender according to College (Science and Engineering –SCE, Medicine and Vet. Medicine –MVM, and Human and Social sciences –HSS) and the respective Schools. [pivot from performance_comboV2]

Independently from their degree subjects, it is evident that the distribution of courses taken by students who register to read psychology 1 or psychology 2 is skewed toward courses in the social sciences (table 6.2). However, it is interesting to note that the variety of students from sciences degree who decide to take psychology1 or psychology 2 is about 20%. This is quite remarkable considering that overall, about 43% of the sample are registered for a psychology degree in one of its flavours (MA, BSc and joint degrees). We will investigate in the next chapter whether there might be interesting differences in performance for these sub-samples.

Following from figure 6.4 above, we can look at the graduates of the past few years from the different subject areas and colleges who took psychology at some point in their degrees. We can see that the proportion of students taking psychology from other subject areas compared to psychology degrees (both single and joint) is indeed quite varied and the gender distribution for the various subjects is quite predictable. In this distribution we ensured that degrees are associated with the respective Schools (i.e. Sociology and Psychology *belongs* to the School of Social and Political Sciences)

The overall depiction of the courses taken by students in our sample and the type of degrees obtained are insightful in determining both predictable patterns (i.e. the gender distributions in the subject areas) and surprising ones (i.e. the fact that about 40% of students from areas other than psychology willingly decide to take the courses).

Knowing that so many take psychology as an 'outside' subject allows us to test specific hypotheses about students' motivations and abilities. For example are students from sciences taking psychology as it is considered an 'easier' subject or are they stronger students anyway or is lower motivation leading to lower grades?

6.1.3. Students in depth: from entry grades to graduation

Earlier we indicated that prior academic achievement is normally a good predictor of performance at University. Satisfactory grades obtained at A Levels or Highers are essential to obtain an offer from most Universities in the UK. In the case of Psychology at the University of Edinburgh, the prospectus makes it very clear that the psychology degrees are quite popular and entry requirements demanding.

Table 6.16 provides an overview of the minimum requirements and *typical* entry grades for the various degrees advertised. There is little doubt that this University has a good reputation and rates highly in league tables, making it a very desirable plcaement for many students. From personal communications with the admission officer in charge for the psychology degrees, it emerged that the ratio of applicants every year is between 10 and 15 to 1 (there are about 300 places available), making the minimum advertised requirement much lower than the *real* typical figures.

Degree title	Required Qualifications	Typical
Psychology (MA)	SQA Highers: BBBB (or more if two sittings). with Standard Grade Mathematics at Grade 1 or Mathematics and Physics at Grade 2 or Mathematics Higher at Grade C.	AABB
	GCE A Levels: BBB. with GCSE Mathematics at Grade A or Mathematics and Physics/Double Science at Grade B.	AAA
Psychology (BSc)	SQA Highers: BBBB (or more if two sittings). Two of Biology, Chemistry, Mathematics, Physics Higher at Grade B: with Standard Grade Mathematics at Grade 1 or Mathematics and Physics at Grade 2. Chemistry at Grade 2.	AABB
	GCE A Levels: BBB. Two of Biology, Chemistry, Mathematics, Physics A Level at Grade B: with GCSE Mathematics at Grade A or GCSE Mathematics and Physics/Double Science at Grade B. GCSE Chemistry/Double Science at Grade B if no Chemistry at A Level.	AAA
Artificial Intelligence & Psychology (BSc)	SQA Highers: ABBB (or more if two sittings). to include Mathematics. Sixth-year work in Mathematics is recommended.	AABB
	GCE A Levels: ABB (or more if two sittings). to include Mathematics; or ABB to include one of Biology, Chemistry or Physics plus AS-level Mathematics at Grade A. (the science subject at A level Grade B may be Psychology.)	AAB
Cognitive Science (MA)	SQA Highers: ABBB (or more if two sittings). to include Mathematics. Sixth-year work in Mathematics is recommended.	AABB
	GCE A Levels: ABB (or more if two sittings). to include Mathematics; or ABB to include one of Biology, Chemistry or Physics plus AS-level Mathematics at Grade A.	AAB
Mind & Language (MA)	SQA Highers: BBBB (or more if two sittings). Standard Grade Mathematics at Grade 1 or Mathematics and Physics at Grade 2 or Mathematics Higher at Grade C.	AABB
	GCE A Levels: BBB. GCSE Mathematics at Grade A or GCSE Mathematics and Physics/Double Science at Grade B.	AAA
Philosophy & Psychology (MA)	SQA Highers: BBBB (or more if two sittings). Standard Grade Mathematics at Grade 1 or Mathematics and Physics at Grade 2 or Mathematics Higher at Grade C.	AABB
	GCE A Levels: BBB. GCSE Mathematics at Grade A or Mathematics and Physics at Grade B.	AAA
Psychology & Business (MA)	SQA Highers: BBBB (or more if two sittings). with Standard Grade Mathematics at Grade 1 or Mathematics and Physics at Grade 2 or Mathematics Higher at Grade C.	AABB
	GCE A Levels: BBB. with GCSE Mathematics at Grade A or Mathematics and Physics/Double Science at Grade B.	AAA
Psychology & Linguistics (MA)	SQA Highers: BBBB (or more if two sittings). with Standard Grade Mathematics at Grade 1 or Mathematics and Physics at Grade 2 or Mathematics Higher at Grade C.	AABB
	GCE A Levels: BBB. with GCSE Mathematics at Grade A or Mathematics and Physics/Double Science at Grade B.	AAA
Sociology & Psychology (MA)	SQA Highers: BBBB (or more if two sittings). with standard Grade Mathematics at Grade 1 or Mathematics and Physics at Grade 2 or Mathematics Higher at Grade C.	AABB
	GCE A Levels: BBB. with GCSE Mathematics at Grade A or Mathematics and Physics/Double Science at Grade B.	AAA

Table 6.3. Required qualifications and typical entry grades for students who were offered a place in the Psychology and related courses.

I was quite interested in finding out what are the *realistic* entry grades necessary to obtain an offer for the Psychology courses at this University. This proved to be impossible: in fact, because data from UCAS and data for enrolled students are kept separate for obvious reasons of confidentiality, I was only able to obtain entry grades for students officially matriculated leading to the typical entry grades for the various degrees reported in table 6.3. The following figures are computed with the awareness that there is missing data for about 20% of the matriculated students, making it only a partial picture.

	Females			Males			
	mean	SD	median	mean	SD	median	max
A grades	2.5	2.2	2.0	2.3	2.2	2.0	11
B grades	1.7	1.7	1.0	1.7	1.7	1.0	8
C grades	0.8	1.3	0.0	0.8	1.2	0.0	8
D or less	0.1	1.3	0.0	0.8	1.2	0.0	5

Table 6.4. Overall features of entry grades for the sample 2002-08.

Table 6.4, shows that a median of 2 As and 2 Bs is a more realistic minimum entry prospect for applicants, with some exceptional cases applying with up to 11 As in their records. It is possible that these figures are still short of the real minimum requirements or slightly inflated due to the missing data, but it is important to highlight that overall, the students considered in this thesis have an excellent record of prior performance.

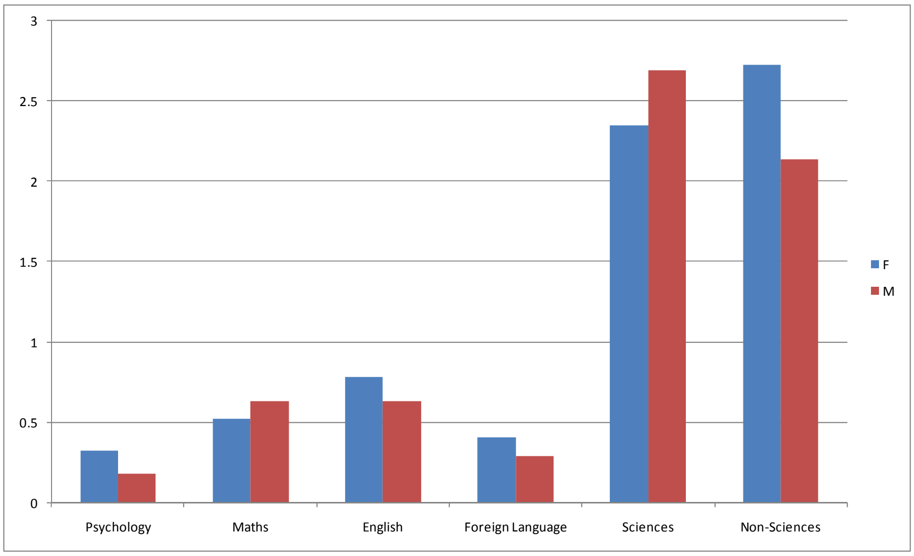


Figure 6.5. Average number of recorded As (entry grades) by gender and subject..

At the other end of the evaluation, looking at the end of the degree (end point) and consider the patterns of termination for students enrolling into psychology courses, it is evident that retention rates are not a major issue compared to other UK Institutions[12]. Of the 2467 active students between 2002-09, only 265 (10.7%) dropped out. Table 6.18 provides a breakdown of the possible recorded reasons for withdrawing from University.

---

[12] As a reference, according to HESA, in 2009 the average dropout rate after the first year of study is 9%. The University of Edinburgh recorded the lowest dropout rate in Scotland 4.2% and the figures reported of 10.7& are including all students abandoning or leaving during the entirety of the expected 4-years period.

Reason for withdrawing	Percent
Academic	3.02%
Financial Reasons	1.89%
Gone into employment	0.38%
Health/Medical Reasons	5.66%
Lapse of time so written off	30.57%
Other Reasons	11.32%
Personal Reasons	26.42%
Returning to a new programme of study	3.77%
Transfer to another institution	10.57%
Unknown reason	6.42%
TOT N	**265**

Table 6.5. Distribution of the reason for withdrawing from the University

Nonetheless, it is very important to explore in more detail *why* and *how* students with extremely good records of performance prior to university entry might fail or under-perform at University level. Figures alone are not enough to give answers. There is likely to be a wide range of external factors that could explain poor outcomes. In fact, as highlighted in the table above, many of these factors are not explicitly accounted for as metrics in this research. However, we can explore some of the reasons for under-performance using personality metrics and the behavioural data recorded.

We mentioned earlier the facts that about 40% of students reading psychology 1 take this course as an outside subject, and a the small proportion of students who withdraw from university at some point during their degrees. An interesting view of the various cohorts is the survival rate of students irrespective of the reason for not progressing from one year to the next (i.e. caused by a change of degree or the fact that they took only one psychology course as an ‘outside subject’ are not differentiated).

The survival plots identify that, on average, only about 40% of the students starting in year 1 continue to pursue a degree in psychology. This is an extremely interesting prospect from the pedagogical point of view when designing the introductory courses: the interest and motivation of these starting students might be diametrically different.

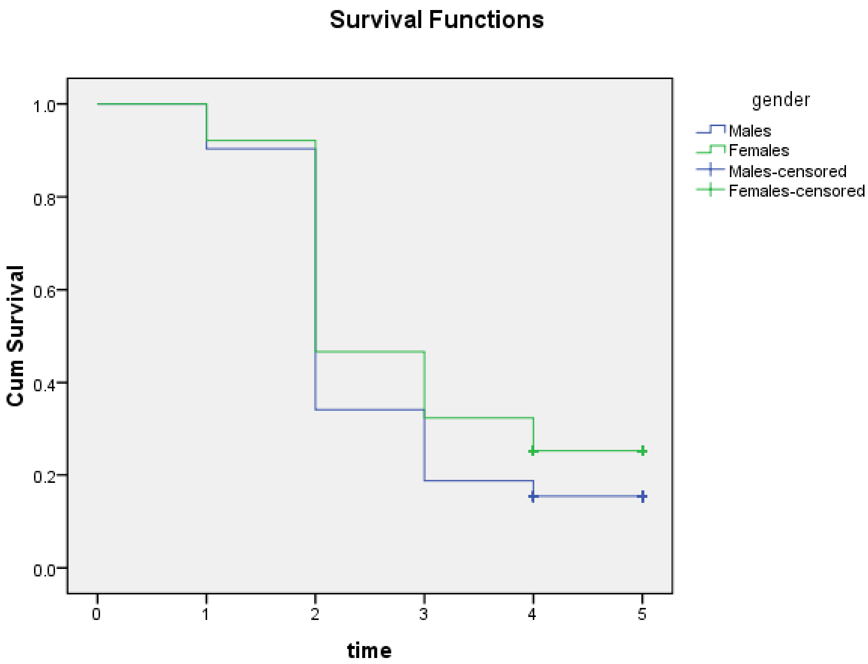


Figure 6.6. Survival rates of students taking psychology courses divided by gender: time point 0 is the start of the degree, 1 the end of year 1 up to 5 the end of the 4-years degree.

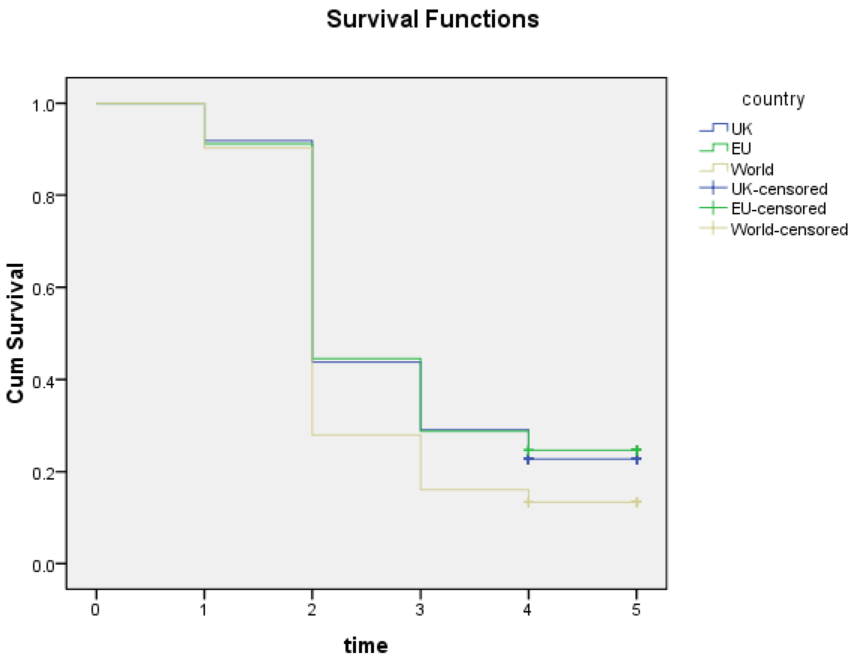


Figure 6.7. . Survival rates of students taking psychology courses divided by country of origin: time point 0 is the start of the degree, 1 the end of year 1 up to 5 the end of the 4-years degree.

Figures 6.6 and 6.7 illustrate the drop in the number of students across the duration of the degree and differentiates between genders and country of origin. Considering the time point zero as the matriculation point, each time interval represents the end of each academic year. Figure 6.8 gives a further insight in the distributions of students continuing their psychology degrees: there is an higher probability that students enrolled for a joint degree in Psychology with another subject are the most likely to opt out from psychology, whilst there are a certain number of students (about 5%) who start from joint degrees and switch to a psychology courses at honours level.

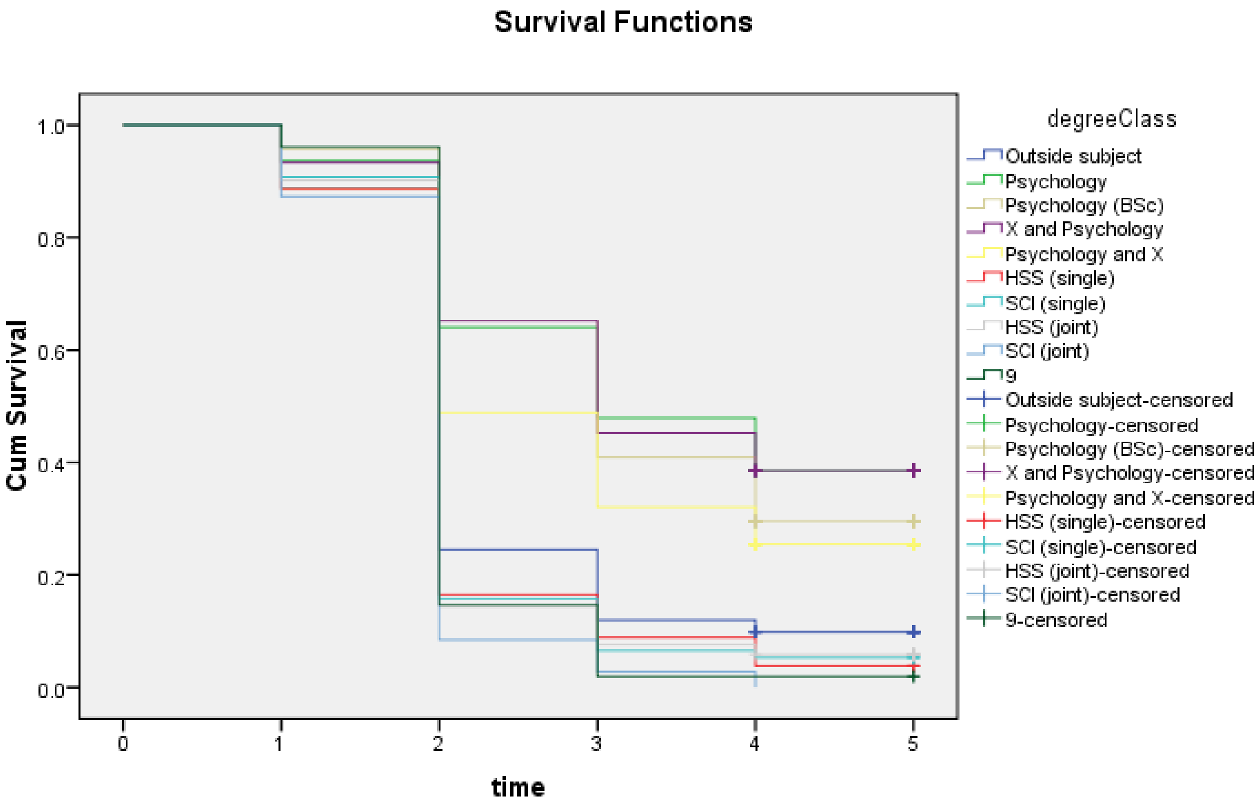


Figure 6.8. Survival rates of students taking psychology courses divided by degree type: time point 0 is the start of the degree, 1 the end of year 1 up to 5 the end of the 4-years degree.

6.2. Performance trends: an overview

In this section, we present the trends characterising the marks in the psychology 1 and psychology 2 courses in the timeframe 2002-09 and relate the results to major changes in the courses. We focus particularly on three changes: a common marking scheme affecting all courses, the change of the exam structure in year 1 and merging of the statistics component of the assessment in the two end-of-semester examinations in year 2.

6.2.1. Degree paths and diversity of performance

Earlier we questioned whether students with different degree paths showed substantial differences in their course performance based on their abilities and aptitudes. We hinted that the *number* of A grades prior to entry could play a more important role than the *types* of A grades and this will be explored later in this section to evaluate the predictive power of entry grades.

Considering their degree choice as a starting point, one could ask whether students taking the psychology courses in year one, from very different disciplines and pursuing different qualifications, also perform differently over their degrees. Figure 6.9 provides a better view of the average grades for all courses taken at university according to the qualification sought (and obtained, since only graduated students are included in this analysis).

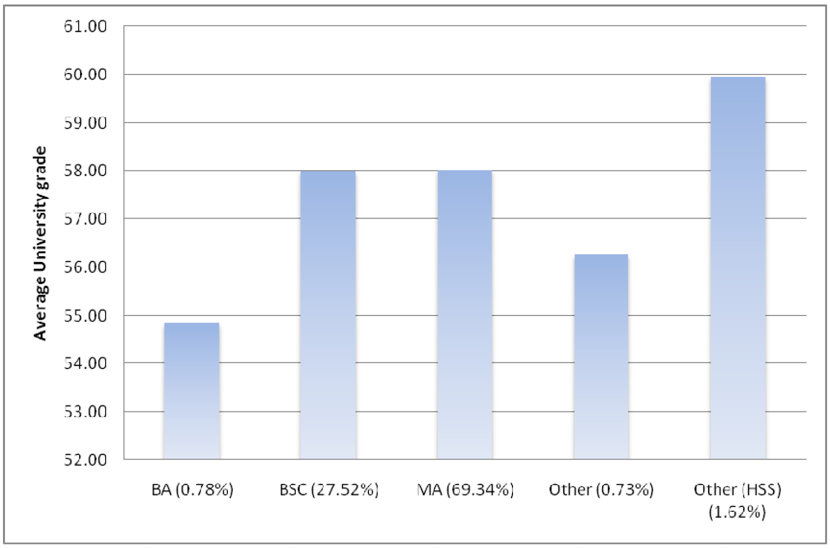


Figure 6.9. Average University grades by degree qualification for all graduated students. The percentage value identifies the number of students obtaining each degree type. As 'other' we intended non-psychology degrees.

Given the small group size of the groups other than the BSc and MA students, we will focus the discussion particularly on these two. The two groups include the obvious division of psychology degrees offered (see table 6.2). This figure shows that performance in psychology courses for students enrolled in the MA programmes is similar to those in the BSc programmes.

If we break down performance based on the year of study a very interesting picture emerges. Focusing on the MA/BSc differences, whilst there is very little difference between the grades in psychology, it looks like those students in the BSc programmes perform slightly poorer, but the difference is not statistically significant.

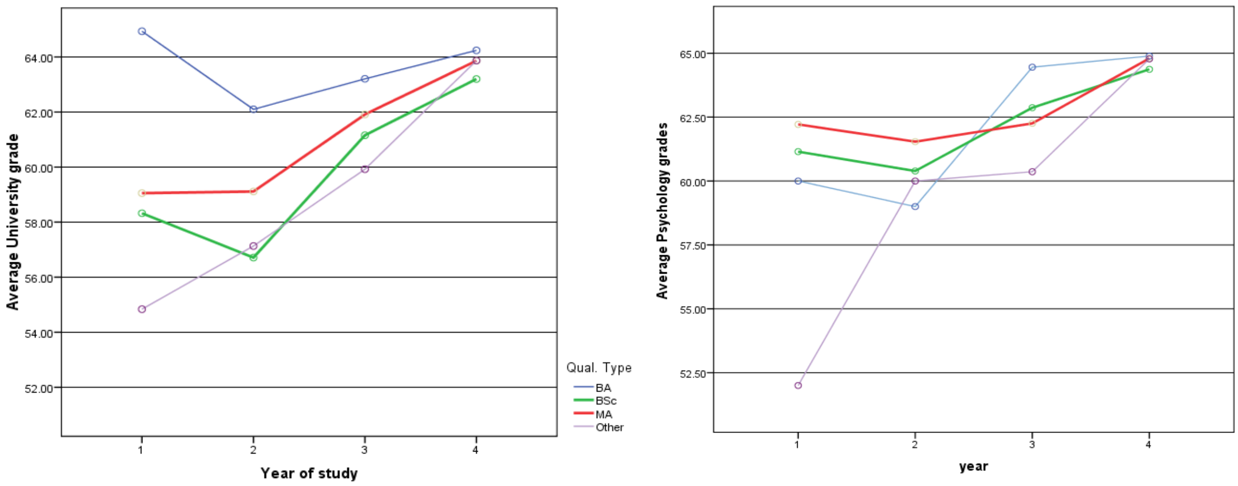


Figure 6.10. Average University grades (left) and psychology grades (right) by degree qualification and year of study for all graduated students. The percentage value identifies the number of students obtaining each degree type.

Overall there seems to be an improvement of grades over the four year degrees, however the rate of improvement differs depending on the degree types and this is true for both the average university grade and for the average grades in psychology courses. In the next chapter we will investigate in more details whether performance differs based on the subject taken. From figure 6.10 it looks like this is plausible as the rate of increase in the psychology courses is less steep than the overall university mark, which means that improvement in certain modules must be more drastic.

A mixed model ANOVA on the year of study with qualification type –BA, BSc, MA, Other-) was performed to look at the differences in average grades represented in Figure 6.10. In both performance measures there is a significant difference in the grades based on the year of study. The difference for the average grade at university is significant for year of study $F_{(3,2295)}$=3.28, p<.05 (p=0.02). There is also an interaction effect between the type of degree and year of study $F_{(4,2295)}$=1.89, p< 0.05 (p=0.03). Post-hoc t-tests using the Bonferroni correction for type I error did not reveal significant differences between either year or degree types. The differences in the average grade in psychology are only significant in the different year of study $F_{(3,2298)}=3.59$ $p<0.05$ ($p=0.013$). Post-hoc t-tests using the Bonferroni methods showed a significant difference between year 4 and year 2 ($p=0.03$).

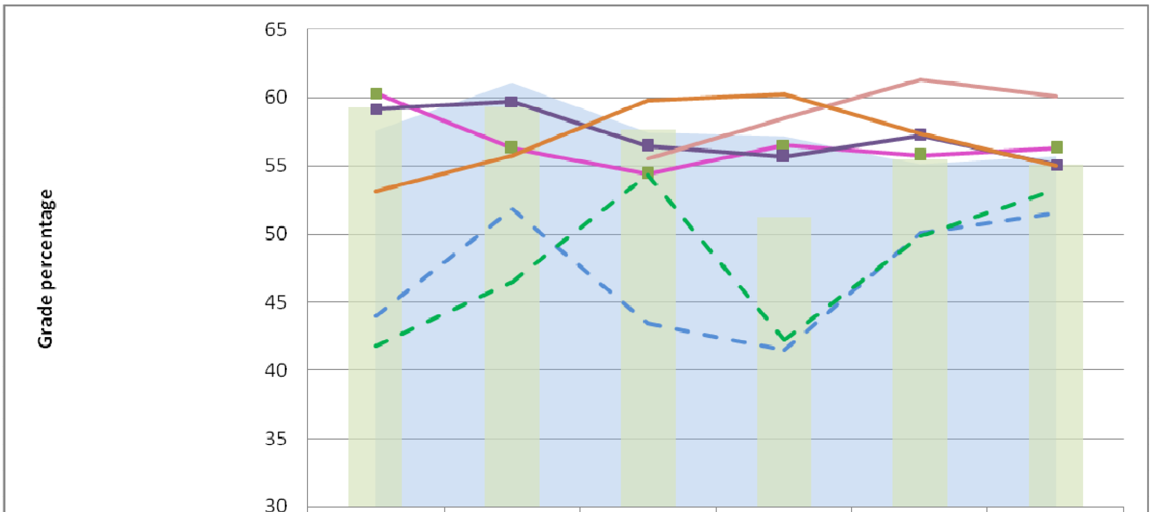
The most interesting fact of this analysis is that grades in psychology are generally stifled in year 2 which we will consider in details in the next section and in the following chapter, based on types of assessment. Even though there are differences in grades between degree subjects, these are not statistically significant.

6.2.2. Students' performance: assessment methods over time

The psychology 1 and psychology 2 courses have seen a number of substantial changes from 2002. Some were dictated by University-wide changes and others pushed in by the rotation of six different course organisers (responsible for the modules). Although the methods of teaching and content taught remained largely the same (apart for the introduction of e-learning support), the methods of assessment shaped the changes over time.
The most important changes, which caused some alignment problems with the extraction of marks for this evaluation, were the abolition of the 'Pass by exemption' in Y1 and the integration of a statistics section in the Y2 degree examination.

In year 1, prior to the year 2004, a student earning a B grade (60% or more) in the December exam (assessed with multiple choice questions only) and averaged with the submitted coursework of the first semester, was exempted from sitting the April examination. From 2004 onward, two equivalent end-of-term exams were used. In 2004/05 the first semester exam consisted only of multiple choice questions; from 2005/06, each sitting had a section containing multiple choices questions and one with essay question. These changes are reflected in figure 6.11.

Before 2004, most people who sat the second exam were students performing poorly in the first sitting; however, there were some 'good' students who for whatever reason missed the first opportunity. The patterns do not seem to be affected and are not surprising, as they required a much higher performance in the second sitting to obtain a satisfactory grade. The year 2005 shows a very interesting pattern: performance in the essay section is the same in the two semesters, but the grades obtained in the MCQs sections are very different. This was caused by two major issues: for the first time with the December exam, a new formula to counteract guessing was introduced to adjust the marks. The second issue was the implementation of the first e-pack in support of students' learning from January 2005, which helped students perform better in the second semester. However, this effect is confounded by the likely change in attitudes of students who performed badly in the first sitting who *needed* to improve. In 2006 a switch to a different textbook and e-learning support system induced a gradual increase in performance in the MCQs section, with very little difference in the average performance over the two sittings. We will explore this in more details in chapter 8.



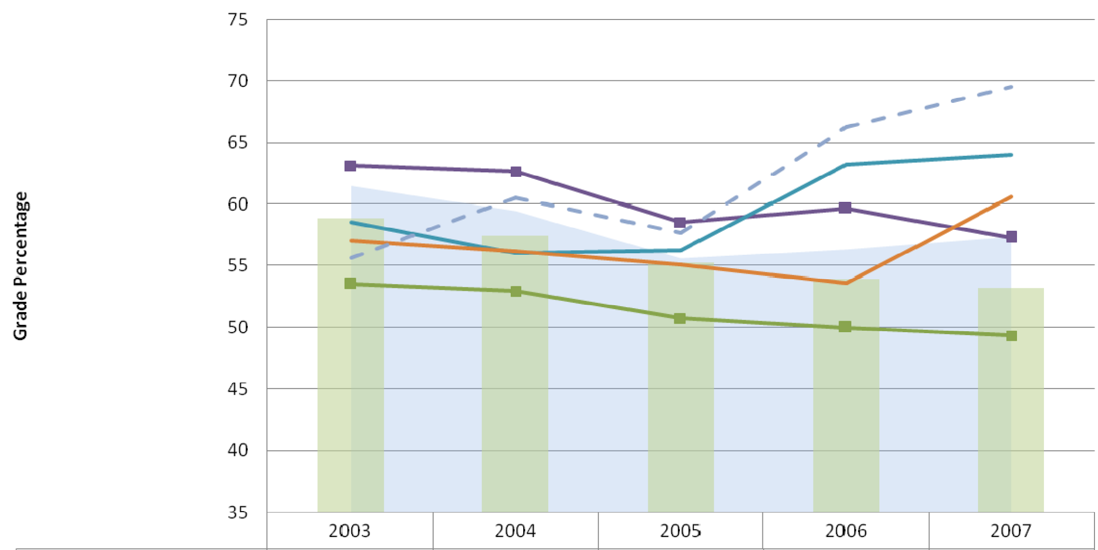
	2003	2004	2005	2006	2007	2008
Y1 Final grade	57.57	61.05	57.42	57.13	55.09	55.70
Y1 Coursework Avg	59.24	59.37	57.66	51.24	55.46	55.09
Y1 Essay1	60.30	56.31	54.41	56.50	55.77	56.32
Y1 Essay3	59.16	59.70	56.46	55.72	57.21	55.04
Y1 Exam Sem1 (Essay)			55.56	58.48	61.33	60.07
Y1 Exam Sem2 (Essay)	53.09	55.75	59.75	60.21	57.36	54.99
Y1 Exam Sem1 (MCQ)	43.99	51.84	43.48	41.48	50.09	51.49
Y1 Exam Sem2 (MCQ)	41.75	46.51	54.31	42.26	49.83	53.22

Figure 6.11. Changes in grades in the Psychology 1 course over time for different forms of assessment. The cohort starting in 2003 is the first for which we have full access to grades.

A similar graph is shown in figure 6.12 referring to the trends in the psychology 2 course. Here the major change was introduced in 2005-06: previously a statistics exam was run as a separate class exam in the second half of the year: a minimum B (60%) grade was required for admission into honours with at least a C (50%) in the statistics component.

The inclusion of the statistics section as part of a 3-hours exam meant that students had to be able to dedicate appropriate amounts of time to the essay questions and the stats (about one hour per section). The change is noticeable in the graph with the lowest grade in the statistics assessment.

Another significant influence worth noticing was a University-wide adoption of a new common marking scheme in 2005. This was intended to provide a common characterization of grades in an attempt to use a wider spectrum of marks. From the figures from both courses, it is possible to detect a wider spectrum of grades from the year of introduction which possibly produced a small decrease in the average course marks; however a study of the frequencies of grades obtained by students showed a non-significant statistical difference over the years.



	2003	2004	2005	2006	2007
Y2 Final grade	61.45	59.43	55.59	56.32	57.35
Y2 Coursework Avg	58.84	57.45	55.26	53.88	53.16
Y2 Report1	53.53	52.94	50.73	49.97	49.32
Y2 Report4	63.13	62.62	58.54	59.69	57.31
Y2 Exam Sem1 (Essay)	58.52	56.01	56.27	63.17	63.98
Y2 Exam Sem2 (Essay)	57.08	56.19	55.07	53.56	60.62
Y2 Exam Stats Avg	55.63	60.54	57.71	66.20	69.52

Figure 6.12. Changes in grades in the Psychology 2 course over time for different forms of assessment. Note that the year refers to their starting year as in table 5.4 (i.e. the 2007/08 cohort read the Psychology 2 course in 2008/09)

From the analysis in the last section, there is a significant difference in the grades over the years, especially in year 2 of the psychology courses.

To find out where this difference might be, the type of assessment for the period considered was considered shown in figure 6.12 with a breakdown of performance. The graph highlights a considerable improvement in the assessed coursework which is consistent over time (Report 1 is submitted in the first semester and Report 4 at the end of the second semester).

Furthermore, considerable changes occurred in the marks for the statistics exam and generally, performance in the coursework has been declining.
It is difficult to assess whether this decline is an effect of change in management (i.e. I was responsible for the moderation of marks from 2005 –cohort starting in 2004/05) or a decline in abilities, especially considering that the performance in the exams seems to follow a different trend. Given the consistency of the gap between the two reports marking the students' improvement in the coursework, it is possible that the decline is caused by abilities.

6.2.3. Differences in performance and types of assessment

The greater variance in table 6.6 lends some support for the claim made earlier that students in years 1 and 2 are working toward different goals when taking the same courses: this might be the reason justifying the larger spectrum of grades. With the progression to Honours, students are more focused (i.e. their degree paths converge to subject-specific courses) and as result, students tend to perform within higher and narrower brackets.

	N	Minimum	Maximum	Mean	Std. Deviation
Y1 Essays avg	1610	0.00	80.00	55.36	12.55
Y1 Exam sem 1 (essay)	1629	0.00	90.00	37.87	29.41
Y1 Exam sem 1 (MCQs)	1629	0.00	91.00	45.08	15.75
Y1 Exam sem 2 (essay)	1629	0.00	90.00	49.19	22.02
Y1 Exam sem 2 (MCQs)	1620	0.00	100.00	44.70	22.25
Y2 Report 4 (project)	712	0.00	90.00	58.17	15.39
Y2 Reports avg	725	0.00	79.00	54.25	15.14
Y2 Exam sem 1 (essay)	725	0.00	81.70	57.44	14.38
Y2 Exam sem 2 (essay)	725	0.00	79.70	53.16	15.77
Y2 Exam Statistics Avg	725	0.00	97.00	59.33	18.34
Y3 Group Project 1	406	21.00	95.00	64.43	8.86
Y3 Group Project 2	427	0.00	88.00	63.24	10.34
Y3 Exam methodology 1	461	19.00	88.00	57.38	12.66
Y3 Exam methodology 2	383	40.00	83.00	62.54	6.92
Y3 Literature Review	382	0.00	89.00	65.87	10.26
Y4 Dissertation	387	32.00	86.00	69.01	6.71
Y1 Psychology Avg	1685	21.00	80.00	57.53	8.22
Y2 Psychology Avg	964	17.00	77.00	57.87	7.88
Y3 Psychology Avg	515	9.50	75.50	60.94	7.15
Y4 Psychology Avg	398	49.89	75.78	64.81	4.57

Table 6.6. Summary of the significant differences in the various types of assessment over the period 2002-08 in the various academic years.

The breakdown of marks based on types of assessment (Figure 6.11 and 6.12), showed a difference across the years. A one-way ANOVA was used to determine if this difference is significant. Table 6.8 offers a summary of the statistical tests for each type of assessment.

Considering that the grades in different class-years were obtained by different students, it seems that the overall decline in years 1 and 2 is also present in year 3 methodology courses and in the average year 4 grades. This is further evidence that there might be a general decrease in ability demonstrated by students, despite a general increase in the performance reported by UCAS prior to admission.

	df	F	p
Y1 Essay 1	1575	8.596	0.000
Y1 Essay 3	1498	7.208	0.000
Y1 Essays avg	1573	23.599	0.000
Y1 Exam sem 1 (essay)	1042	12.339	0.000
Y1 Exam sem 1 (MCQs)	1569	56.57	0.000
Y1 Exam sem 2 (essay)	1401	13.923	0.000
Y1 Exam sem 2 (MCQs)	1404	49.289	0.000
Y1 Average	1678	16.605	0.000
Y2 Report 1	676	2.401	0.027
Y2 Report 4 (project)	677	5.244	0.000
Y2 Reports avg	694	4.342	0.000
Y2 Exam sem 1 (essay)	698	11.013	0.000
Y2 Exam sem 2 (essay)	677	9.229	0.000
Y2 Exam Statistics Avg	697	14.348	0.000
Y2 Average	957	3.896	0.001
Y3 Group Project 1	401	1.465	ns
Y3 Group Project 2	422	0.167	ns
Y3 Exam methodology 1	456	16.375	0.000
Y3 Exam methodology 2	378	4.365	0.002
Y3 Literature Review	377	1.849	ns
Y3 Average	509	1.674	ns
Y4 Dissertation	383	0.47	ns
Y4 Average	394	2.732	0.044

Table 6.7. Summary of the significant differences in the various types of assessment over the period 2002-08 in the various academic years.

By looking at the different types of the assessment, one might wonder if the decrease is associated specifically with the content: most psychology students dread the research methodology and statistics component of their courses. In fact, the final report in year 2 (group project) and the statistics mark are strongly correlated to later performance. This is an interesting finding as the level achieved in these analytic assessments is strongly correlated with all other assessment over the years. Table 6.8 show the intercorrelations between the various types of assessment and the average grades for each year as well as the final grade. It is possible to observe that there is a medium to large correlation between all grades. A correlation between assessments was expected; what was not expected is the strength of the relations between them. Such strong correlations will need to be taken into account in any regression analysis.

		Y1_E1	Y1_E3	Y1_Eavg	Y_EX1	Y1_MCQ1	Y1_EX2	Y1_MCQ2	PSY_Y1	Y2R1	Y2R4	Y2_Ravg	Y2_EX1	Y2_EX2	Y2Stats	PSY_Y2	Y3_group1	Y3_group2	Y3_meth1	Y3_meth2	Y3_litReview	PSY_Y3	Y4_diss	PSY_Y4	UniAverage
Y1_E1	Pearson Correlation	1	.317**	.531**	.162**	.249**	.141**	.112**	.428**	.233**	.192**	.196**	.185**	.212**	.161**	.297**	0.001	.128*	-0.03	.129*	.249**	.234**	0.112	.158**	.345**
	Sig. (2-tailed)		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.989	0.017	0.561	0.021	0	0	0.05	0.005	0
	N	1582	1493	1562	1028	1551	1389	1392	1440	660	661	678	681	660	680	799	340	347	378	321	321	422	304	314	1567
Y1_E3	Pearson Correlation		1	.643**	.167**	.303**	.172**	.186**	.522**	.311**	.355**	.377**	.203**	.281**	.194**	.381**	.186**	.139**	.124*	.195**	0.07	.258**	.151**	.155**	.456**
	Sig. (2-tailed)			0	0	0	0	0	0	0	0	0	0	0	0	0	0.001	0.01	0.016	0	0.213	0	0.009	0.006	0
	N			1503	987	1492	1350	1350	1402	655	654	669	670	654	671	785	337	345	375	318	319	418	303	313	1504
Y1_Eavg	Pearson Correlation			1	.225**	.481**	.295**	.392**	.716**	.383**	.391**	.422**	.252**	.376**	.218**	.472**	.149**	.211**	.211**	.238**	.263**	.429**	.216**	.250**	.587**
	Sig. (2-tailed)				0	0	0	0	0	0	0	0	0	0	0	0	0.006	0	0	0	0	0	0	0	0
	N				1030	1558	1405	1408	1455	665	666	683	686	665	686	805	340	348	379	321	321	423	305	315	1576
Y_EX1	Pearson Correlation				1	.409**	.213**	.256**	.501**	.184**	.147**	.146**	.258**	.248**	.189**	.304**	0.103	0.052	0.015	0.085	0.15	.201**	0.078	.220*	.325**
	Sig. (2-tailed)					0	0	0	0	0.001	0.006	0.006	0	0	0	0	0.216	0.534	0.852	0.312	0.078	0.006	0.466	0.031	0
	N					1047	1017	1017	942	352	348	358	361	354	358	435	147	147	167	144	138	185	90	96	1044
Y1_MCQ1	Pearson Correlation					1	.254**	.591**	.727**	.337**	.316**	.325**	.298**	.469**	.373**	.489**	.182**	.153**	.381**	.227**	.209**	.405**	.209**	.337**	.511**
	Sig. (2-tailed)						0	0	0	0	0	0	0	0	0	0	0.001	0.004	0	0	0	0	0	0	0
	N						1391	1393	1446	662	663	680	683	662	683	799	339	347	378	320	320	421	305	315	1573
Y1_EX2	Pearson Correlation						1	.387**	.533**	.167**	.126**	.155**	.238**	.238**	.236**	.260**	.124*	0.108	.244**	0.069	0.049	.288**	0.071	0.116	.413**
	Sig. (2-tailed)							0	0	0	0.002	0	0	0	0	0	0.04	0.074	0	0.269	0.438	0	0.283	0.072	0
	N							1408	1305	569	572	586	586	574	588	692	274	276	308	257	256	345	231	240	1408
Y1_MCQ2	Pearson Correlation							1	.595**	.174**	.174**	.214**	.233**	.319**	.296**	.318**	0.1	0.101	.364**	0.116	-0.017	.357**	0.052	0.104	.491**
	Sig. (2-tailed)								0	0	0	0	0	0	0	0	0.098	0.093	0	0.063	0.781	0	0.429	0.108	0
	N								1307	570	573	587	587	575	589	693	274	276	308	257	256	345	231	240	1411
PSY_Y1	Pearson Correlation								1	.453**	.446**	.447**	.382**	.470**	.356**	.551**	.275**	.253**	.399**	.299**	.237**	.488**	.289**	.390**	.734**
	Sig. (2-tailed)									0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	N									671	671	688	691	671	691	828	401	420	454	378	377	499	381	391	1685
Y2R1	Pearson Correlation									1	.477**	.555**	.321**	.348**	.292**	.523**	.263**	.194**	.233**	.296**	.200**	.380**	.152*	.154**	.488**
	Sig. (2-tailed)										0	0	0	0	0	0	0	0	0	0	0	0	0.01	0.008	0
	N										662	676	673	655	673	661	319	326	355	303	301	386	284	292	683
Y2R4	Pearson Correlation										1	.879**	.287**	.412**	.300**	.668**	.322**	.229**	.326**	.333**	.372**	.529**	.304**	.349**	.575**
	Sig. (2-tailed)											0	0	0	0	0	0	0	0	0	0	0	0	0	0
	N											684	677	673	683	676	324	330	360	307	306	396	287	295	684
Y2_Ravg	Pearson Correlation											1	.360**	.458**	.397**	.759**	.410**	.353**	.342**	.383**	.426**	.574**	.305**	.327**	.634**
	Sig. (2-tailed)												0	0	0	0	0	0	0	0	0	0	0	0	0
	N												692	677	696	683	325	331	362	308	307	398	287	296	701
Y2_EX1	Pearson Correlation												1	.417**	.386**	.612**	.199**	.220**	.276**	.202**	.160**	.437**	.290**	.406**	.522**
	Sig. (2-tailed)													0	0	0	0	0	0	0	0.005	0	0	0	0
	N													677	694	683	324	329	360	306	305	397	284	293	704
Y2_EX2	Pearson Correlation													1	.452**	.784**	.226**	.247**	.328**	.382**	.263**	.541**	.277**	.385**	.600**
	Sig. (2-tailed)														0	0	0	0	0	0	0	0	0	0	0
	N														682	679	320	326	358	303	302	394	284	293	684
Y2Stats	Pearson Correlation														1	.558**	.300**	.324**	.475**	.259**	.281**	.453**	.338**	.387**	.486**
	Sig. (2-tailed)															0	0	0	0	0	0	0	0	0	0
	N															687	324	330	361	307	306	397	287	296	704
PSY_Y2	Pearson Correlation															1	.388**	.384**	.454**	.458**	.375**	.685**	.389**	.533**	.770**
	Sig. (2-tailed)																0	0	0	0	0	0	0	0	0
	N																406	427	461	383	382	506	387	398	964
Y3_group1	Pearson Correlation																1	.347**	.365**	.321**	.277**	.597**	.335**	.376**	.525**
	Sig. (2-tailed)																	0	0	0	0	0	0	0	0
	N																	360	393	371	369	406	321	322	406
Y3_group2	Pearson Correlation																	1	.422**	.443**	.346**	.659**	.414**	.448**	.563**
	Sig. (2-tailed)																		0	0	0	0	0	0	0
	N																		415	365	360	426	338	342	427
Y3_meth1	Pearson Correlation																		1	.339**	.222**	.734**	.407**	.521**	.673**
	Sig. (2-tailed)																			0	0	0	0	0	0
	N																			377	373	460	367	373	461
Y3_meth2	Pearson Correlation																			1	.236**	.646**	.286**	.495**	.620**
	Sig. (2-tailed)																				0	0	0	0	0
	N																				370	383	308	309	383
Y3_litReview	Pearson Correlation																				1	.577**	.283**	.324**	.487**
	Sig. (2-tailed)																					0	0	0	0
	N																					382	305	307	382
PSY_Y3	Pearson Correlation																					1	.529**	.714**	.885**
	Sig. (2-tailed)																						0	0	0
	N																						370	377	515
Y4_diss	Pearson Correlation																						1	.553**	.560**
	Sig. (2-tailed)																							0	0
	N																							387	387
PSY_Y4	Pearson Correlation																							1	.816**
	Sig. (2-tailed)																								0
	N																								398
UniAverage	Pearson Correlation																								1
	Sig. (2-tailed)																								
	N																								

****. Correlation is significant at the 0.01 level (2-tailed).**
***. Correlation is significant at the 0.05 level (2-tailed).**

Table 6.8. Correlation table of the average grades achieved in the different forms of assessment.

6.2.4. Perpetuating stereotypes? Gender differences and degree subjects

One of the aspects explored in this chapter was the bias toward a higher presence of females in the sample. If we look in more detail to the distribution of genders in the various degree types (table 6.9) this shows a statistically significant difference ($X^2_{(9)}$=114.147, p<.001). Interestingly, however, for this sample we did not find a greater presence of male students associated with the scientific degrees.

degree Type		sex: Males	sex: Females	Total
Psychology (MA)	Count	105	470	575
	% within degree type	18.3%	81.7%	100.0%
	% within sex	17.2%	33.8%	28.7%
Psychology (BSc)	Count	40	122	162
	% within degree type	24.7%	75.3%	100.0%
	% within sex	6.6%	8.8%	8.1%
Psychology with...	Count	29	96	125
	% within degree type	23.2%	76.8%	100.0%
	% within sex	4.8%	6.9%	6.2%
Other with Psychology	Count	92	138	230
	% within degree type	40.0%	60.0%	100.0%
	% within sex	15.1%	9.9%	11.5%
Philosophy	Count	31	48	79
	% within degree type	39.2%	60.8%	100.0%
	% within sex	5.1%	3.5%	3.9%
Cog Science, Linguistics, AI	Count	29	46	75
	% within degree type	38.7%	61.3%	100.0%
	% within sex	4.8%	3.3%	3.7%
Social and political studies	Count	64	150	214
	% within degree type	29.9%	70.1%	100.0%
	% within sex	10.5%	10.8%	10.7%
Economy and management	Count	47	24	71
	% within degree type	66.2%	33.8%	100.0%
	% within sex	7.7%	1.7%	3.5%
English, languages and cultures	Count	33	69	102
	% within degree type	32.4%	67.6%	100.0%
	% within sex	5.4%	5.0%	5.1%
Sciences	Count	140	228	368
	% within degree type	38.0%	62.0%	100.0%
	% within sex	23.0%	16.4%	18.4%
Total	**Count**	**610**	**1391**	**2001**
	% within degree type	**30.5%**	**69.5%**	**100.0%**
	% within sex	**100.0%**	**100.0%**	**100.0%**

Table 6.9. Distribution of genders across degree types within the sample considered.

In general, females outperform males both in terms of average grades ($t_{(1909)}$=-6.77 p<.001) and final degree classes. Table 6.10 shows the distribution of the graduates from this period with the degree class achieved by males and females.

degree Class		sex Males	Females	Total
1st	Count	33	109	142
	% within degree class	23.2%	76.8%	100.0%
	% within sex	5.4%	7.8%	7.1%
2nd Div 1	Count	161	413	574
	% within degree class	28.0%	72.0%	100.0%
	% within sex	26.4%	29.7%	28.7%
2nd Div 2	Count	43	75	118
	% within degree class	36.4%	63.6%	100.0%
	% within sex	7.0%	5.4%	5.9%
3rd	Count	6	5	11
	% within degree class	54.5%	45.5%	100.0%
	% within sex	1.0%	.4%	.5%
Pass	Count	72	133	205
	% within degree class	11.8%	9.6%	10.2%
	% within sex	35.1%	64.9%	100.0%
	Total Count	**315**	**735**	**1050**

Table 6.10. Distribution of degree classed achieved by students between 2002-08.

	sex	N	Mean	Std. Dev.	T	df	p
Y1 Essays avg	Males	485	52.35	14.81			
	Females	1125	56.66	11.19	-6.40	1608	0.001
Y1 Exam sem 1 (essay)	Males	491	38.41	28.67			
	Females	1138	37.64	29.73			ns
Y1 Exam sem 1 (MCQs)	Males	491	42.41	16.11			
	Females	1138	46.23	15.46	-4.52	1627	0.001
Y1 Exam sem 2 (essay)	Males	491	47.53	21.48			
	Females	1138	49.91	22.23	-1.99	1627	0.046
Y1 Exam sem 2 (MCQs)	Males	490	44.21	22.75			
	Females	1130	44.92	22.04			ns
Y2 Report 4 (project)	Males	172	53.17	19.85			
	Females	540	59.76	13.30	-4.97	710	0.001
Y2 Reports avg	Males	180	48.75	19.83			
	Females	545	56.07	12.74	-5.74	723	0.001
Y2 Exam sem 1 (essay)	Males	180	53.08	17.12			
	Females	545	58.88	13.05	-4.72	723	0.001
Y2 Exam sem 2 (essay)	Males	180	48.65	19.81			
	Females	545	54.65	13.89	-4.48	723	0.001
Y2 Exam Statistics Avg	Males	180	54.75	21.37			
	Females	545	60.84	16.97	-3.90	723	0.001
Y3 Group Project 1	Males	71	64.94	6.95			
	Females	335	64.32	9.22			ns
Y3 Group Project 2	Males	79	63.77	8.55			
	Females	348	63.12	10.71			ns
Y3 Exam methodology 1	Males	86	59.27	12.21			
	Females	375	56.94	12.73			ns
Y3 Exam methodology 2	Males	64	61.17	7.02			
	Females	319	62.82	6.88			ns
Y3 Literature Review	Males	64	67.52	7.92			
	Females	318	65.53	10.65			ns
Y4 Dissertation	Males	77	68.87	6.11			
	Females	310	69.04	6.85			ns
Y1 Psychology Avg	Males	499	55.58	8.68			
	Females	1186	58.35	7.89	-6.37	1683	0.001
Y2 Psychology Avg	Males	248	55.34	9.46			
	Females	716	58.74	7.05	-5.97	962	0.001
Y3 Psychology Avg	Males	105	60.68	6.38			
	Females	410	61.01	7.34			ns
Y4 Psychology Avg	Males	83	63.86	4.79			
	Females	315	65.06	4.49	-2.13	396	0.034

Table 6.11. Differences in performance between genders considering the various forms of assessment in the curriculum.

Looking at the effect of gender in the level of achievement with different types of assessment, table 6.11 provides a detailed overview of the statistically significant differences. In a later section we performance will be related to stylistic differences to investigate whether males show particular stylistic patterns *causing* the lower performance.

6.2.5. The predicting power of prior performance.

The final consideration regarding overall performance trends is the concern that the sample is characterised by high performing students and this affects dramatically the success rate of the students enrolled in the courses at the University of Edinburgh.

To test this hypothesis we conducted a regression analysis to establish which parameter in the prior performance profile is able to predict the average performance levels at the university. A number of regressors have been considered and are shown in the graphs below (Figure 6.13 to 6.15). The first graph (figure 6.13) depicts the relation between the number and class of entry grades recorded.

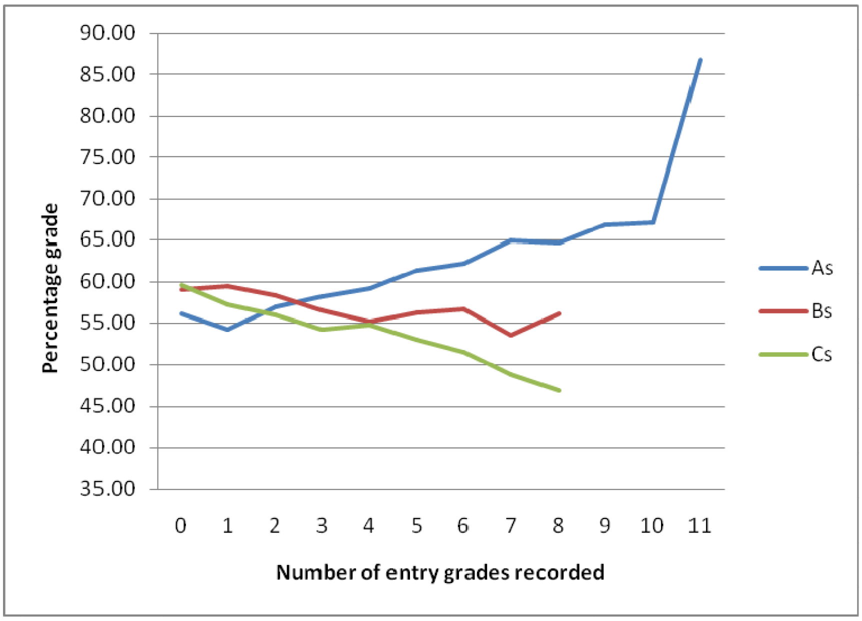


Figure 6.13. Types of entry grades and marks achieved in year one.

It is evident that the number of As, irrespective of the subject in which they were achieved, is a good indicator of performance in year 1. In fact, the average increment with increasing number of As is 1.2, for the number of Bs is -0.45 and -1 for the number of Cs.

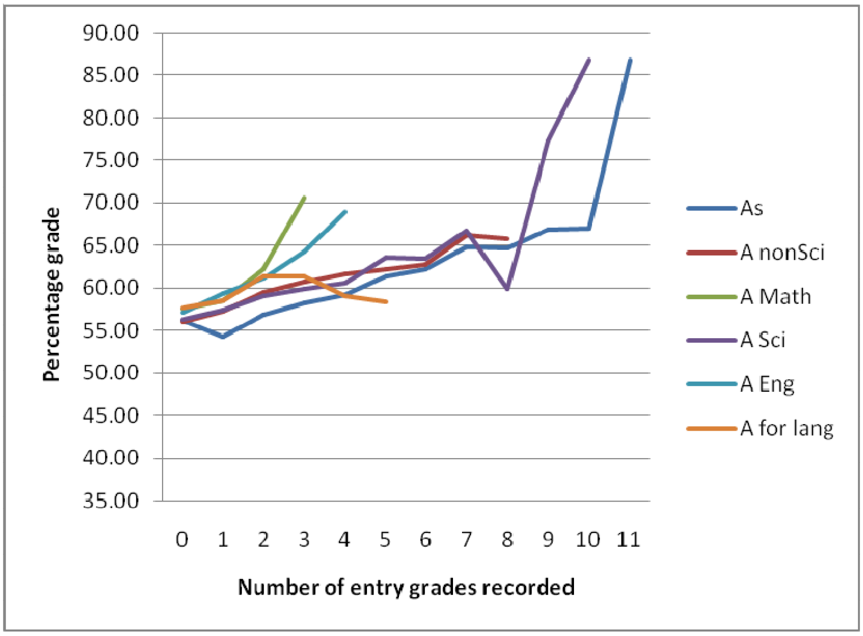


Figure 6.14. Types of A grades prior to entry and marks achieved in year one.

Figure 6.15. Average increments of marks between year 1 and year 4(delta) and entry grades.

More complicated is the relation between the number of As in different subjects and performance at university. From figure 6.14 the two core subjects (Maths and English) are better predictor of performance than foreign languages. However when the number of As increases comparing the sciences and non-sciences subjects, it seems that expected performance has similar trends. It should be noted that these graphs are used specifically to show the bias of outliers on the traditional methods of analysis and we will compare these with mining techniques later in this chapter.

If we look in more detail at the relation between the number and types of As prior to university and the overall university performance, we conducted a stepwise linear regression. The model comprising the combination of the total number of As in science, non-science and in psychology predicts the average university grades, b=.232, $t_{(1890)}$=179.15, p<.001 and accounts for about 10% of the variance, r^2=.10, $F_{(4,1885)}$=52.35, p<.001.

Furthermore, as expected from the literature, prior performance (the proportion of A, B and C grades over the total marks recorded) also predict the university average grades, b=.28, $t_{(1907)}$=149.94, p<.001 and account for about 13% of the variance (with As accounting for close to 9% alone), r^2=13.4, $F_{(3,1907)}$=98.11, p<.001.

The final observation regards the margin of improvement once students are at university. Even though Figure 6.13 presented a gloomy picture of the expected decrease in performance proportional to the number of C grades a student has, Figure 6.15 shows the average increase in grades between the year 1 marks and year 4 marks for the psychology courses (hence considering only students completing a psychology degree). This reveals a very encouraging trend: whilst students with a higher number of As tend to improve very little, it seems that lower performing students at entry, *could* improve considerably, and this, despite the odd peak with 4 Cs, is linked to the number of Cs.

6.3. The ground rules of data mining

In chapter 4 a visual approach to mining the literature on styles was introduced to familiarise the reader with the general concepts. In chapter 5 we presented a coarse methodology for data mining and web usage mining, however we kept the practical details at a minimum.

In this chapter we will dig deeper into the application of mining techniques and expand this in the next chapter when looking at styles, therefore it is necessary to provide more details on how data mining is accomplished. As a further incremental step, in this section we will address some of the core ideas, which will be adopted later in the analysis of the measures of style and usage data.

A detailed guide to mining techniques, algorithms and techniques can be found elsewhere (Hand et al. 2001, Witten & Frank 2005, Han & Kamber 2006). As a reminder of chapter 5 there are four steps involved: pre-processing, pattern discovery, data visualization and data analysis.

Pre-processing is similar in many aspects to data preparation and exploratory data analysis presented in traditional research methods textbooks: data need to be cleaned, summarised, integrated and/or transformed as suitable, then need to be aggregated or reduced to be prepared for statistical testing. In data mining this step is crucial: contrary to statistical techniques in which data is collected to test specific hypotheses, in data mining variables (or instances) need to be selected appropriately. As well as the Simpson's paradox exemplified in chapter 6, there is evidence that the effect of additional variables affects negatively the results of mining techniques (Witten and Frank 2005), Therefore the importance of this step should not be underestimated.

Pattern discovery allows to individuate useful knowledge about the data using one of many algorithms. In our case we will benefit from the use PASW v17 for a first look at clustering solutions, and use WEKA (Waikato Environment for Knowledge Analysis), which is a workbench of software tool to aid data mining, to replicate the clustering where possible. The tools offer 3 categories of algorithms: *classification*, *association* and *clustering* algorithms. The workbench is a powerful tool to exploit the effectiveness of the most recent advances in data mining making the use of complex techniques easier.

Data visualization is twofold: it is the process of making sense of the data by visually exploring the data, but is also used to provide a visual enhancement to facilitate the understanding of complex relations. Both will be used at various stages of the analysis.

Data analysis is the process of testing hypotheses and comparing the effectiveness of different techniques: this is more similar to traditional statistical testing. In data mining this is usually done by comparing the performance of learning algorithms in particular applications.

In general terms, the three families of algorithms have different approaches, expect different data structures and suit particular problems. In particular, *classification* is used mainly to test predictions related to a class of outcomes: for example, a decision tree or a Bayes classification are used to test if given a set of instances with particular attributes, one can predict a given target class.

Association techniques are an extension to classification, in the sense that rather than predicting a class, these techniques allow to predict any attribute. Social scientists are familiar with correlation and regression methods which are the basis for PCA (principal component analysis).

Unlike the other two, which are mainly predictive in scope, *clustering* is the process of grouping objects into classes of similar objects: the features are the instances (variables) and the groups are created based on the appropriate selection of particular occurrences (nominal/ordinal variable) or magnitude (ratio/scale) of features.

In this chapter we already used correlation and regression to quantify the strength of the relations between variables. In the next section clustering is applied to prior performance to provide an effective alternative to understand non-linear relations between past grades on university achievement.

6.3.1. Clustering prior performance

To test whether it was possible to identify patterns in the grades of admitted students we performed a Two-steps cluster analysis using the built-in algorithm in PASW v17.

The total number of As, Bs and Cs in Psychology, Maths, Other sciences (aggregated average), English, Foreign Languages and Other social sciences (aggregated average) were entered in the algorithm with gender, country of origin and degree type as categorical variables which returned a classification of three clusters.

For clarity we report in table 6.12 a partial view of the clusters and the distribution of students in each cluster tabulated according to number of A grades in Psychology, Maths, English and a Foreign Language. Although this is a partial view, the composition of clusters becomes clearer.

In addition to the table, if we were to describe verbally the composition of C2, for example, in this cluster we find the majority of students who did not take psychology. Generally this group has also the majority of students with at least 1 A in Maths, English and Foreign languages. The number of combinations increases with the number of variables, making it more difficult to provide a succinct verbal account of each cluster.

Average number of grades recorded for each student

		tot As		tot Bs		tot Cs	
		Mean	sd	Mean	sd	Mean	sd
Cluster	1	0.29	89.8%	0.19	60.5%	0.07	34.7%
	2	3.17	206.8%	2.20	158.1%	1.10	138.7%
	3	3.19	214.5%	2.29	157.3%	1.06	135.9%
	Combined	2.45	224.6%	1.72	165.4%	0.83	128.3%

tot As Psychology

		0		1		2	
		Freq.	%	Freq.	%	Freq.	%
Cluster	1	509	29.6%	0	.0%	0	.0%
	2	1063	61.8%	0	.0%	0	.0%
	3	148	8.6%	221	100.0%	81	100.0%
	Combined	1720	100.0%	221	100.0%	81	100.0%

Tot As Maths

		0		1		2		3	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%
Cluster	1	509	33.0%	0	.0%	0	.0%	0	.0%
	2	677	43.9%	318	80.3%	64	82.1%	4	80.0%
	3	357	23.1%	78	19.7%	14	17.9%	1	20.0%
	Combined	1543	100.0%	396	100.0%	78	100.0%	5	100.0%

Tot As English

		0		1		2		3		4	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Cluster	1	509	35.5%	0	.0%	0	.0%	0	.0%	0	.0%
	2	625	43.6%	370	75.7%	64	67.4%	3	100.0%	1	100.0%
	3	300	20.9%	119	24.3%	31	32.6%	0	.0%	0	.0%
	Combined	1434	100.0%	489	100.0%	95	100.0%	3	100.0%	1	100.0%

Tot As foreign language

		0		1		2		3		4		5		6	
		Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Cluster	1	509	30.1%	0	.0%	0	.0%	0	.0%	0	.0%	0	.0%	0	.0%
	2	806	47.7%	178	77.1%	63	81.8%	10	66.7%	4	66.7%	1	100.0%	1	100.0%
	3	376	22.2%	53	22.9%	14	18.2%	5	33.3%	2	33.3%	0	.0%	0	.0%
	Combined	1691	100.0%	231	100.0%	77	100.0%	15	100.0%	6	100.0%	1	100.0%	1	100.0%

Table 6.12. Partial overview of the defining features of the 3 clusters obtained from prior performance.

To prove the utility of such a categorization *emergent* from the data, we conducted a one-way ANOVA using the cluster assignment as grouping factor on the key performance indicators throughout for the first and second year of the psychology courses.

Table 6.13 shows a summary of the findings which are quite surprising. This method, in fact, demonstrated that the emergent patterns from the prior performance produce three groups, which enable discriminating between performances.

Assessment point	df	F	p	C1 avg	C2 avg	C3 avg
Y1 coursework	2, 1607	27.736	.000	52.4	54.8	58.9
Y1 final grade	2, 1797	19.106	.000	57.3	56.7	57.8
Y2 Coursework	2, 705	8.504	.000	51.8	52.9	57.6
Y2 statistics component	2, 705	5.509	.004	56.2	58.3	62.4
Y2 Final grade	2, 823	8.943	.000	57.6	57.4	60.1

Table 6.13. Significant differences in grades for the three clusters and their centroids.

The example just presented shows the practical value of a rudimentary application of data mining to prior performance to produce a data-driven classification of prior performance. Similar techniques will be used to reduce the complexity of the results generated from styles measures and online usage.

6.4. Chapter summary

In this chapter we provided a general overview of the samples of students taking the courses in psychology. A number of demographic factors were identified and described, largely matching the national averages. Overall performance in the courses was considered and we tested possible differences caused by demographic features and sample-specific trends.
Then we looked at prior performance and evaluated the predictive power of prior performance on the average grades at university level. We reported a strong effect of the proportion of A grades obtained before entry, able to account for about 10% of the variance of the grades obtained during the university degree.
Finally, we demonstrated how applying data mining techniques to the prior performance led us to generate a classification of 3 groups which characterised students. Significant differences were found in the grades at key points in year 1 and 2, showing the practical effectiveness of the technique.

In the next two chapters, we step deeper in the database: in the next chapter, we look at the individual differences in performance and the relations between instruments. In chapter 8, we examine the patterns of online usage and relate them back to academic performance and stylistic differences.

Chapter 7. Differential aspects of performance: styles in practice

The last chapter confirmed the predictable trend that prior performance is a fairly powerful indicator of academic performance (AP) at university. However, in the first four chapters, it was argued that the trend does not explain how and why some students do not fit in this prediction and under perform[13].

In the review chapters, we intimated that some differential aspects quantified by styles measures could provide a possible explanation, especially when one's styles mismatch the methods of teaching and assessment imposed by the university curriculum. In this chapter we will explore whether any of the styles measures used can account for a higher/lower performance than expected and/or if any combination of styles is more effective in succeeding in the psychology courses.

Firstly, we will consider each metric independently to provide a detailed analysis of the validity and reliability of the instruments used. Then, interrelations between metrics will be evaluated bringing further evidence for the styles debate as identified in chapter 4. We will also look at the possible relations and dependencies between measures of styles and academic indicators (AP, participation, attendance and punctuality).
Finally we will test whether the clusters emergent from styles metrics are a useful method to differentiate performance in the students of the sample studied.

---

[13] Given the high entry level of students in these courses it is not possible to comment on the opposite trend in which students with mediocre grades before entry, mature and perform very well at university.

7.1. Styles Measures: reliability, validity and their interrelations

To maintain a coherent argument, the various measures of styles are considered following a similar structure in the presentation for all sections: 1) brief summary of the application of the tool, 2) validity and reliability of the instruments (both at item and structure level where suitable), 3) overview of the results and group differences.

Following the guidance in Tabachnik and Fidell (1998) each presentation of the test with different samples has been treated as separate for the purpose of testing the reliability of the scales and to verify the factor structure. Each year-class is also treated independently. Although pooling participants would provide better power for particular statistical analysis, the authors indicated that this is preferable in order to demonstrate the internal consistency of the tests and avoid confounding sample-specific patterns.

7.1.1. Approaches and Study Skills Inventory for Students (ASSIST)

Use of the test

The ASSIST questionnaire is the only tool that was used in all iterations of this project. This choice was simple: the relevance of the resulting profiles had a particular appeal to psychology students who were more motivated in completing the questionnaire to learn about their own approaches to learning.
This allowed us to collect data from 3 different starting cohorts (since 2006/07) and, for 103 students, also to repeat the test in the second year, providing a rare opportunity to explore possible changes in the measures over time.

Validity and reliability

As observed in chapter 6, the ASSIST questionnaire provides metrics for three approaches to learning with 13 subscales. Using a similar format to the one presented in McCune and Entwistle (2000, also adopted in table 5.7), in the next few pages we present an overview of the factor loadings and reliability of all the subscales in each sample, all measured using

Cronbach alpha. The format adopted should simplify the comparisons with the result reported by the authors.

These results show a consistent replication of the factor loadings presented in previous research for each subscale and a good reliability of the ASSIST scales with alphas for the three major dimensions of approaches (deep, surface and strategic) generally above .8.

The subscales also have a fair to high reliability: most scales have alphas above .6. There are however some exceptions, particularly for the first year cohort starting in 2007-08, in which unacceptable reliability values are present in the subscales (as low as .16 to .3). It is not clear why the results from this class are anomalous as in the year 2 class (2008-09), which is partly the same cohort, the reliability values are high.

The interrelations between the constructs show expected patterns with the deep approach positively correlated to the strategic approach (ranges between .27 and .47) and negatively correlated with the surface approach (ranges -.18 to -.36, with one exception in the Y1 2008 sample in which the correlation was not significant).
The two scales to assess preferences for the learning environments also supported the expected patterns of relations with the three approaches to learning.

Y1 STUDENTS (2006-07 CLASS), N=78, 31% of the class				
	DEEP	STRATEGIC	SURFACE	ALPHA
DEEP Approach				**.91 (.91)**
Seeking meaning	**.86****	.50**	-.32**	.57 (.58)
Relating ideas	**.91****	.43**	-.36**	.73 (.74)
Use of evidence	**.91****	.37**	-.29*	.72 (.72)
Interest in ideas	**.84****	.37**	-.30**	.80 (.81)
STRATEGIC Approach				**.80 (.79)**
Organised studying	.24*	**.73****	-.11	.38 (.38)
Time management	.29*	**.82****	-.10	.78 (.78)
Alertness to assessment demands	.12	**.42****	.15	.49 (.49)
Achieving	.42**	**.81****	-.09	.49 (.49)
Monitoring effectivness	.54**	**.57****	-.09	.59 (.58)
SURFACE/APATHETIC Approach				**.76 (.76)**
Lack of purpose	-.28*	-.17	**.54****	.63 (.65)
Unrelated memorising	-.32**	.020	**.74****	.59 (.58)
Syllabus boundness	-.40**	-.14	**.62****	.56 (.56)
Fear of failure	-.003	.064	**.72****	.80 (.80)
Preferences for learning environments				
Supporting understanding(DEEP)	.62**	.28*	-.45**	.73 (.74)
Transmitting information (SURFACE)	-.20	-.10	.37**	.65 (.66)
	DEEP	STRATEGIC	SURFACE	
DEEP Approach	1	.47**	-.36**	
STRATEGIC Approach		1	-.08	
SURFACE/APATHETIC Approach			1	

Y2 STUDENTS (2006-07 CLASS), N=89, 54% of the class				
	DEEP	STRATEGIC	SURFACE	ALPHA
DEEP Approach				**.79 (.79)**
Seeking meaning	**.73****	.52**	-.30**	.54 (.54)
Relating ideas	**.70****	-.01	-.27*	.57 (.57)
Use of evidence	**.78****	.32**	-.24*	.56 (.57)
Interest in ideas	**.76****	.38**	-.26*	.71 (.71)
STRATEGIC Approach				**.86 (.87)**
Organised studying	.21*	**.86****	-.08	.65 (.64)
Time management	.04	**.79****	.03	.74 (.75)
Alertness to assessment demands	.38**	**.58****	.01	.62 (.62)
Achieving	.46**	**.83****	-.26*	.62 (.64)
Monitoring effectivness	.58**	**.70****	-.15	.56 (.56)
SURFACE/APATHETIC Approach				**.79 (.78)**
Lack of purpose	-.38**	-.38**	**.65****	.69 (.70)
Unrelated memorising	-.37**	.11	**.78****	.52 (.52)
Syllabus boundness	-23*	-.2	**.55****	.50 (.49)
Fear of failure	.08	.11	**.77****	.80 (.80)
Preferences for learning environments				
Supporting understanding(DEEP)	.43**	.04	-.30**	.61 (.60)
Transmitting information (SURFACE)	-.21*	.13	.32**	.75 (.76)
DEEP Approach	1	.41**	-.36**	
STRATEGIC Approach		1	-.11	
SURFACE/APATHETIC Approach			1	

Table 7.1. Factor loadings and alphas (standardised in brackets) for the various scales for the year 1 and year 2 classes of the year 2006-07. For all values ** indicates a significance at 0.01 (two tailed) and * at 0.05 level.

Y1 STUDENTS (2007-08 CLASS), N= 158, 55.1% of the class				
	DEEP	STRATEGIC	SURFACE	ALPHA
DEEP Approach				**.72 (.72)**
Seeking meaning	**.68****	.16*	-.13	.38 (.38)
Relating ideas	**.80****	.15	-.05	.49 (.49)
Use of evidence	**.70****	.18*	-.06	.21 (.23)
Interest in ideas	**.76****	.22**	-.29**	.38 (.40)
STRATEGIC Approach				**.76 (.76)**
Organised studying	.07	**.77****	.07	.34 (.31)
Time management	.10	**.74****	.21**	.71 (.71)
Alertness to assessment demands	.08	**.60****	.08	.49 (.48)
Achieving	.26**	**.68****	-.01	.16 (.19)
Monitoring effectivness	.37**	**.65****	.16	.33 (.33)
SURFACE/APATHETIC Approach				**.65 (.65)**
Lack of purpose	-.08	-.01	**.67****	.77 (.77)
Unrelated memorising	-.16*	.19*	**.66****	.31 (.31)
Syllabus boundness	-.19*	-.04	**.40****	.36 (.37)
Fear of failure	-.04	.19*	**.71****	.52 (.54)
Preferences for learning environments				
Supporting understanding(DEEP)	.40**	.07	-.16*	.66 (.66)
Transmitting information (SURFACE)	-.14	.11	.20*	.72 (.73)
DEEP Approach	1	.24**	-.18*	
STRATEGIC Approach		1	.12	
SURFACE/APATHETIC Approach			1	

Y2 STUDENTS (2007-08 CLASS), N= 98, 73.2% of the class				
	DEEP	STRATEGIC	SURFACE	ALPHA
DEEP Approach				**.82 (.83)**
Seeking meaning	**.76****	.38**	-.15	.66 (.66)
Relating ideas	**.77****	-.02	-23*	.49 (.47)
Use of evidence	**.85****	.21*	-.09	.51 (.55)
Interest in ideas	**.76****	.17	-.21*	.68 (.69)
STRATEGIC Approach				**.89 (.89)**
Organised studying	.21*	**.85****	-.18	.69 (.69)
Time management	.09	**.85****	-.18	.81 (.81)
Alertness to assessment demands	.18	**.58****	.02	.66 (.66)
Achieving	.19	**.83****	-.18	.70 (.69)
Monitoring effectivness	.30**	**.74****	-.14	.69 (.69)
SURFACE/APATHETIC Approach				**.80 (.79)**
Lack of purpose	-.21*	-.33**	**.58****	.70 (.72)
Unrelated memorising	-.22*	.02	**.79****	.56 (.56)
Syllabus boundness	-.20*	-.19	**.63****	.65 (.64)
Fear of failure	-.01	.02	**.75****	.82 (.82)
Preferences for learning environments				
Supporting understanding(DEEP)	.56**	-.01	-.38**	.72 (.72)
Transmitting information (SURFACE)	-.07	.13	.34**	.71 (.72)
DEEP Approach	1	.24*	-.22*	
STRATEGIC Approach		1	.18	
SURFACE/APATHETIC Approach			1	

Table 7.2. Factor loadings and alphas for the various scales for the year 1 and year 2 classes of the year 2007-08. For all values ** indicates a significance at 0.01 (two tailed) and * at 0.05 level.

Y1 STUDENTS (2008-09 CLASS), N= 221, 71% of the class				
	DEEP	STRATEGIC	SURFACE	ALPHA
DEEP Approach				**.85 (.85)**
Seeking meaning	**.79****	.47**	-.02	.53 (.53)
Relating ideas	**.87****	.28**	-.08	.66 (.66)
Use of evidence	**.82****	.33**	.09	.58 (.58)
Interest in ideas	**.79****	.42**	-.12	.72 (.72)
STRATEGIC Approach				**.83 (.83)**
Organised studying	.26**	**.81****	.05	.51 (.51)
Time management	.20**	**.80****	.02	.72 (.72)
Alertness to assessment demands	.41**	**.62****	.09	.46 (.44)
Achieving	.38**	**.82****	-.05	.65 (.65)
Monitoring effectivness	.51**	**.73****	-.02	.63 (.62)
SURFACE/APATHETIC Approach				**.72 (.72)**
Lack of purpose	-.13	-.19**	**.58****	.62 (.63)
Unrelated memorising	-.07	.10	**.78****	.61 (.60)
Syllabus boundness	-.08	-.06	**.56****	.55 (.55)
Fear of failure	.12	.16*	**.75****	.78 (.78)
Preferences for learning environments				
Supporting understanding(DEEP)	.60**	.23**	-.29**	.65 (.65)
Transmitting information (SURFACE)	-.21*	.01	.26**	.69 (.70)
DEEP Approach	1	.46**	-.04	
STRATEGIC Approach		1	.02	
SURFACE/APATHETIC Approach			1	

Y2 STUDENTS (2008-09 CLASS), N= 120, 94.4% of the class				
	DEEP	STRATEGIC	SURFACE	ALPHA
DEEP Approach				**.83 (.84)**
Seeking meaning	**.75****	.46**	-.22*	.57 (.57)
Relating ideas	**.84****	.24**	-.25**	.64 (.65)
Use of evidence	**.75****	.33**	-.04	.53 (.54)
Interest in ideas	**.78****	.38**	-.37**	.68 (.70)
STRATEGIC Approach				**.89 (.89)**
Organised studying	.29**	**.85****	-.22*	.52 (.53)
Time management	.26**	**.85****	-.10	.80 (.81)
Alertness to assessment demands	.24**	**.62****	-.09	.69 (.69)
Achieving	.54**	**.80****	-.35**	.73 (.74)
Monitoring effectivness	.46**	**.83****	-.21*	.61 (.61)
SURFACE/APATHETIC Approach				**.82 (.83)**
Lack of purpose	-.36**	-.35**	**.77****	.81 (.82)
Unrelated memorising	-.14	-.15	**.72****	.58 (.59)
Syllabus boundness	-.24**	-.29**	**.68****	.61 (.59)
Fear of failure	-.09	.07	**.72****	.78 (.79)
Preferences for learning environments				
Supporting understanding(DEEP)	.59**	.29**	-.38**	.74 (.74)
Transmitting information (SURFACE)	-.26**	.06	.29**	.62 (.61)
DEEP Approach	1	.45**	-.29**	
STRATEGIC Approach		1	-.24**	
SURFACE/APATHETIC Approach			1	

Table 7.3. Factor loadings and alphas for the various scales for the year 1 and year 2 classes of the year 2008-09. For all values ** indicates a significance at 0.01 (two tailed) and * at 0.05 level.

Group differences

When at the overall scores for the three approaches are examined, there are two aspects to be noted: the differences between groups within the samples and the variation of scores across year 1 and year 2 for a longitudinal sub-sample.

Tables 7.4 and 7.5 show a summary of the descriptive statistics and relevant tests to identify differences between genders, country of origin and year classes for the samples considered.

Year 1	Males				Females						
	N	Range	Mean	SD	N	Range	Mean	SD	df	one-way F	p
DEEP Approach	104	62.5	74.17	12.31	338	85	70.1	11.75	1, 444	9.32	0.002
STRATEGIC Approach	104	59	63.16	11.34	338	76	64.51	11.46			ns
SURFACE/APATHETIC Approach	104	56.25	60.6	11.79	338	77.5	62.29	10.96			ns
Preferences for learning environments	104				338						
Supporting understanding(DEEP)		14	16.05	3.16		11	17.35	2.39			ns
Transmitting information (SURFACE)		16	14.63	3.37		15	14.14	2.84	1, 440	19.95	0.001

Year 2	Males				Females						
	N	Range	Mean	SD	N	Range	Mean	SD	df	one-way F	p
DEEP Approach	72	47.5	74.79	9.87	228	66.25	72.37	10.51			ns
STRATEGIC Approach	72	50	61.93	12.27	228	60	69.21	12.12	1, 300	21.09	0.001
SURFACE/APATHETIC Approach	72	61.25	60.9	12.12	228	61.25	60.06	11.9			ns
Preferences for learning environments	72				228						
Supporting understanding(DEEP)		13	16.64	2.72		9	17.46	2.37			ns
Transmitting information (SURFACE)		12	14.99	2.81		16	14.36	3.1	1, 298	6.17	0.014

Table 7.4. Average scores of the three approaches grouped by year-level and gender. Differences tested using one-way ANOVAs are reported in the rightmost column

Year 1	UK				EU				World						
	N	Range	Mean	SD	N	Range	Mean	SD	N	Range	Mean	SD	df	one-way F	p
DEEP Approach	370	67.5	70.92	11.43	42	52.5	76.25	10.24	30	51.25	70.96	10.34	2, 443	4.44	0.012
STRATEGIC Approach	370	64	64.99	10.73	42	53	61.59	12.43	30	46	63	9.69			ns
SURFACE/APATHETIC Approach	370	52.5	62.71	10.07	42	50	60.8	12.95	30	47.5	59.25	10.16			ns
Preferences for learning environments	370				42				30						
Supporting understanding(DEEP)		13	17.2	2.55		14	16.3	3.32		9	16.07	2.46	2, 439	4.19	0.016
Transmitting information (SURFACE)		16	14.14	2.99		11	15.5	2.65		12	13.93	2.86	2, 439	4.26	0.015

Year 2	UK				EU				World						
	N	Range	Mean	SD	N	Range	Mean	SD	N	Range	Mean	SD	df	one-way F	p
DEEP Approach	264	67.5	72.24	10.38	21	36.25	78.63	9.32	15	28.75	77.58	8.72	2, 299	3.8	0.025
STRATEGIC Approach	264	58	67.34	12.56	21	41	69.24	10.36	15	56	67.13	15.22			ns
SURFACE/APATHETIC Approach	264	60	63.05	11.21	21	55	56.55	14.84	15	52.5	62.08	14.44	2, 299	4.29	0.014
Preferences for learning environments	264				21				15						
Supporting understanding(DEEP)		11	17.31	2.42		13	16.43	3.32		7	17.73	2.12			ns

Table 7.5. Average scores of the three approaches grouped by year-level and country of origin. Differences tested using one-way ANOVAs are reported in the rightmost column

An interesting result is the fact that males' scores are significantly higher than females in the deep approach (year 1) and surface approach (year 2). The higher score in the preference for the learning environment tending toward transmission of information was also significantly higher in males in both year 1 and year 2.

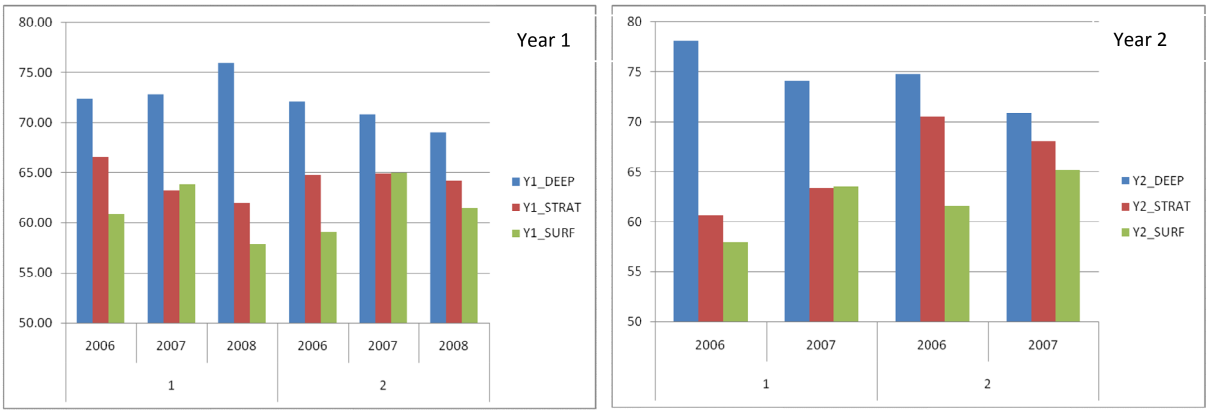


Figure 7.1. Change in the scores of the different approaches to learning for the cohort starting in 2006-08 differentiated by gender

(1=males, 2=females). On the left the scores in year 1 and the corresponding scores in year 2 on the right. (Comparisons are within subjects, N=103).

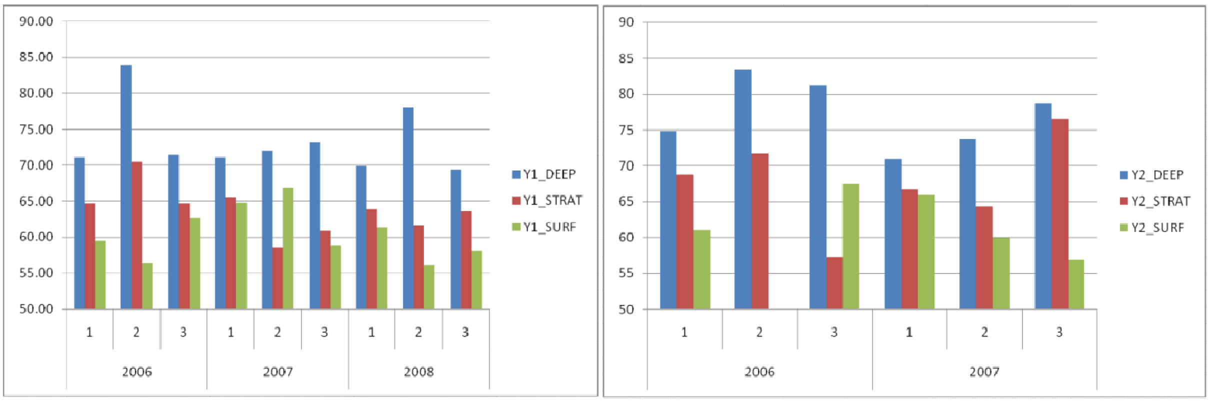


Figure 7.2. Change in the scores of the different approaches to learning for the cohorts starting in 2006-08 differentiated by country of origin

(1=UK, 2= EU, 3=World). On the left the scores in year 1 and the corresponding scores in ye ar 2 on the right (N=103).

As demonstrated in Figure 7.1, looking at each year in turn, there is a difference between the way in which males and females approach studying psychology. The general decline in the deep approach to learning over the years in females with an apparent increase of the surface approach is a worrying feature. Males present the opposite trend, with an increase in the scores for the deep approach.

In the year 2, similar patterns to year 1 emerge for females, with a steady decline of the type of approach reported. Looking at the differences between the two years, there is a small improvement in the expected direction of a more focused approach to learning in which a deep approach is favoured.

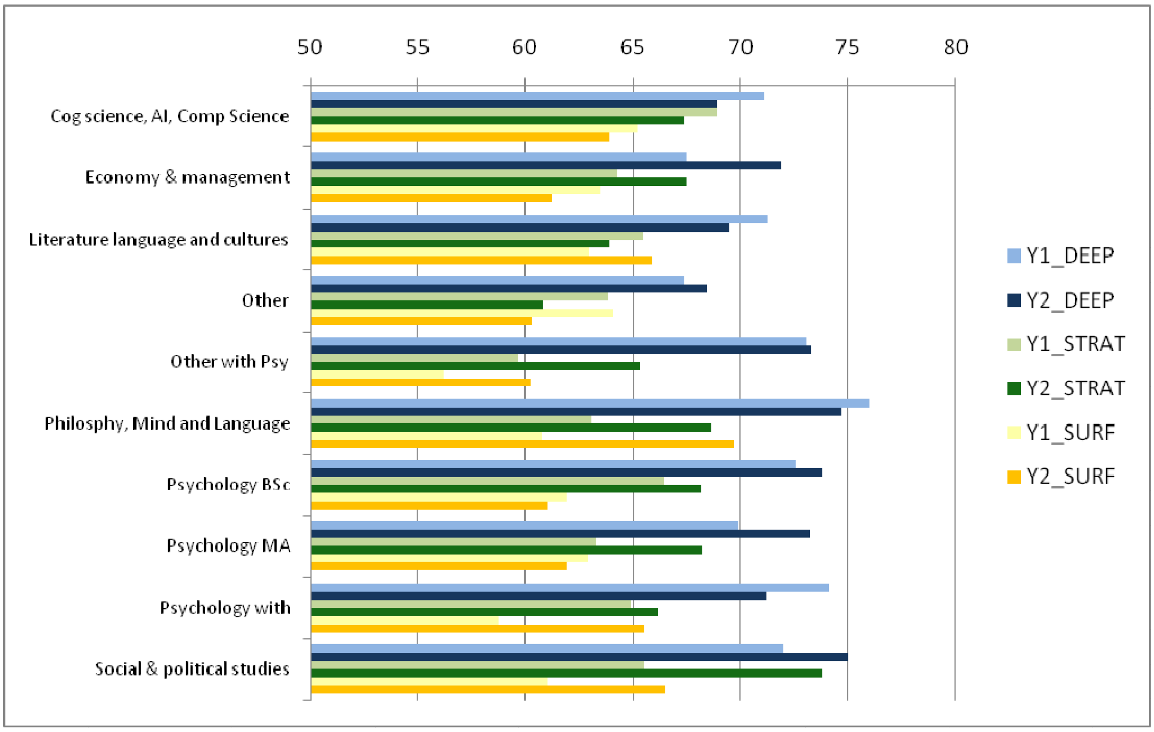


Figure 7.3. Change in the scores of the different approaches to learning for the cohorts starting in 2006-08 differentiated by type of degree.

It should be noted an increase of the scores in both the deep and strategic approach for psychology students between year 1 and year 2 (N=103).

Degree Paths	Year 1		Year 2	
	one-way F	p	one-way F	p
DEEP Approach	1.92	0.048	-	ns
STRATEGIC Approach	1.87	0.055	-	ns
SURFACE/APATHETIC Approach	2.55	0.007	-	ns
Preferences for learning environments				
Supporting understanding(DEEP)	3.49	0.001	-	ns
Transmitting information (SURFACE)	6.05	0.001	-	ns

Table 7.6. Differences in the scores of the three approaches according to the degree types.

When looking at the mean scores between different countries of origin (Figure 7.2), it is quite evident that EU students have higher scores in the deep approach than their UK counterparts, trend which is replicated in the Y2 students. These were found to be significant (table 7.6).

The pattern could be a useful indicator of the fact that foreign students might approach their studies in a different way: their motivation and dedication is expected to be higher due to the pressure to do well (this might be elicited by both parents and government agencies providing funding), but it could be equally argued that students going abroad might already be more dynamic in the first place.

A ‘deep approach’ does not necessarily guarantee better performance, and might not be the most efficient way to achieve good results either, but certainly is an indicator that different *types* of students have different ways to achieve their goals.
By looking at the scores in the different dimensions according to the degree path, trends are not at all clear. As can be seen in Figure 7.3 there is an increase in scores in most cases in both the Deep and Strategic approach and a decline in the surface approach. This is partially expected as students in 2nd year need to attain specific performance levels to be allowed to enter into Honours. Not only do they need to become more efficient in the use of resources and to manage their workload, but they have to focus on specific minimum targets which promote the acquisition of more effective learning approaches.

From the observed trends, there is therefore no obvious characterization of the typology of students when considering the approaches to learning. On one hand there are differences between genders and origin. On the other there are significant differences between degree paths (table 7.6), especially in year 1, which confirm that the year 1 cohort is quite heterogeneous.

	Year 1 N	Range	Mean	SD	Year 2 N	Range	Mean	SD	df	t	p
DEEP Approach	**446**	**85.00**	**71.06**	**11.99**	**302**	**67.50**	**72.77**	**10.60**	**102**		**ns**
Seeking meaning	446	16.00	14.30	2.72	302	14.00	14.56	2.68	102		ns
Relating ideas	446	16.00	13.80	3.15	302	15.00	14.11	2.72	102		ns
Use of evidence	446	18.00	14.20	2.82	302	13.00	14.45	2.59	102		ns
Interest in ideas	446	18.00	14.55	3.26	302	16.00	15.10	2.96	102		ns
STRATEGIC Approach	**446**	**80.00**	**64.19**	**11.43**	**302**	**63.00**	**67.31**	**12.65**	**102**	**-3.95**	**0.001**
Organised studying	446	16.00	12.11	3.07	302	16.00	13.08	3.36	102	-4.2	0.001
Time management	446	19.00	11.26	3.80	302	16.00	11.53	3.97	102		ns
Alertness to assessment demands	446	18.00	13.78	2.87	302	14.00	14.44	3.04	102	-4.65	0.001
Achieving	446	17.00	12.63	3.03	302	14.00	13.58	3.10	102	-2.97	0.004
Monitoring effectivness	446	16.00	14.41	2.94	302	16.00	14.69	2.86	102		ns
SURFACE/APATHETIC Approach	**446**	**78.75**	**61.89**	**11.17**	**302**	**62.50**	**62.41**	**11.85**	**102**		**ns**
Lack of purpose	446	18.00	9.98	3.84	302	15.00	9.29	3.44	102		ns
Unrelated memorising	446	18.00	11.68	3.01	302	15.00	12.12	2.97	102	-2.41	0.018
Syllabus boundness	446	17.00	14.13	3.07	302	16.00	14.06	2.96	102	-2.76	0.007
Fear of failure	446	18.00	13.73	3.88	302	16.00	14.46	4.01	102	-2.66	0.009
Preferences for learning environments											
Supporting understanding(DEEP)	442	14.00	17.04	2.65	300	13.00	17.27	2.48	102		ns
Transmitting information (SURFACE)	442	16.00	14.25	2.98	300	16.00	14.51	3.04	102		0.04

Table 7.7. Average scores of the three approaches and subscales divided by year-level (longitudinal sample N=103).

Very similar patterns of scores are present in both years. Post-hoc paired-sample t-tests feature on the rightmost column.

To determine whether these differences are important in the creation of clusters for the sample, we focused specifically on the longitudinal sample of 103 students. Table 7.7 shows the descriptive statistics for each subscale. After conducting a mixed ANOVA with the class-years, gender, country and degree as between subjects factors, it was found that for this sample only the interaction between approaches and country of origin was significant ($F_{(2,94)}$=3.49, p<.05) and although the differences between year-classes was marginally not significant ($F_{(1,94)}$=2.98, p=.08), post-hoc, paired-sample T-tests allowed the identification of significant differences between some of the subscales and a difference in the strategic approach (table 7.7).

	DEEP Mean	DEEP Median	DEEP SD	STRATEGIC Mean	STRATEGIC Median	STRATEGIC SD	SURFACE Mean	SURFACE Median	SURFACE SD
Cluster 1	63.96	65.0	10.09	61.85	62.0	10.41	67.75	67.5	9.00
Cluster 2	70.72	71.3	7.96	63.18	64.0	10.42	63.89	63.8	10.00
Cluster 3	70.11	70.0	10.50	65.53	67.0	11.52	63.10	63.8	9.28
Cluster 4	66.59	67.5	9.65	64.31	65.0	9.73	66.87	66.3	9.13
Cluster 5	80.36	80.0	7.68	65.57	66.0	11.22	56.46	56.3	11.30
Cluster 6	81.70	81.3	6.57	81.64	80.0	7.50	55.34	55.6	10.51

Table 7.8. Cluster centroids for the three approaches to learning scales.

Mining approaches to learning

The knowledge of some significant differences and variations between the subscales prompted us to consider the subscales rather than the constructs for data mining. The purpose was to identify a number of suitable groups to categorise students. Originally, the need for using clustering techniques was dictated by the fact that McCune and colleagues stressed the fact that the scores should not be used prescriptively as normative measures. Furthermore, the variation of scores across multiple samples (justified mainly with the variation caused by the type of courses and the context) was another issue in the debate. Differences were found in our samples, making it difficult to provide a simple representation of the approaches without a more analytic study of the context. The question, however, is whether styles, as measured by the ASSIST, could be useful in practice as a tool to discriminate between different *types* of students. A first attempt to provide a meaningful categorization was achieved running a Two-steps cluster analysis on the subscales of the ASSIST using PASW 17[14]. This generated a solution with 6 groups. The mean scores of both scales (Fig 7.8) and subscales (Fig. 7.9) for the clusters are provided in table 7.8.

A visual representation of the mean scores of each cluster is provided in figure 7.5 (next page). The radial graphs are quite useful in identifying a ‘typical’ profile for a student in each cluster. It should be noted that clusters in this procedures are created using gender, country of origin, degree type and year class in addition to all the subscales. A log-likelihood distance measure between scores was used as it can handle both continuous and categorical variables. The distances between centroids of each cluster are tested using either t-tests or chi square. A visual representation of the t-scores is useful to gain an insight in the process (Figure 7.4). This clearly shows the direction of the differences present in the two clusters and the colour codes are marking the three approaches.

This solution using clusters is considered a better option than a statistical binning based on fixed threshold values using, for example, a median or quartile split (Witten & Frank 2005, Hand et al. 2006). In Vigentini (2009) the latter criterion was adopted successfully to determine differences in groups of performance for both online usage and styles. Whilst differences were found in the comparison of the top and bottom groups, a large middle section of the sample could not be easily characterised. In a later section we will be able to

---

[14] A more detailed explanation of the Two-Steps cluster algorithm is provided in Appendix 2.

demonstrate if the use of the six groups generated with this technique will be effective in providing greater discrimination between students' types.

Centroids of the subscales for clusters			1	2	3	4	5	6
DEEP Approach	*Seeking meaning*	*DE1*	13.45	13.80	14.03	13.58	15.65	16.94
	Relating ideas	*DE2*	12.07	13.98	13.69	12.45	16.15	15.21
	Use of evidence	*DE3*	13.34	14.06	13.56	13.56	15.86	15.91
	Interest in ideas	*DE4*	12.30	14.73	14.81	13.68	16.63	17.30
STRATEGIC Approach	*Organised studying*	*ST1*	11.61	12.03	12.34	12.64	12.42	15.98
	Time management	*ST2*	11.40	10.31	10.93	11.66	10.90	15.56
	Alertness to assessment demands	*ST3*	13.34	14.19	14.40	13.47	14.15	15.89
	Achieving	*ST4*	12.02	12.68	13.05	12.39	13.07	17.06
	Monitoring effectivness	*ST5*	13.49	13.98	14.82	14.15	15.02	17.14
SURFACE/APATHETIC Approach	*Lack of purpose*	*SU1*	12.34	9.87	8.77	10.75	8.94	6.39
	Unrelated memorising	*SU2*	12.61	12.45	12.58	12.51	10.36	11.33
	Syllabus boundness	*SU3*	14.62	14.65	15.10	14.50	13.37	12.27
	Fear of failure	*SU4*	14.64	14.14	14.03	15.74	12.50	14.27
Preferences for learning environments	*Supporting understanding(DEEP)*	*UND*	12.89	14.24	13.89	13.29	15.86	16.24
	Transmitting information (SURFACE)	*TRA*	17.03	17.33	18.05	17.92	15.90	17.09

Table 7.9. Average scores (standardised scales) for the three approaches to learning in the different clusters.

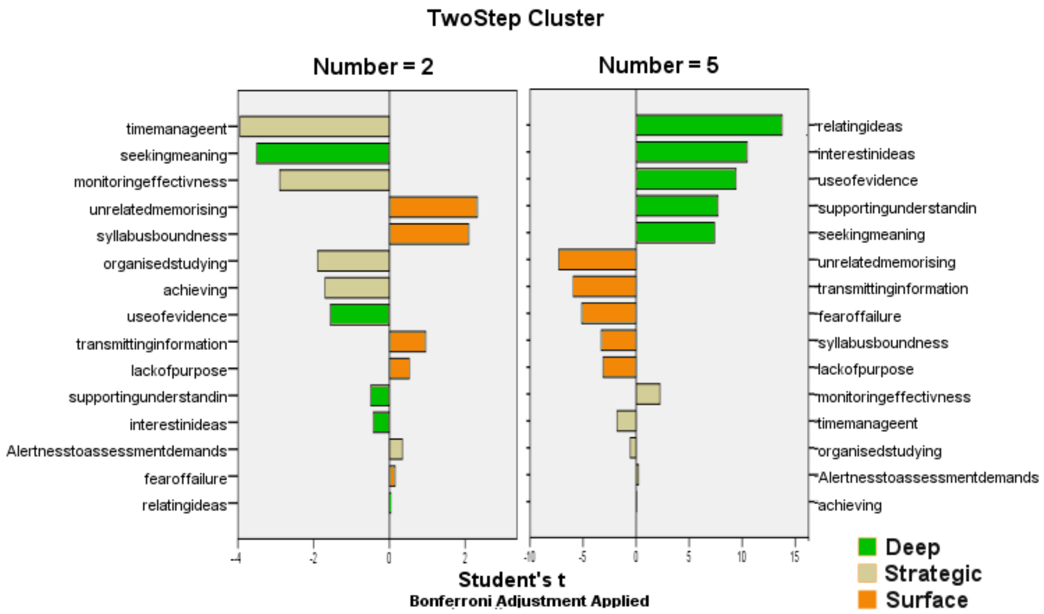


Figure 7.4. Visual representation of the direction and size of the differences between groups using paired-samples t-tests values (Bonferroni adjustment) for the ASSIST subscales relevant to cluster number 2 and 5 (t values ranked, therefore left and right variables might be offset).

As clusters are explicitly formed to maximise differences, this graph is only useful to show which attributes might be describing the specific cluster. In the example on the left features are not as clearcut as in the example on the right in which key subscales discriminate the clusters.

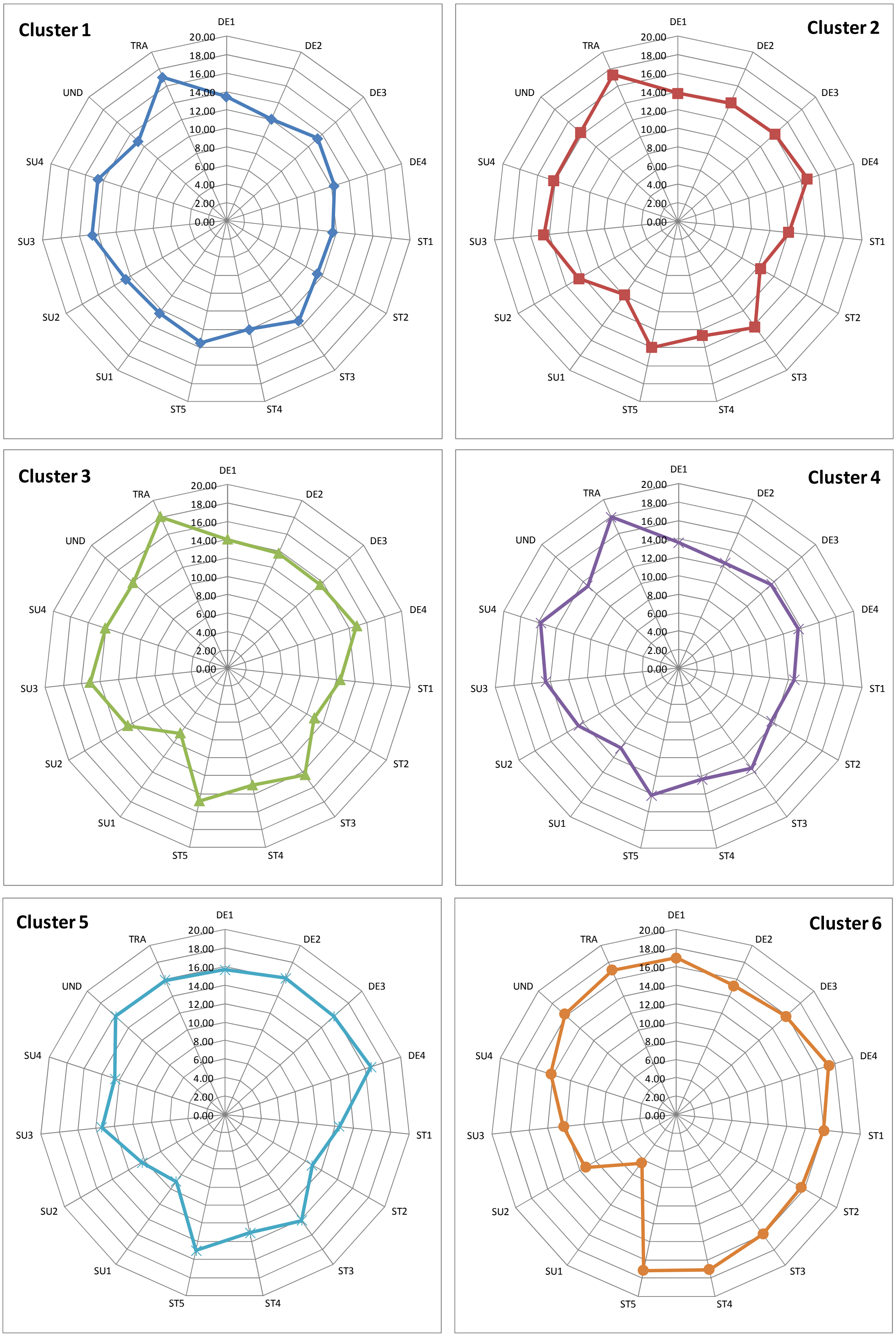


Figure 7.5. Characterisation of the means scores for the ASSIST subscales in the six clusters

Summary for the findings with the ASSIST

The ASSIST proved to be a reliable instrument to measure differences in the approaches to learning. Results largely overlapped with the findings reported by McCune et al. (2000), but interesting patterns were pointed out regarding the differences between genders, country of origin and year-class.

Using a data mining approach to data reduction, we also identified six groups based on the intrinsic features of the student's and suggested a visual (radial graphs) and numerical (Two-steps cluster analysis using sex, country, degree path and year as well as all the ASSSIST subscales) way to determine the *typicality* of a students' approaches to learning. This grouping will be used in a later section to compare both performance and usage and determine if such grouping is effective in characterising the students.

The same technique will be applied to the other measures of styles offering a richer discretisation of styles.

7.1.2. Mental self-government theory (MSG)

The MSG is certainly the most ambitious of all the theories considered in this thesis because it is presented as a comprehensive theory of intellectual styles. The work started by Sternberg was continued assiduously by Zhang who disseminated findings from research with a clear cross-cultural stance. The original appeal was that the Thinking Styles Inventory (TSI) has been used at university level in conjunction to a number of other learning styles measures. The debate about its reliability, however, is quite heated, with researchers outside the Sternberg circle more cautious about its validity (see chapter 4 for more details).

In the context of this thesis, the TSI has been used with two Y1 classes and in one Y2 class therefore there is no longitudinal data available for this test, however it was used concurrently with ASSIST, CSI and VICS-WA providing a unique opportunity for cross-validation. Furthermore we will look more closely to the support for the threefold model of intellectual styles presented in chapter 4.

Validity and reliability

As observed in chapters 4 and chapter 5, claims about the high reliability of the instrument seem to come mainly from figures reported by Sternberg and associates, but Coffield et al. (2004) were dismissive of the instrument.

Table 7.10 provides an overview of the reliability scores for the subscales in our samples. Only 5 of the figures were considered unacceptable (alpha below .5), many are average reliabilities (below .7) and about half for each sample can be considered good. Given that the samples would be expected to be similar (especially in the comparison of the two year 1 classes), scales like the Monarchic and Democratic, in which reliability is consistently below par, could offer evidence to argue that there are weaknesses in the item structure of the questionnaire.

The TSI is very long (112 statements) compared with other styles measures and it was argued that the length could be a major drawback for the reliability of the tool. Looking at the overall performance of the test, by removing the democratic scale (not included in the original TSI) the average alphas range between .38 and .85.
These figures are very similar in magnitude to those reported by Zhang (2001, 2008) and confirm a fair reliability for the instrument as a whole.

			Scale statistics						Reliability analysis		
			Y1 class 2007/08		Y2 class 2007/08		Y1 class 2008/09		alpha 1 (N=177)	alpha 2 (N=110)	alpha 3 (N=164)
dimension	scale	items	Mean	SD	Mean	SD	Mean	SD			
Function	Legislative	8	4.54	0.31	4.69	0.41	4.79	0.23	0.799	0.796	0.819
	Executive	8	4.65	0.31	4.70	0.54	4.81	0.34	0.710	**0.649**	0.845
	Judicial	8	4.51	0.46	4.72	0.43	4.57	0.57	**0.544**	**0.452**	0.712
Level	Global	8	4.44	0.23	4.70	0.36	4.39	0.67	**0.640**	**0.509**	**0.683**
	Local	8	4.04	0.50	4.10	0.74	4.14	0.53	**0.528**	**0.560**	**0.606**
Leaning	Progressive	8	4.69	0.30	4.90	0.42	4.45	0.39	0.787	0.757	0.888
	Conservative	8	4.63	0.49	4.77	0.65	4.49	0.36	**0.666**	**0.641**	0.810
Form	Hierarchical	8	4.89	0.35	5.04	0.49	5.02	0.16	**0.681**	**0.512**	0.795
	Monarchinc	8	4.37	0.53	4.48	0.64	4.24	0.55	**0.434**	**0.507**	**0.257**
	Oligarchic	8	3.95	0.46	4.00	0.50	4.16	0.40	0.779	**0.630**	0.802
	Anarchic	8	4.46	0.38	4.58	0.39	4.42	0.51	0.763	0.805	**0.581**
	Democratic	8	4.14	0.37	4.27	0.43	4.76	0.40	**0.408**	**0.258**	**0.653**
Scope	Internal	8	4.44	0.34	4.57	0.41	4.36	0.49	0.818	0.820	0.752
	External	8	4.69	0.23	4.95	0.26	4.70	0.30	0.851	0.804	0.843

Table 7.10. Reliability and descriptive statistics for the dimensions of the MSG.

Cronbach alpha values are reported on the left highlighting unacceptable levels in red (<.4) and average levels in green (<.7 and > .5).

With regard to the interrelation between the scales, table 7.11 provides full details for each subscale in the three samples. These correlations are useful to identify common patterns of overlapping measures (i.e. those with a high correlation) because these could arguably be measuring the same constructs. There seem to be some cases with moderate correlations, however these do not appear dissimilar from the figures reported by Zhang in both American and Chinese samples. To test the overall structure of the TSI, we conducted an exploratory factor analysis on the scales. Using an un-rotated principal component analysis all three procedures resulted in a 5 factors model of which we provide factor loadings in table 7.12.

Y1 class 2007/08	n=182	LEGL	EXEC	JUDI	GLOB	LOCL	PROG	CONS	HIER	MONA	OLIG	ANAR	DEMO	INTR	EXTR
Function	LEGL														
	EXEC														
	JUDI		.436**												
Level	GLOB			.302**											
	LOCL	.191**	.261**	.186*	.177*										
Leaning	PROG			.354**	.345**										
	CONS		.472**	.219**			.208**								
Form	HIER	.151*	.481**	.292**				.404**							
	MONA		.257**	.176*	.278**		.245**	.406**	.395**						
	OLIG				.274**	.283**	.187*	.309**							
	ANAR		.291**	.527**	.197**		.448**	.252**		.230**					
	DEMO	.210**	.193**			.342**	-.156*		.211**			-.241**	.251**		
Scope	INTR	.518**			.200**	.473**						-.320**			
	EXTR	-.161*		.205**	.425**		.444**	.212**		.210**	.410**	.242**	-.202**		

Y2 class 2007/08	n=114	LEGL	EXEC	JUDI	GLOB	LOCL	PROG	CONS	HIER	MONA	OLIG	ANAR	DEMO	INTR	EXTR
Function	LEGL														
	EXEC														
	JUDI		.341**												
Level	GLOB			.185*											
	LOCL				.211*										
Leaning	PROG			.397**	.217*										
	CONS		.364**				.271**								
Form	HIER	.279**	.233*	.216*				.324**							
	MONA			.319**	.228*			.453**	.314**						
	OLIG		.184*		.296**	.315**	.207*	.364**							
	ANAR		.367**	.613**			.507**			.257**					
	DEMO		.274**		.207*			.223*			.454**			.286**	
Scope	INTR	.403**	-.426**	-.251**	.281**	.337**					.195*	-.292**			
	EXTR				.342**		.500**	.361**		.212*	.457**				

Y1 class 2008/09	n=169	LEGL	EXEC	JUDI	GLOB	LOCL	PROG	CONS	HIER	MONA	OLIG	ANAR	DEMO	INTR	EXTR
Function	LEGL														
	EXEC														
	JUDI	.503**													
Level	GLOB	.153*	-.205**												
	LOCL		.494**		-.439**										
Leaning	PROG	.650**	-.317**	.522**	.166*										
	CONS	-.317**	.733**	-.161*	-.209**	.471**	-.460**								
Form	HIER	.376**	.444**	.471**		.207**	.250**	.213**							
	MONA		.290**		.217**	.218**		.308**	.250**						
	OLIG		.264**			.316**		.307**		.202**					
	ANAR	.462**		.276**	.186*		.546**	-.169*	.176*						
	DEMO	.348**		.374**	.169*	.153*	.484**		.439**		.366**	.549**		.591**	
Scope	INTR	.455**				.163*			.173*	.245**	-.231**	.169*			
	EXTR				.174*		.287**		.245**		.445**	.311**	-.564**		

Table 7.11. Intercorrelations of the TSI subscale in the three samples considered. Only significant values are reported. p values are two-tailed (p<.01, *p<.05levels).**

This result does not match with the 3-components structure proposed by Zhang and the inconsistency of the factor loadings questions whether we could talk about a coherent underlying structure at all.

Arguably, a confirmatory factor analysis could be tried in order to attempt to fit the model to the one proposed by Zhang and Sternberg. However, because of the very different structure in the two samples considered it is difficult to imagine how forcing the model on our data could lead to the confirmation of the theory.

		Y1 class 2007/08					Y1 class 2007/08					Y1 class 2007/08				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Function	Legislative		.617		.457				.683			.758				
	Executive	.544					.436				.517		.668			
	Judicial	.676					.670					.546	.470			
Level	Global	.479		.583			.404	.497							.786	
	Local		.603					.492			.638		.623			
Leaning	Progressive	.555			.421		.671					.839				
	Conservative	.625				.421	.625			.612			.500			
Form	Hierarchical	.484							.627	.456			.735			
	Monarchinc	.595				.424	.556						.460		.682	
	Oligarchic	.417		.449				.532					.455	.551		
	Anarchic	.608					.667					.575				.491
Scope	Internal		.792					.784	.410							
	External	.533		.425			.556	.400						.827		

Table 7.12. Factor loading of the scales of the TSI with a resulting 5-components model, all loadings less than .4 have been left out.

Post-hoc	T-test	Y107 v Y207			Y207 v Y108			Y107 v Y108		
		t	df	p	t	df	p	t	df	p
Function	LEGL							-2.77	349	.006
	EXEC									
	JUDI	-2.87	294	.004	1.83	281	.069			
Level	GLOB	-3.25	294	.001	4.04	281	.000			
	LOCL									
Leaning	PROG	-2.67	294	.008	4.60	281	.000	2.71	349	.007
	CONS				3.18	281	.002	2.14	349	.033
Form	HIER	-2.06	294	.041						
	MONA				3.38	281	.001	2.03	349	.043
	OLIG							-2.28	349	.023
	ANAR				-6.49	281	.000	-9.05	349	.000
	DEMO				*1.94*	*281*	*.054*			
Scope	INTR				2.11	281	.036			
	EXTR	-2.70	294	.007	2.70	281	.007			

Table 7.13. Post-hoc t-tests on the differences between the scores in each subscale for the 3 samples considered.

ANOVA		sex		sample		degree		country		sex*sample		sex*degree		sex*country		sex*country*degree	
		F	p	F	p	F	p	F	p	F	p	F	p	F	p	F	p
Function	LEGL	8.90	.003	10.65	.000			6.38	.002	6.87	.001					3.02	.050
	EXEC																
	JUDI	10.62	.001	6.77	.001					3.16	.043						
Level	GLOB			3.54	.030												
	LOCL																
Leaning	PROG	6.73	.010	4.60	.011			3.07	.048	3.23	.041						
	CONS			6.27	.002	*1.89*	*.052*										
Form	HIER			4.43	.012			5.20	.006	*2.99*	*.051*						
	MONA			4.64	.010	2.18	.023							*2.86*	*.058*		
	OLIG																
	ANAR	14.37	.000	5.47	.005					3.64	.027						
	DEMO							4.95	.008	4.11	.017						
Scope	INTR	5.74	.017									2.25	.018	4.36	.013		
	EXTR			24.33	.000												

Table 7.14. Significant differences in the scores of the various MSG subscales. On the left the results of the ANOVA comparing the scores for each scale according to gender and the sample. (Note in italics are marginally non-significant values)

Group differences

As for the ASSIST, we looked for differences in the samples. Table 7.14 shows a summary of the significant differences between groups after running a mixed ANOVA on the scores with sex, sample, degree and country of origin as between subjects' factors. Table 7.13 provide the details for the post-hoc paired sample t-tests to identify where the differences between the scores are to be found. Although the differences in gender, country and degree were partially expected, more unsettling are the differences between the samples which, given the similarity of the students taking the course are puzzling. It is possible that the way of thinking in the two cohorts starting in 2007 and in 2008 is different in nature, but the question then is how much of what the TSI measure is explained by generational factors.

What remain certain from the analysis conducted so far is that there is plenty of variation, not only within subjects, but also between subjects. Whilst this could provide reasons for concern for traditional statistical methods (i.e. in regression, when considering the relations between AP and styles), the variation could be useful in the identification of suitable groupings using clustering.

Mining intellectual styles

Greater variance means that a wider range of scores is present: this is not ideal for a linear model, but could turn out to be an advantage for clustering. Following a similar procedure and based on the previous analysis, we conducted a Two-steps cluster analysis including gender, sample and country of origin as well as all the subscales. Results are shown in table 7.15 and figure 7.6.

	Gender				Samples					
	Males		Females		Y1 class 2007/08		Y2 class 2007/08		Y1 class 2008/09	
Clusters	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
1	0	.0%	138	38.4%	0	.0%	1	.9%	137	81.1%
2	106	100.0%	10	2.8%	64	35.2%	20	17.5%	32	18.9%
3	0	.0%	118	32.9%	118	64.8%	0	.0%	0	.0%
4	0	.0%	93	25.9%	0	.0%	93	81.6%	0	.0%
Combined	106	100.0%	359	100.0%	182	100.0%	114	100.0%	169	100.0%

Table 7.15. Relative frequencies and percentages of the fit for the 4 different clusters.

The choice of the attributes for clustering was informed by the differences found in the ANOVA in which the sample was the major contributor to the variation in scores and gender differences had an impact on interactions with other variables. Country of origin was removed after a first test as the solution provided the same number of clusters and did not affect the overall distributions of the groups (table 7.13).

The radial charts provide an insight into the typical profiles of students in each of the four clusters. One of the most striking aspects of the profiles presented is that all males are fitting in one profile (cluster 2) whilst females seem to have a wider range of scores. Such pattern is very different from the clusters emerging from the ASSIST in which genders are more evenly distributed over all clusters.

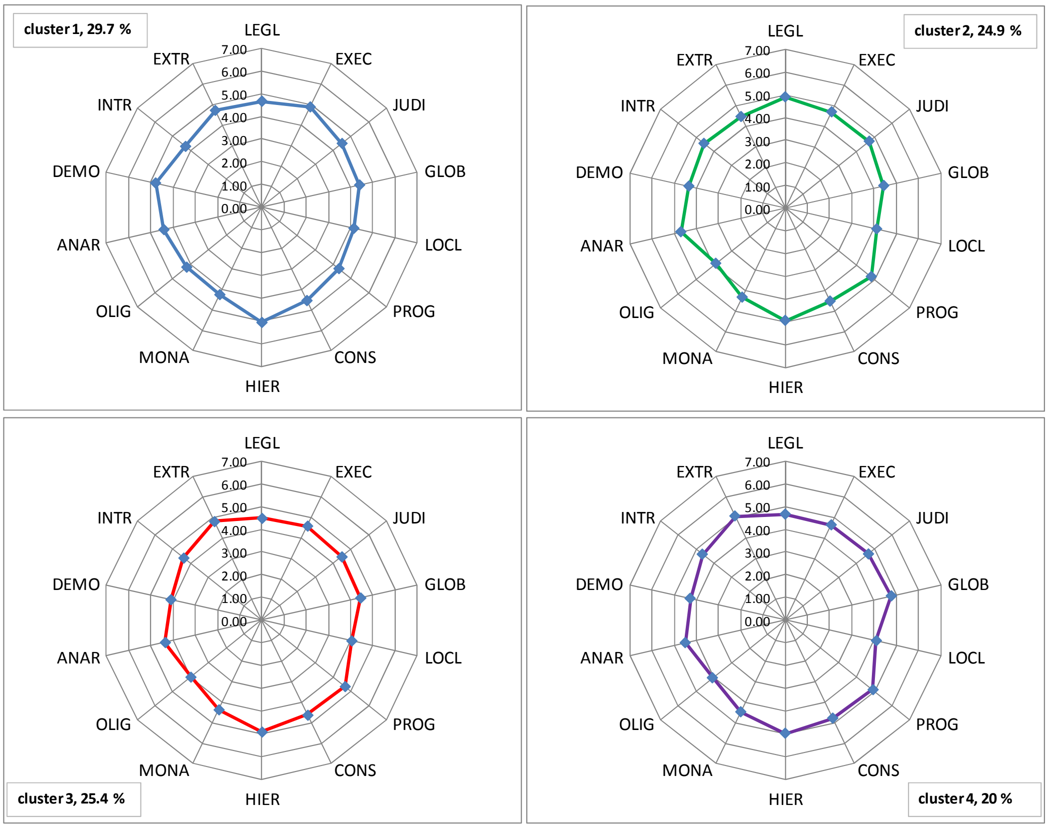


Figure 7.6. Classification of cognitive styles according to gender, samples and TSI scales.

The radial graphs provide a characterization of the mean value scores for a typical students fitting in one of the four clusters.

Summary for the findings with the MSG

Despite the quick dismissal of Coffield et al., the TSI proved to be a fairly reliable instrument. Some doubts are cast on the validity of the constructs measured given the fact that we were unable to replicate the factor structure suggested by Zhang and Sternberg and the fact that we found a number of significant differences between the scores of the subscales in different samples which were expected to be similar (i.e. between first year cohorts starting their courses in 2007 and 2008).

With moderate intercorrelations between scales and the inability to consistently replicate the factor structure suggested, calls for a more in-depth review of the items. Perhaps reducing the number of items would not only be beneficial for participants responding to a shorter inventory, but it could increase the overall reliability of the instrument. One cannot avoid wondering whether the TSI has too many subscales measuring closely related concepts: from our data we cannot confirm that a coherent framework is emerging nor that the constructs are stable. Furthermore, because we didn’t have a longitudinal sample for this test we cannot confirm whether the styles measured are stable over time.

Mining the dataset, however, allowed us to determine 4 clusters for further analysis. As for the ASSIST, we hope that these grouping will provide an effective differentiation between students. From this analysis, at this stage, it is apparent that there is more variation in the profiles of female students compared to men who fit in a single cluster.

7.1.3. Cognitive Styles Inventory (CSI)

Similar to the MSG in the previous section, the CSI was used in two consecutive Y2 classes and in one Y1 class therefore there is no longitudinal data available for this test. As we have seen in chapter 5, although the test is considered reliable, there is an ongoing debate on the complexity of the dimensions measured, with Allinson & Hayes proposing a unidimensional scale encompassing intuition at one end and analysis at the other, and Hodgkinson & Sadler-Smith suggesting that a solution with two orthogonal dimensions is a more appropriate representation of the conceptual framework for this test.

In this section we will present the results using both methods in calculating the scores and testing for reliability of the scales.

Validity and reliability

Starting from the comparison with normative data provided by Allinson & Hayes, the mean scores in our samples are very similar. They are lower than the norms provided for Canadian Students (42.5, who attend a very similar university system) and higher than the other EU counterparts (Germans 35.64, French 37.79).

The alpha levels reported for each scale can be considered satisfactory using the original method to compute the scores. However, if we follow the alternative method as indicated by Hodgkinson et al. the reliability of the scales is increased taking all the alpha values above 0.7.

This seems to provide further support of a two-dimensional construct suggested by the latter. Although the reliability of the scales increases using a two-factors model, the interrelation between the scales is quite high (Pearson $r_{(336)} = -.59$, $p<.01$). This is not clarifying the debate as such correlation does not support the idea of orthogonal measures.

CSI (Allinson & Hayes)

class	N	mean	med	SD	alpha
Y1 (2008-09)	225	41.33	42	7.95	0.65
Y2 (2007-08)	66	40.8	40.5	7.59	0.62
Y2 (2008-09)	64	41.3	42	8.82	0.72

CSI Analysis (Hodgkinson & Sadler-Smith)

class	N	mean	med	SD	alpha
Y1 (2008-09)	225	27.6	28	7.53	0.78
Y2 (2007-08)	66	26.92	28	7.99	0.8
Y2 (2008-09)	64	27.31	27.5	7.79	0.8

CSI Intuition (Hodgkinson & Sadler-Smith)

class	N	mean	med	SD	alpha
Y1 (2008-09)	225	18.93	19	6.58	0.72
Y2 (2007-08)	66	18.35	18.5	6.62	0.74
Y2 (2008-09)	64	18.31	18.5	7.07	0.75

Table 7.16. Descriptive statistics for the CSI including the two accepted computations of styles.

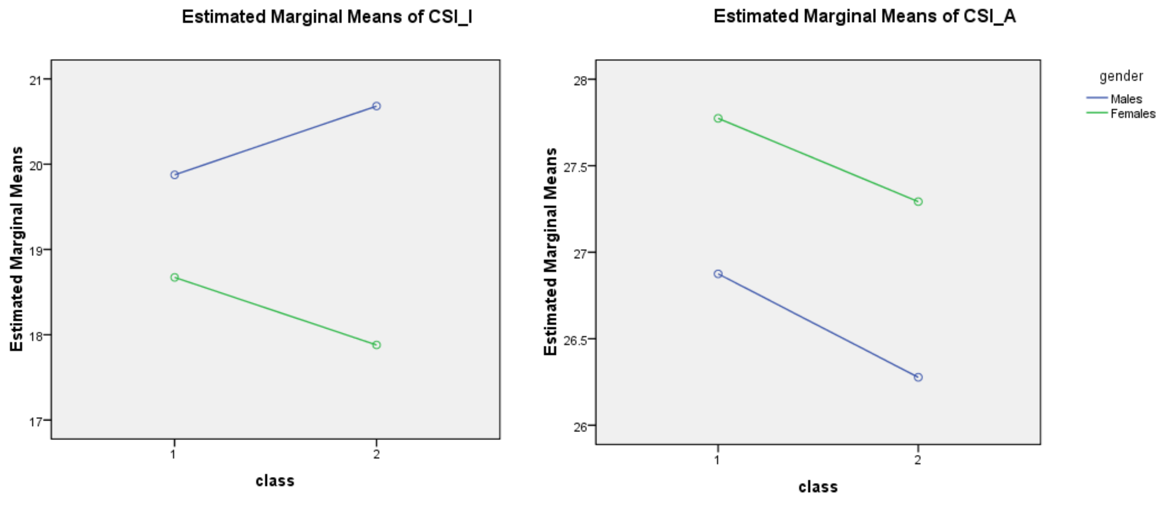


Figure 7.7. Trends in the differences between the Analytic scale (CSI_A) and the intuition scale (CSI_I) computed according to the alternative method (2 scales) for males and females in the Y1 and Y2 classes.

Group differences

Looking at table 7.14, which reports the descriptive statistics for the scales, we can also see that there are very little difference in the scores between years.

In fact, conducting a mixed ANOVA on the three styles measures (in turn the scale as a whole and in the two subscales), we ascertained that there were no significant differences between the class level (i.e. first or second year), cohort (2007 and 2008), country of origin, degree or gender. The only exception was a significant difference in the intuition scale for both gender and cohort. The trend shows that males in Y2 are more intuitive than the females ($F_{(1,349)}$=3.7, p=.055) and that they are generally slightly less analytical than females (non statistically significant). It is difficult to make sense of these differences because of the 4:1 ratio between females to males in the courses considered, but it is worthy of further investigation.

Mining approaches to learning

As for the previous examples, we conducted a Two-steps cluster analysis using the two scales, gender, degree, country of origin, class and cohort. The first solution resulted in 6 clusters, however, when looking at the centroids we realised that the metrics were largely overlapping: for this reason the solution was rejected.

Informed by the previous analysis, we eliminated the variables which did not produce significant differences (keeping only sex and country) and repeated the procedure.

Figure 7.8 shows two possible solutions: one with four clusters and one with three clusters (using only sex and the scores) which have a reasonable number of students in each cluster. Contrary to the previous two examples which required a multidimensional representation with the radial graphs, the two scores of the CSI can be easily represented in a familiar Cartesian plane. The figure shows the centroids and the size of the bubble represents the frequencies for each cluster. Whilst clusters 1 and 2 are well defined (1 has high scores in analysis and low in intuition and 2 has fairly similar scores in both), it is interesting to know that these two account for over 60% of the samples, and membership is exclusive of females for these two clusters. This explains why group 3 and 4 (the latter contains only males) are therefore more powerful in discriminating individuals compared to a simple 2-clusters solution. As the clusters 3 and 4 are largely overlapping we removed the variable country from the model as in the ANOVA it did not produce significant differences between groups.

This example clearly demonstrates in practice how the exploratory phase can inform the choice of instances used by the clustering algorithm. An interesting effect is that cluster 2 absorbed many cases which were included in cluster 1 and that the centroids of the clusters also shifted slightly with the effect that the standard deviation also decreased. It is difficult to assess which solution is *right*. Because we identified some variation in performance based on country of origin before, we will test both options when comparing the groups' AP and usage.

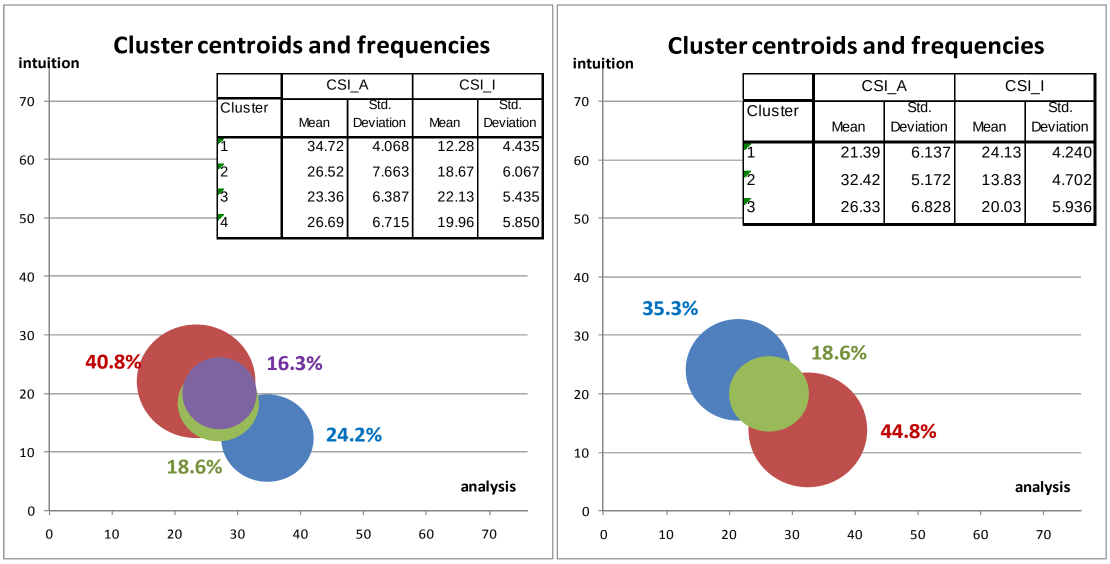


Figure 7.8. Classification of the CSI representing centroids of the clusters with frequencies. Whilst a 6-clusters solution has been considered, a 3-clusters solution provides a more parsimonious solution keeping into account the ANOVA results. The side-by-side comparison demonstrates how an additional instance (i.e. country in this case) can unnecessarily modify the resulting clustering.

Summary for the findings with the CSI

Results from the CSI produced the expected patterns for what concerns both the reliability of the tool and the scores. The analysis of reliability showed that the alternative method for computing the scores as two different scales brings further support to Hodgkinson et al. and the scores computed as two separate dimensions also improved the reliability of the measures. However the average negative correlation between the 2 scales doesn't make the debate about the complexity of the analysis-intuition any simpler.

We showed how the removal of a categorical variable from the clustering algorithm could modify the solution. Because only two dimensions are present with the CSI it seems clear that the 4-clusters solution misclassify some cases. Whilst there was no doubt that the class-year and degree types caused unnecessary noise with largely overlapping classes, the decision in favour of a 3 or 4 clusters solution, is demonstrated in figure 7.8.

7.1.4. Visual/Imagery and extended Wholistic/Analytic Cognitive Style (VICS-WA)

As indicated in chapter 5, the VICS-WA was only used in one iteration of the project with both first and second year students. The interesting aspect of this tool was that unlike the other measures of styles considered, this test is closer to cognitive tasks, and reaction times in responding to a simple choice between stimuli are at the core of the test. More details about the value of the test and the improved reliability of the test compared to the CSA (Riding 1991) can be found in Peterson et al. (2003, 2005, 2006). Here we adopted a similar strategy to evaluate the reliability of the measure at both structural and item levels.

Overall performance

The first step in the analysis for this task was to evaluate if participants' results were acceptable by performing an error analysis. Following a procedure similar to the one performed by Peterson (2003), we looked at the number of errors and the correlations between sections of the test. We also tested for significant differences in the average reaction times based on the type of stimuli and presentations. As in the previous sections, the year-classes were treated as separate samples.

Verbal-Imagery test

In the VICS test, whilst the overall error rates are quite low, it is evident that with verbal stimuli the 'mixed' category present the highest error rate and there is a difference between the first and second year students (table 7.17 and 7.18). This could be partially explained by the different commitment and motivation of the students which affected participation and attention. (I.e. first year students participated to earn course credits; second year students performed the tasks as part of a class exercise)

Task and stimuli	errors Y1	error Y2	Task and stimuli	RTs Y1	RTs Y2
Verbal	**2.39%**	**3.76%**	**Verbal**	**2.32**	**1.94**
Pictures	2.19%	3.58%	Pictures	2.10	1.77
Words	2.59%	3.95%	Words	2.53	2.12
Imagery	**2.77%**	**4.40%**	**Imagery**	**2.38**	**2.12**
Pictures	2.95%	4.32%	Pictures	2.10	1.87
Words	2.59%	4.48%	Words	2.66	2.37
Grand Total	**2.58%**	**4.08%**	**Grand Total**	**2.35**	**2.03**

Table 7.17. Percentage of errors and average RTs (in seconds) in the Verbal tasks

Task and stimuli	errors Y1	error Y2	Task and stimuli	errors Y1	error Y2
Verbal	**2.39%**	**3.76%**	**Imagery**	**2.77%**	**4.40%**
Pictures	2.19%	3.58%	Pictures	2.95%	4.32%
Natural	0.64%	1.25%	Smaller	1.92%	3.30%
Man-made	2.40%	3.69%	Equal	2.43%	3.33%
Mixed	7.99%	13.19%	Bigger	4.09%	5.58%
Words	2.59%	3.95%	Words	2.59%	4.48%
Natural	2.40%	2.44%	Smaller	1.68%	3.46%
Man-made	2.40%	3.08%	Equal	4.17%	7.64%
Mixed	4.17%	14.31%	Bigger	3.13%	4.78%

Table 7.18. Breakdown of the errors based on stimulus type in the Verbal tasks.

Year 1			**dfs**	**F**	**Sig.**
	TaskType	RTs	1,5568	5.141	**.023**
		Accuracy	1,5568	1.577	ns
	Stimulus Format	RTs	1,5568	306.201	**.000**
		Accuracy	1,5568	0.004	ns
	TaskType * StimulusFormat	RTs	1,1	4.701	**.030**
		Accuracy	1,1	1.577	ns
Year 2					
	TaskType	RTs	1,13919	119	**.001**
		Accuracy	1,13919	7.264	**.007**
	Stimulus Format	RTs	1,13919	721.147	**.001**
		Accuracy	1,13919	1.256	ns
	TaskType * StimulusFormat	RTs	1,1	23.127	**.001**
		Accuracy	1,1	0.206	ns

Table 7.19. Summary of the significant differences in RTs and accuracy for the different types of stimuli in the Verbal tasks.

To evaluate the differences between the conditions we performed a multivariate ANOVA (task type: Verbal, Imagery; stimulus type: Words, pictures) on both the number of errors and the average reaction times. In both class-years there is a significant difference in the RTs for the task type and the stimulus type (summarised in table 7.19). Contrary to what was expected, the increase in the number of errors identified above was not significant, except for tasks in Y2.

Wholistic-Analytic test

Error rates in the second part of the VICS-WA test are also similar to the ones reported by Peterson. Interestingly table 7.20 shows that there is a decrease in performance (more errors) between the first and second presentation of the tasks, but there is a decrease in reaction times, which probably suggests that fatigue elicited a quicker and more inaccurate response (the full test takes about 35-40 minutes).

		Year 1			Year 2	
Task and stimuli	errors 1st	errors 2nd	**errors tot**	errors 1st	errors 2nd	**errors tot**
Wholistic	**1.05%**	**1.80%**	**2.85%**	**1.48%**	**2.65%**	**4.13%**
Not the same	1.33%	2.50%	3.83%	1.63%	3.33%	4.96%
The same	0.78%	1.09%	1.88%	1.33%	1.96%	3.29%
Analytic	**2.37%**	**1.78%**	**4.14%**	**2.63%**	**2.13%**	**4.75%**
Not contained	3.05%	1.84%	4.89%	3.29%	2.13%	5.42%
Contained	1.68%	1.72%	3.40%	1.96%	2.13%	4.08%
Grand Total	**1.71%**	**1.79%**	**3.50%**	**2.05%**	**2.39%**	**4.44%**
		Year 1			**Year 2**	
Task and stimuli	RTs 1st	RTs 2nd	**RTs Mean**	RTs 1st	RTs 2nd	**RTs Mean**
Wholistic	**2.14**	**1.78**	**1.96**	**1.78**	**1.58**	**1.68**
Not the same	2.12	1.69	1.90	1.80	1.54	1.67
The same	2.16	1.87	2.01	1.76	1.61	1.68
Analytic	**1.79**	**1.46**	**1.62**	**1.61**	**1.36**	**1.49**
Not contained	1.95	1.55	1.75	1.72	1.42	1.57
Contained	1.62	1.36	1.49	1.51	1.30	1.40
Grand Total	**1.96**	**1.62**	**1.79**	**1.69**	**1.47**	**1.58**

Table 7.20. Breakdown of the errors and RTs based on stimulus type and presentation in the Visual tasks

A mixed ANOVA similar to the one conducted above for the visual task was performed. In this second task differences between the presentation, tasks type and type of stimulus were significant for either the speed of response and accuracy or both. As we can see from the summary table 7.21, however the patterns are not exactly the same for the two year classes. What is apparent is that although one could have expected an improvement in performance caused by a learning effect on the tasks, it looks like performance deteriorated significantly for this test over the two presentations. The tasks were presented in the same order for both cohorts and the fact that in both groups the deterioration is present can only be justified by the fatigue and loss of interest in the participants.

A careful review of Peterson's data showed very similar patterns. In her thesis she presents also the results of intercorrelations between sections of the test with Pearson r above .9. Table 7.22 presents similar calculations, but the correlation values, although significant, are somewhat less (range .57 to .9).

The results from the analysis of validity and reliability of the VICS-WA are not very different from the original study conducted by Peterson. Error rates are in line with her results even if there are some significant differences in speed and accuracy between conditions. The final aim of the task is to produce ratios of the median verbal and median imagery and a second using the median from the wholistic and analytic tasks.

Mixed ANOVA			**dfs**	**F**	**Sig.**
	TaskType	RTs	1,5119	12.715	**.001**
		Accuracy	1,5119	132.169	**.001**
	Presentation	RTs	1,5119	0.193	ns
		Accuracy	1,5119	139.521	**.001**
	Category of Stimulus	RTs	1,5119	22.51	**.001**
		Accuracy	1,5119	6.65	**.010**
Year 1	TaskType * Presentation	RTs	1,5119	13.383	**.001**
		Accuracy	1,5119	0.248	ns
	TaskType * Category of Stimulus	RTs	1,5119	0.418	ns
		Accuracy	1,5119	39.89	**.001**
	Presentation * CategoryofStimulus	RTs	1,5119	0.29	ns
		Accuracy	1,5119	5.905	**.015**
	TaskType * Presentation * Category of Stimulus	RTs	1,5119	8.474	**.004**
		Accuracy	1,5119	0	ns
	TaskType	RTs	1,4799	2.219	ns
		Accuracy	1,4799	102.945	**.001**
	Presentation	RTs	1,4799	2.524	ns
		Accuracy	1,4799	143.667	**.001**
	Category of Stimulus	RTs	1,4799	12.779	**.001**
		Accuracy	1,4799	17.297	**.001**
Year 2	TaskType * Presentation	RTs	1,4799	15.777	**.001**
		Accuracy	1,4799	1.602	ns
	TaskType * Category of Stimulus	RTs	1,4799	0.158	ns
		Accuracy	1,4799	22.57	**.001**
	Presentation * CategoryofStimulus	RTs	1,4799	0.089	ns
		Accuracy	1,4799	6.677	**.010**
	TaskType * Presentation * Category of Stimulus	RTs	1,4799	8.293	**.004**
		Accuracy	1,4799	0.153	ns

Table 7.21. Summary of the significant differences between conditions

Table 7.23 shows the descriptive statistics for the core measures derived by the aggregation of data in the various conditions. As in Peterson's original study we used the medians rather than the mean values to compute the two ratios.

Y1, N=128	Medians corr.	sig.	Y2, N=120	Medians corr.	sig.
Verbal vs Imagery	0.69	(.001)	Verbal vs Imagery	0.57	(.001)
(picture vs words)	0.9	(.001)	(picture vs words)	0.66	(.001)
Wholistic vs Analytic	0.78	(.001)	Wholistic vs Analytic	0.8	(.001)
V/I vs W/A	**0.14**	**ns**		**0.01**	**ns**

Table 7.22. Summary of the correlations between cognitive dimensions.

VICS		**n**	**V (avg. med RT)**	**SD**	**I (avg. med RT)**	**SD**	**V/I ratio**	**SD**
Year 1		128	1.80	0.46	1.86	0.39	0.97	0.17
Year 2		120	1.67	0.32	1.82	0.34	0.93	0.16
	total	248	1.73	0.40	1.84	0.37	0.95	0.16
CSA			**W (avg. med RT)**	**SD**	**A (avg. med RT)**	**SD**	**W/A ratio**	**SD**
Year 1		128	1.58	0.50	1.39	0.30	1.13	0.21
Year 2		120	1.42	0.32	1.32	0.26	1.08	0.14
	total	248	1.50	0.43	1.35	0.28	1.11	0.18

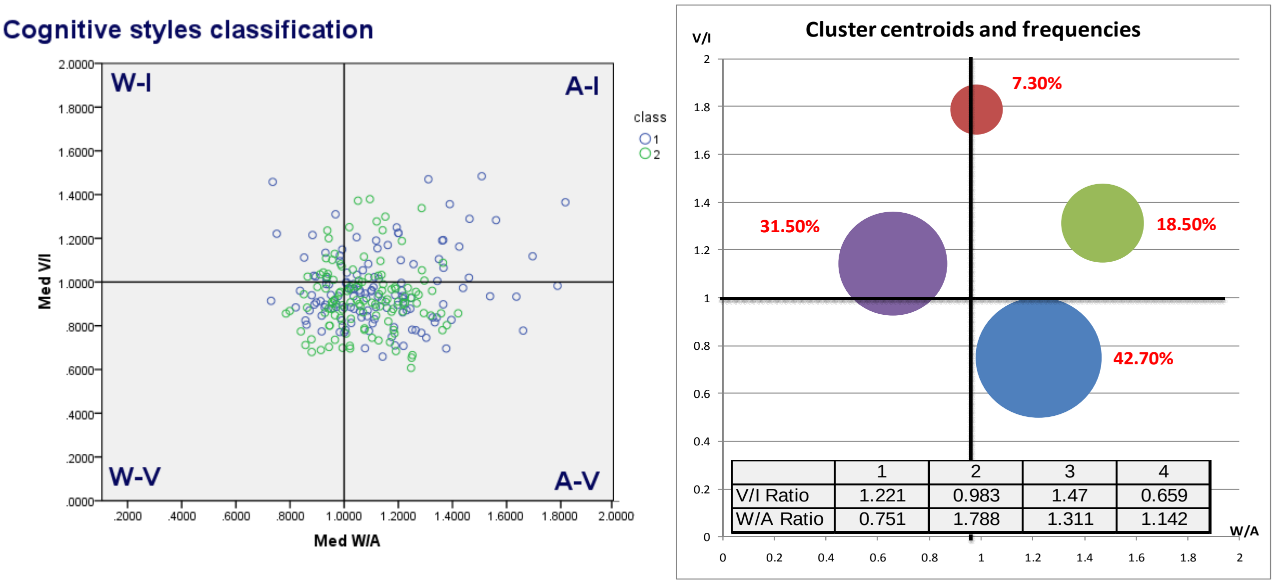
Table 7.23. Descriptive statistics for the mean values and standard deviations for the core metrics in the two parts of the test.

As identified by both Riding et al. and Peterson et al., the ratios are used to differentiate between a measure of intelligence (and/or ability) and a measure of style, which is supposed to provide a preference. In his original work Riding indicated that to determine the style of a participant one could either divide the range of values for each ratio in three categories (i.e. Analytic, middle and wholistic) or validate the scores on a sample of 999 people in the UK. Both of these are quite artificial and in fact, by plotting the ratios in a scattergram (7.9 left) we can see that most people are clustered in the middle. By looking at the ranges proposed in Peterson et al. there is little difference. One could argue that the compression of the distribution is caused precisely by the procedure applied in the selection of the stimuli which, like for a cognitive task or intelligence test, and unlike a questionnaire, was attempting to minimise the variance to improve the reliability of the measure.

Mining VI and WA dimensions.

Figure 7.9 is useful to identify two core features of the dimensions in the aggregated samples used. On one hand the distribution of ratios is very much close to the centre, making it difficult to provide a metric able to clearly discriminate between styles. On the other hand such narrow distributions make factor analytic techniques less useful to determine group fit without heavy transformations of the variables (i.e. a rotation could provide a better fit at the cost of interpretability).

After conducting a two-steps cluster analysis using the two ratios, gender and the year classes to generate a possible solution, a 3-clusters solution was rejected because the clusters seem to be too close and overlapping.. A 4-clusters option was forced in the solution using k-nearest neighbourhood technique (default value of k=3), which seems to fit better the data providing groups which, as it is shown in figure 7.9 (right), don't match precisely with the ideal quadrants of the figure on the left, but provide enough scope for differentiation.



	1	2	3	4
V/I Ratio	1.221	0.983	1.47	0.659
W/A Ratio	0.751	1.788	1.311	1.142

Figure 7.9. Classification of cognitive styles using the V/I and W/A ratios.
On the right the centroids of the clusters with frequencies (mean distance from centroids is ranging between .1 and .21)

Summary for the findings with the VICS-WA

Although the tasks are fairly simple and the results confirm previous research, there are two reasons explaining why this measure of style was only used once. On one hand the difficulty of administration (the completion of the tasks requires about 35-40 minutes per participant), on the other the fact that the results of the test provided a very narrow distribution, which, although it is in line with previous finding makes one wonder if the test is a good enough tool to discriminate between cognitive styles. Nevertheless, in the next section, we will explore the relations of this instrument with other measures of styles and we managed to generate a number of clusters which might be better in assigning participants to styles than a purely statistical solution.

7.1.5. Interrelations between measures

One of the most interesting aspects of the research conducted was the fact that multiple measures of styles were used concurrently with the students, opening up the possibility to relate the different constructs and measures. In chapter 4 we identified this as a need to bring more clarity to the field of learning styles, and therefore this is an excellent opportunity to provide useful empirical evidence to the debate.

ASSIST and TSI

The same problem which affected the evaluation of the TSI as a useful tool to measure styles, affected the analysis of the relations between the TSI and the ASSIST. Potentially this could have provided very valuable insights, particularly because of the threefold model. However, when comparing the various subscales of the ASSIST with the TSI the year 1 sample produced a number of moderate correlations (12 for the Judicial scale, 10 for the Conservative and 9 each for the Legislative and Anarchic scales as examples), however the year 2 sample did not produce similar patterns.

To minimise the effects caused by too much variation between the subscales, we therefore decided to look at the correlation coefficients of the TSI scale with the three approaches to learning plus the preferences for the learning environment. Table 7.24 presents the summary of the significant findings. As indicated with the highlights in yellow, there are some parallels between the trends in the Y1 and Y2 samples which are of particular interest. For example, the fact that the three approaches have stronger associations with the three different functions is predictable; however it is more difficult to make sense of the relations between the approaches and the forms of government.

Year 1 Samples, N = 340	Function			Level		Leaning		Form					Scope	
	Legislative	Executive	Judicial	Global	Local	Progressive	Conservative	Hierarchical	Monarchinc	Oligarchic	Anarchic	Democratic	Internal	External
DEEP Approach	.229** .000		.380** .000			.276** .000	-.152** .005	.134* .013			.340** .000			
STRATEGIC Approach	.109* .045	.193** .000	.146** .007		.207** .000		.120* .027	.279** .000		.119* .028				.173** .001
SURFACE/APATHETIC Approach	-.169** .002	.111* .041	-.184** .001			-.158** .003	.178** .001				-.123* .024			
Transmitting information (SURFACE)	-.126* .021	.151** .005	-.152** .005		.170** .002	-.186** .001	.224** .000			.194** .000	-.272** .000		.112* .040	.123* .023
Supporting understanding(DEEP)	.202** .000		.328** .000			.349** .000	-.172** .002	.134* .014		-.157** .004	.393** .000			
Year 2 Samples, N = 176	LEGL	EXEC	JUDI	GLOB	LOCL	PROG	CONS	HIER	MONA	OLIG	ANAR	DEMO	INTR	EXTR
DEEP Approach		.250** .001	.323** .000	.165* .029		.329** .000	.184* .015	.227** .002		.166* .028	.368** .000	.181* .016		.175* .020
STRATEGIC Approach		.202** .007			.311** .000	.152* .044				.338** .000			.287** .000	.240** .001
SURFACE/APATHETIC Approach	-.154* .042					-.269** .000								
Transmitting information (SURFACE)													.183* .016	
Supporting understanding(DEEP)			.275** .000			.313** .000					.350** .000			

Table 7.24. Correlations between the subscales of the ASSIST and the TSI.

Only significant correlations are reported with ** at the .01 level and * at the .05 level. Highlighted in yellow are the common features, i.e. the correlations consistent in both the year 1 and year 2 consistent in both year and

ASSIST and CSI

To our knowledge the ASSIST and the CSI have never been used with the same samples, giving a unique opportunity to evaluate the relations of the two instruments.

Because the ASSIST and CSI were used with two different samples (in the first and second year level) correlations between the score were computed separately. The resulting patterns, however, are quite similar. In particular, no significant correlations were found between the scales of the CSI (no matter if these are computed with the traditional or alternate method). However a number of consistent correlations were uncovered specifically related to the Strategic approach and its subscales. Table 7.25 shows the magnitude and direction of the relations.

Although the size of the relations is quite small, it is interesting to observe some expected relations between what could be defined an analytic/rational executive and a negative relation with particular activities and the intuitive dimension. The fact that relations are stronger with the two separate scales brings further evidence to the fact that a single dimension with intuition and analysis at the two extremes might not be suitable, weakening the associations with other tools.

	Year 1, N=261			Year 2, N=122		
	CSI_all	CSI_A	CSI_I	CSI_all	CSI_A	CSI_I
STRATEGIC Approach	.260**	.367**	-.223**	.380**	.461**	-.269**
	.000	.000	.000	.000	.000	.003
relating ideas [DEEP]			.131*			.207*
			.035			.022
organised studying	.195**	.324**	-.215**	.308**	.396**	-.197*
	.002	.000	.000	.001	.000	.029
time management	.227**	.308**	-.197**	.191*	.300**	-.237**
	.000	.000	.001	.036	.001	.008
Alertness to assessment demands	.160**	.196**		.401**	.333**	
	.010	.001		.000	.000	
achieving	.201**	.286**	-.177**	.247**	.334**	-.204*
	.001	.000	.004	.006	.000	.024
monitoring effectivness	.141*	.186**		.290**	.359**	-.206*
	.023	.003		.001	.000	.023
fear of failure [SURFACE]		.143*			.258**	-.227*
		.021			.004	.012

Table 7.25. Summary of the significant correlations between the ASSIST and CSI.

The focus is particularly on the Strategic approach (green) which provides consistent significant relations.

ASSIST and VICS-WA

From the analysis conducted, we found virtually no correlations, only the subscale ‘Relating ideas’ negatively correlated with the VI ratio ($r_{(159)}$=-.16 p<.5) and the ‘Use of evidence’ correlated with the WA ratio ($r_{(159)}$=.18, p<.05). Both values associations are only found in the Y1 sample and the size of the correlations is small. The highest recorded correlation is the ‘Unrelated memorising’ correlating .22 (p<.05) with the Median analysis RTs, but it could be completely incidental.

TSI and CSI

The TSI and CSI were used together on a single sample of Y1 students. As indicated in table 7.26, patterns are predictable and the direction of the association seems to provide further evidence in support of the CSI (the simpler to interpret). The correlations are small, but consistent with the theory. So, for example, the fact that the ‘executive’ function correlates positively with analysis and negatively with intuition is something that was seen in managers. Further research in different domains might provide further useful interpretations for these relations.

		CSI_All		CSI_Analysis		CSI_Intuition	
dimension	**scale**	**r sample1**	**p**	**r sample 2**	**p**	**r sample 3**	**p**
Function	Legislative						
	Executive	.135*	0.028	.267**	0.000	-.236**	0.000
	Judicial						
Level	Global	-.147*	0.017	-.254**	0.000	.224**	0.000
	Local	.191**	0.002	.299**	0.000	-.192**	0.002
Leaning	Progressive			-.154*	0.012	.298**	0.000
	Conservative	.171**	0.005	.258**	0.000	-.230**	0.000
Form	Hierarchical	.173**	0.005	.256**	0.000	-.146*	0.017
	Monarchinc					.149*	0.015
	Oligarchic	.175**	0.004	.168**	0.006		
	Anarchic			-.185**	0.002	.378**	0.000
	Democratic						
Scope	Internal						
	External	.140*	0.022			.128*	0.038

Table 7.26. Summary of significant correlations between the TSI and CSI.

TSI and VICS-WA

The VICS-WA ratios produced a negligible association with the MSG. In fact the only significant correlations was Anarchic scale with the WA ratio ($r_{(232)}$=.16 p=.015). Yet again it seems that the RT base of this task didn't produce enough variation to be comparable with the inventories.

VICS-WA and CSI

No correlations were found between the VICS-WA and the CSI. This is the most surprising result as both tools are supposed to measure the analytic construct.
In fact, this result is more useful than many other correlations explored thus far. Given the fairly high validity and reliability of the two individual instruments, one cannot fail to question whether the label of *analytic* style is misused in one of the two.

The fact that the VICS-WA seemed to have so little in common with the other metrics is also providing evidence that the tasks might tap more into the domain of abilities rather than styles.

7.1.6. An overview of styles classification

After successfully determining clusters for each of the measures of styles, a number of practical issues had to be taken into account to evaluate the possible avenues for further analysis.
The aim of the clustering was to reduce the complexity of the dataset with a rich enough description of each group. One option would be to try to reduce *types* even further.
To explore the possibilities we attempted a further clustering procedure on the classifications assignment for each student. This however produced a 2-clusters solution. The reason justifying this result is the list-wise exclusion of cases which reduced the usable sample to 30 students, therefore this option was rejected.

Although the missing values turned out to be a problem for clustering, a second option was explored to look at the variation of *types* in the sample. After aggregating the cluster assignments for each person in 4-digits strings in which missing values were substituted with zeros, we found that there were 260 different types of combinations and that many profiles were represented only a few times. Again, this was not an optimal solution, because types need to be representative.

For the above reasons we opted to consider the instruments separately rather than attempting to create a *super-profile* (or meta-profile) which, instead of taking advantage of the richness provided by the different measures tried to over simply the characterisation.
In further support of this method is the fact that it would allow us to test whether clustering per se, is useful in this context thanks to the availability of four distinct measures of styles.

7.2. Styles and academic performance

7.2.1. Interrelations between styles and AP

The most striking relations found are recorded by the ASSIST (tables 7.27 and 7.28) and some of the figures are fairly controversial. In the first year, the strategic approach correlates positively with most of the forms of assessment and even if all three approaches seem to be correlated with the final Y1 grade, this could be feasible as we have highlighted the existence of 6 different clusters of students with significantly different features. Recalling Laurillard, there might be equally valid paths to learning, although not all might be the most effective or efficient. Notable is also the difference in the relation with the performance in the MCQs section of the second exam, compared with the semester 1 exam, with which the surface approach is strongly related. It is not surprising that the subscales highlights possible reasons why this might be the case (unrelated memorising and fear of failure). In Vigentini (2008) we hinted at the fact that the additional resources (i.e. the test banks) were used by students to 'learn the answers' rather than to gain a deeper understanding of the concepts. This could be evidence in support of the statement. Fear of failure was also very much present as we showed that on average students performed much lower in the MCQs section than in the Essay section in the first exam.

In the year 2 results are particularly interesting. This is in fact a core course and performance becomes the criterion to continue into honours. Unlike the year 1 course, however, a heavy practical component with statistics and project reports characterises the activities. The strong correlation between the subscale 'lack of purpose' and the statistics exam is demonstrating a fact which was explained in chapter 1: psychology students don't like this element of the course and do not understand why they need this element at all. The strategic approach with an extremely high correlation with the Y2 grade (>.9), demonstrates that the ability of adapting to the course demands is an essential element for success.

Puzzling are the relations between the overall psychology average grade (across the 4 years) and the various approaches, as this seems to indicate that either a deep or strategic approaches are less correlated with achievement. For us, the very big shift in the value of the surface approach between year 1 and year 2 is a clear indicator that the very big increase in demands of the second year course, causes a fundamental change in the way students approach their workload and priorities which might ultimately produce a differentiation in grades between those who are more capable than others. It would have been extremely interesting to assess how approaches to learning change in the honours course in year 3 and 4.

ASSIST and Academic indicators in Year 1, N = 444	Coursework Average	Exam 1 (essay)	Exam 1 (MCQs)	Exam 2 (essay)	Exam 2 (MCQs)	Final Grade	Psychology Average	Research partecip.	Attendance	Late Essays
					.105*	**.204****	**-.142***			
DEEP Approach					**.027**	**.001**	**.001**			
Seeking meaning					.328**	.126**	.902**			
					.044	.008	.007			
Relating ideas						.142*	-.151**			
						.029	.032			
Use of evidence	-.127*					-.182**	.718**			
	.008					.005	.004			
Interest in ideas						-.157**	.144**		.097*	
						.003	.002		.041	
	.173**	**.328****		**.571****	**.305****	**.673****	**.482****	**.198****	**.182****	**-.177****
STRATEGIC Approach	**.000**	**.043**		**.012**	**.007**	**.000**	**.000**	**.000**	**.000**	**.000**
Organised studying	.256**					-.141**	.139**	.173**	-.228**	-.143**
	.009					.001	.001	.001	.000	.002
Time management	.138**			.101*		.202**	.442**	.177**	.768**	.297**
	.004			.033		.020	.014	.001	.038	.000
Alertness to assessment demands					.768**					.460**
					.036					.010
Achieving	-.142*	.328**		.524**	-.127*	.150*	.615**	.367**	-.117**	-.150**
	.001	.042		.002	.007	.000	.000	.002	.000	.021
Monitoring effectivness	.126**				.410**	-.141**	-.240*	-.164**	.199**	
	.008				.010	.001	.001	.001	.004	
		.699*	**-.144****		**.614****	**.317****	**-.136***			
SURFACE/APATHETIC Approach		**.006**	**.002**		**.012**	**.001**	**.001**			
Lack of purpose										
Unrelated memorising		.515**	.319**		.664**	-.239*	-.147*			
		.004	.000		.003	.000	.000			
Syllabus boundness			-.680**			.479**	.225**			
			.016			.002	.002			
Fear of failure		.200*	-.271**		.560**			.141**		
		.029	.007		.010			.006		
Preferences for learning environments										
Transmitting information (SURFACE)										-.105*
										.028
Supporting understanding(DEEP)	.677**		-.278**		-.278**	.587**	-.235*			
	.031		.023		.024	.000	.000			

Table 7.27. Summary of the significant correlations between the dimensions of the ASSIST and Academic performance in Y1.

ASSIST and Academic indicators in Year 2, N = 288	Coursework average	Exam 1 (essays)	Exam 2 (essays)	Statistics Average	Final Grade	Psychology Average	Attendance	Late reports	Class quiz score
DEEP Approach						**-.160****	**-.134***		
						.012	**.040**		
Seeking meaning	.102*								
	.024								
Relating ideas									
Use of evidence						-.202*			
						.041			
Interest in ideas						-.200*	.154*		
						.010	.023		
STRATEGIC Approach	**-.181****	**.571****	**.495****	**.442****	**.993****	**-.194****	**-.574****	**.125****	
	.000	**.043**	**.020**	**.048**	**.000**	**.000**	**.000**	**.000**	
Organised studying	.213**	.150*	.126*	.126*	.280**	-.276**	.147**	.282**	.185**
	.000	.011	.033	.033	.000	.000	.002	.000	.009
Time management	-.228**		.120*		.235**	.173**	-.134*	.556*	.326**
	.002		.042		.000	.003	.040	.000	.041
Alertness to assessment demands	.377**		.065		.182**	.173**	.579**	-.148**	
	.000		.275		.002	.003	.003	.028	
Achieving	-.122**	.162**	-.134*	-.365**	-.117*	-.208**	.706**	.311**	
	.000	.017	.018	.030	.000	.000	.000	.005	
Monitoring effectivness	-.228**				.673**	.179**	.142*		
	.002				.001	.002	.038		
SURFACE/APATHETIC Approach	**.121***				**.311****	**.765****			
	.012				**.006**	**.000**			
Lack of purpose	-.153**			.655**	.187*	-.244*	-.224**		
	.000			.018	.001	.000	.001		
Unrelated memorising					.288**	-.201*			
					.040	.003			
Syllabus boundness									
Fear of failure						-.283**		-.136*	
						.020		.021	
Preferences for learning environments									
Transmitting information (SURFACE)								-.141*	
								.017	
Supporting understanding(DEEP)				.371**				-.235*	.133**
				.015				.004	.004

Table 7.28. Summary of the significant correlations between the dimensions of the ASSIST and Academic performance in Year 2.

The correlations between the CSI and academic indicators seem to make some sense (Table 7.29). The fact that MCQs form of assessment correlates negatively with the intuition scale is not surprising. The fact that there is a relation between coursework (positive) and late submissions (negative) with the analysis scale is also not surprising as an analytical student would be more sensitive to course demands and systematically organise their time to meet the allocated deadlines.

	Exam 1 (MCQs)	Exam 1 (MCQs)	Y2 coursework	Psychology Average	Research partecip.	Late Essays	Year 2 Attendance
CSI Analysis			.235**		.196*	-.113*	
			0.009		0.001	0.039	
			121		266	331	
CSI Intuition	-.172**	-.192**		-.200**	.169*		
	0.002	0.000		0.000	0.006		0.005
	332	332		332	266		127

Table 7.29. Correlations between the CSI and academic indicators (only significant values are reported)

The TSI has provided results difficult to explain: its relations with academic indicators were not any different. In fact, by looking at Sternberg’s summary table of the relations between the scales of the TSI and instruction methods (table 4.6), the correlations emerging from table 7.30 are not particularly easy to explain or justify.

The hierarchical and judicial scales, for example, were associated with lectures and thought based questions: Although it could be argued that essay questions and project reports stimulate students to think, the correlations reported are quite small (if any).

The hierarchical scale, correlating with many of the items of assessment, particularly in year 1, might show that students prefer the lecture as a mode of teaching the subject, which then impacted on the exams.

7.2.2. Correlation of styles and expressed intentions

Proxy measures such as attendance and punctuality are the expression of particular a behavioural patterns. The literature has looked at parameters such as seminar attendance, predicted absenteeism and truancy as precursors of disruptive behaviours and low academic performance (Petrides et al. 2005).

On the rightmost part of the tables Table 7.27 and 7.28 and 7.29 we also indicated how attendance, research participation (only in Y1) and number of late submissions (percentage of the total) related to styles measures and AP (table 7.30).
Of all the styles measures, the ASSIST was the most useful in identifying simple and meaningful relations. For example the correlations between the strategic approach and attendance were small but significant.

For what concerns the relations between AP and behaviours table 7.31 and 7.32 provide an overview of the significant correlations between attendance, participation and punctuality in submitting coursework. Not surprisingly, tutorial attendance in year 1 correlates with essay average grade ($r_{(1336)}$=.34, p<.01) and Y1 final grade ($r_{(1336)}$=.25, p<.01) and research participation ($r_{(542)}$=.38, p<.01).

What is most interesting of these summary tables is the strength of the relations with later assessments points. For example, late essays have a growing impact (N=1339) with Pearson rs=-.15, -.18, -.29 on performance on the final grade (N=1280) r=-.19. The correlation is moderate with the reports in Y2, showing that patterns of behaviours are often carried on in successive years.
The strength between attendance and the statistic grade in Y2 is expected, but the impact of the research participation across all years was not. In fact, although this element of the year 1 course only accounts for 5% of the mark in year 1, it might show that students actually benefit from it. A number of explanations are possible, but one could be that those who fulfil the requirements of participation might be more committed than others and this is ultimately reflected in their overall performance.

TSI and Academic indicators in Year 1, N = 456		Coursework Average Y1	Exam 1 (essay)	Exam 1 (MCQs)	Exam 2 (essay)	Exam 2 (MCQs)	Y1 Final Grade	Research partecip.	Attendance	Late Essays
Function	**Executive**	.132** 0.005			.137** 0.003	.191** 0.000	.115* 0.015			-.095* 0.042
	Judicial				.107* 0.023	.104* 0.027	.129** 0.006			
Level	**Local**									-.106* 0.024
Leaning	**Progressive**									
Form	**Hierarchical**	.147** 0.002	.099* 0.034	0.087 0.064	.170** 0.000	.203** 0.000	.147* 0.000	.108* 0.045	.100* 0.034	
	Oligarchic									
	Anarchic									
	Democratic							.101* 0.024	.121* 0.032	
Scope	**External**			-.111* 0.018		-.096* 0.041				

TSI and Academic indicators in Year 2, N = 199		Coursework Average Y2	Exam 1 (essays)	Exam 2 (essays)	Statistics Average			Psychology Average	Attendance	Late reports
Function	**Executive**							**.093*** **0.047**		
	Judicial	-.200** 0.007			-.157* 0.034			**.115*** **0.015**		
Level	**Local**								.255** 0.000	
Leaning	**Progressive**									
Form	**Hierarchical**							**.186**** **0.000**	.185** 0.009	
	Oligarchic									
	Anarchic								.158* 0.026	.236** 0.001
	Democratic	-.198** 0.007	-.251** 0.001	-.210** 0.004						-.153* .037
Scope	**External**	.166* 0.024								

Table 7.30. Correlations between the TSI and academic indicators (only significant values are reported)

		Essay 1	Essay 2	Essay 3	Coursework Average	Exam 1 (essay)	Exam 1 (MCQs)	Exam 2 (essay)	Exam 2 (MCQs)	Final Grade
Research participation	Pearson Correlation		.302**	.359**	.248**	.221**	.321**	.321**	.337**	.380**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.000
	N		561	561	553	566	566	566	558	543
Tutorials attended	Pearson Correlation	.248**	.337**	.332**	.343**	-.111**	.374**		-.070*	.250**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000		.011	.000
	N	1332	1332	1332	1331	1337	1337		1336	1280
Late essays	Pearson Correlation	-.145**	-.183**	-.298**	-.192**					-.187**
	Sig. (2-tailed)	.000	.000	.000	.000					.000
	N	1340	1340	1340	1331					1280

		Report 1	Coursework Average Y2	Late Reports	Y2 Final Grade	Y2 attend.	Average Y4	Psychology Average
Research participation	Pearson Correlation	.209*	.266**		.326**	.330**	.397**	.390**
	Sig. (2-tailed)	.038	.008		.000	.000	.000	.000
	N	99	99		226	121	543	543
Tutorials attended	Pearson Correlation	.179**	.223**		-.324**	.127*	.223**	.230**
	Sig. (2-tailed)	.000	.000		.000	.044	.000	.000
	N	524	533		691	251	1286	1286
Late essays	Pearson Correlation	-.313**		.367**	.322**	-.164**	-.197**	-.202**
	Sig. (2-tailed)	.000		.000	.000	.009	.000	.000
	N	524		540	209	251	1286	1286

Table 7.31. Correlations academic indicators in year 1 (only significant values are reported)

		Report 1	Report 2	Report 3	Report 4	Coursework Average	Exam 1 (essay)	Exam 2 (essay)	Statistics Average	Final Grade
Year 2 Practicals attended	Pearson Correlation	.179**	.169**	.182**	.179**	.223**	.482**	.404**	.411**	.326**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	524	522	523	520	533	230	230	230	226
Late Reports	Pearson Correlation	-.365**	-.231**	-.271**	-.209**	-.226**	-.182**	-.149**	-.142**	-.324**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	713	712	713	710	721	721	721	721	691
Week 5 Quiz	Pearson Correlation							.178**	.193**	.322**
	Sig. (2-tailed)							.010	.005	.000
	N							211	211	209

		Literature Review Y3	Y3 Research methods	Average Y4	Psychology Average
Year 2 Practicals attended	Pearson Correlation	.369**	.235*	.409**	.400**
	Sig. (2-tailed)	.004	.047	.000	.000
	N	58	72	257	257
Late Reports	Pearson Correlation		-.182**	-.309**	-.295**
	Sig. (2-tailed)		.001	.000	.000
	N		363	730	730
Week 5 Quiz	Pearson Correlation	.232*	.105		
	Sig. (2-tailed)	.050	.419		
	N	72	61		

Table 7.32. Correlations academic indicators in year 2 (only significant values are reported)

7.3. An overall model of performance

One of the most interesting aspects of the research conducted with the available datasets is the possibility to study academic performance in relation to other measures. As in the previous section we examined the correlations between measure of styles and indicators of academic performance; it is now useful to look at the predictive power of all these metrics in relation to performance. Looking back at chapter 4 we have seen two models in which personality metrics and styles were drawn in the prediction of academic performance.
Here we will start from the observation of the predictive power of styles on AP considering the two year-classes in turn, then changing perspective, we will use the clusters generated in the previous analysis to look at differences in AP and attempt to model the interrelations between styles and AP to provide an overall model able to heuristically determine performance from the styles.

7.3.1. Linear models of academic performance

An extension of the correlation analysis conducted in the last section is regression analysis, which allows one to look at the concurrent impact of the variables introduced in the model in the prediction of a third variable. Like other parametric tests, however, the technique is limited, especially when collinearity (i.e. high correlations between the variables) is present or there is limited prior knowledge about the importance of each variable in the literature.
To give an example, if we were to run a simple regression using all parameters considered thus far (all styles and proxy measures) to predict AP, the resulting model accounts for over 66% of the variance. This is great news for researchers of styles, considering that in the last chapter we have seen that only 10% of the variance is predicted by prior performance. However, the problem with this approach is interpretational in nature given that the mathematical prediction is a direct consequence of the number of parameters included in the equation.

At the opposite end, another approach would be to look at the predicting power of each style test in turn. For example, looking at the models generated in this way to predict the Y1 final mark, the VICS-WA accounts for 1% in the prediction of the Y1 grade, the CSI about 5%, the TSI for about 10%, and the ASSIST between 6% and 13% (depending if the three approaches or all the subscales are used).
From the methodological point of view, the former approach is not acceptable because it does not provide a structured approach to hypothesis testing.

The latter, in practice, is what researchers in the field have been doing for decades, exploring only a just a partial view of the *styles spectrum* without the means for comparing tools and instruments against each other (in most cases two or three inventories have been used concurrently).

Yet another familiar approach to data reduction in social sciences has often been factor analysis. The technique is specifically designed to pull together measures that are highly correlated. In fact, by applying a principal component factor analysis without rotation to all our styles metrics[15], a 15 factors model is generated (see table 7.33 in the next page). Using the resulting 15 factors in the regression model, explains about 50% of the variance in the Y1 final mark, 38% of the Y2 final mark and 47% of the average psychology mark at university. These seem quite reasonable results, at least from the mathematical point of view, and match the logic of previous research (i.e. Zhang and Sternberg threefold model of styles mentioned earlier performed similar procedures to check whether the addition of personality to learning styles added anything to the TSI).
The key issue, however, is to explain what the originated factors are, what they mean and what impact they have in practice for both the interpretation of the resulting merged dimensions or the teacher interested to know how styles are related to AP.

7.3.2. Using the clusters to understand students'

A completely different approach offered as a result of the previous data mining applied to the styles measures, is to look if and how natural grouping in the data produce differences in performance. Once a stylistic label is known to exist, determining whether such classification has an impact on AP could be useful in practice in order to implement appropriate teaching strategies.

For example, using the grouping generated we conducted three mixed ANOVAS on the performance in Y1 in Y2 and in successive years, using the clusters of each style measure in turn. Considering that the clusters were generated independently from AP the hypothesis was that if the clusters are discriminating enough the types of students in each group, chances are that their AP performance will also be different.

[15] Note that apart for removal of outliers, no other optimization with variables and or factors was attempted.

In fact only the groups from the VICS-WA didn't produce significant differences in performance (table 7.34).

	Component														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Y2_interestinideas	.678											.326			
Y2_achieving	.671														
Y2_relatingideas	.654	-.345													
Y2_useofevidence	.619								.312						
Y2_seekingmeaning	.618											.368		-.304	
Y1_supportingunderstandin	.610											-.301			
Y1_useofevidence	.572				.502										
ANAR	.566		.340					.375							
Y2_supportingunderstandin	.565	-.357							.318						
Y2_syllabusboundness	-.539				.340										
Y1_relatingideas	.489		-.381			.381									
Y1_interestinideas	.478		-.365	.448											
Y2_monitoringeffectivness	.478	.451			-.426										
Y1_syllabusboundness	-.471	.362	.305												.342
Y1_organisedstudying	.452						.319				.332				
Y2_organisedstudying	.434	.342	-.425			-.323	.324								
Y1_monitoringeffectivness	.415	.301		.388											
CSI_A		.682													
Y1_Alertnesstoassessmentdemands		.587	.337												
Y1_transmittinginformation	-.444	.575													
Y2_Alertnesstoassessmentdemands		.550	.315							-.420					
Y2_transmittinginformation	-.398	.541													
CSI_I		-.529	.344	.332											
Y1_fearoffailure		.514					-.497								
Y1_unrelatedmemorising		.499										-.308		-.467	
OLIG		.459													
Y1_achieving	.367	.442	-.304						-.406		.322				
Y2_fearoffailure	-.367	.439							.323						
EXTR		.438		.433										.310	
EXEC	.329	.429		-.404				.328							
LOCL		.387					.382								
MONA			.673	-.306					-.319						
PROG	.467		.623												
CONS		.438	.501	-.302		.344									
Y2_timemanageent	.376		-.464				.372								
Y1_lackofpurpose				-.627					.363				-.423		
HIER				-.623		.367									
INTR				-.468	.367	-.338	.420								
LEGL	.393			-.342	.476			-.338							
Y2_lackofpurpose	-.364				.475				.323						
Y1_seekingmeaning	.337		-.337		.453										
GLOB			.494			-.516									
VI ratio						-.346	-.405							.310	
JUDI	.376		.394					.534		.324					
DEMO		.329				.305		-.466			.367				
Y2_unrelatedmemorising	-.408				.385				.441						
WA ratio	.304				.337	-.329				-.445					
Y1_timemanageent	.408				.319						.380		-.423		

Table 7.33. Factor analysis of all measures of styles. Values below .3 were excluded and no rotation of the axes was attempted.

This is quite encouraging particularly when looking at the different types of assessment, because the natural groups generated with clustering can identify performance differences. The technique also demonstrated that, apart for the VICS-WA, *despite* the measure of styles which is elected, the resulting classification is *good enough* to discriminate students' performance, providing a valuable method for characterising the students.

ASSIST Y1		**C1**	**C2**	**C3**	**C4**	**C5**	**C6**
Y1 Coursework Average	mean	55.03	61.23	56.02	57.28	55.31	58.03
	sd	8.48	.	6.95	6.40	10.07	6.40
	N	113	1	119	66	126	9
Y1 Exam MCQs average	mean	47.47	52.85	52.60	52.85	51.96	48.24
	sd	13.01	.	9.40	9.95	10.03	8.22
	N	113	1	119	66	126	9
Y1 Exam Essay average	mean	57.94	66.50	58.99	59.64	58.81	59.83
	sd	10.52	.	9.04	7.04	8.27	12.62
	N	113	1	119	66	126	9
Y1 Final grade	mean	54.99	64.00	56.67	58.71	57.64	58.44
	sd	8.45	.	6.76	6.24	7.37	7.70
	N	113	1	119	66	126	9
Psychology Average	mean	54.95	61.73	56.67	59.45	57.74	60.35
	sd	8.44	.	6.76	6.50	7.42	7.78
	N	113	1	119	66	126	9

TSI		**C1**	**C2**	**C3**	**C4**
Y1 Coursework Average	mean	56.28	55.70	53.39	55.15
	sd	8.81	10.12	10.40	6.74
	N	157	65	80	147
Y1 Exam MCQs average	mean	50.74	55.00	48.22	48.34
	sd	10.77	11.73	11.88	10.79
	N	157	65	80	147
Y1 Exam Essay average	mean	60.07	60.14	57.27	58.47
	sd	7.79	7.40	8.33	9.13
	N	157	65	80	147
Y1 Final grade	mean	57.83	59.71	53.76	56.50
	sd	6.76	7.00	8.72	6.99
	N	157	65	80	147
Psychology Average	mean	57.94	59.50	53.79	56.79
	sd	6.86	7.11	9.05	7.07
	N	157	65	80	147

Table 7.34. Summary of descriptive statistics for the means of various indicators of AP in the different clusters for the identified earlier.

Table 7.35 provides an overview of the significant differences in the performance in various assessment points using the clustering. It is apparent that the effectiveness of the clusters is reduced in the year 2, in which only the clusters from the ASSIST are sensitive to the variation of grades.

More work needs to be done to determine if these findings are replicable, and the fact that similar results were obtained in the two class-years suggests that this is possible.

Year 1		F	Sig.
ASSIST clusters (5, 434)	Y1 Coursework Average		ns
	Y1 Exam MCQs average	3.718	.003
	Y1 Exam Essay average		ns
	Y1 Final grade	2.866	.015
	Psychology Average	3.917	.002
TSI clusters (3, 445)	Y1 Coursework Average		ns
	Y1 Exam MCQs average	6.357	.000
	Y1 Exam Essay average	2.629	.050
	Y1 Final grade	9.240	.000
	Psychology Average	8.358	.000
CSI clusters (2, 323)	Y1 Coursework Average		ns
	Y1 Exam MCQs average	2.964	.053
	Y1 Exam Essay average	3.104	.046
	Y1 Final grade	2.913	.056
	Psychology Average	3.563	.029
VICS-WA clusters (3, 238)	Y1 Coursework Average		ns
	Y1 Exam MCQs average		ns
	Y1 Exam Essay average		ns
	Y1 Final grade		ns
	Psychology Average		ns

Table 7.35. Summary of the significant differences in One-way ANOVAs for scores in performance indicators grouped by the styles clusters.

7.4. Summary of findings

This chapter provided a number of valuable insights. Firstly we demonstrated that the measures of styles used had a fairly good reliability overall and to large extent bring further support to the respective authors.

The fact that we had the opportunity of repeating the tests over a number of years and in some cases we had available a longitudinal sample, was an excellent opportunity to investigate the reliability of the instruments further.

The ASSIST, in particular, demonstrated very strong reliability and construct validity. The TSI was more difficult to explain, however we suggested that the instrument might be more subject to the context than initially expected. We considered two different ways of computing the scores for the CSI based on a simple or complex conceptualization of the Intuition-analysis model. From the results it seems that a complex model with a two-dimensional framework is more suitable. The VICS-WA was also shown to be very reliable replicating Peterson’s findings.

The second key result was in defining the relations between the instruments: this was possible because in a number of iterations of this project more than one measure of style was administered. Some associations were in the expected direction, some were not. Most disappointing was lack of correlation with the VICS-WA, which was the measure with the highest expected internal reliability and validity. The presence of bi-directional relations with the CSI computed as two dimensions with some of the other styles measures brought further support to the conceptualisation of Hodgkinson and Sadler-Smith. A set of complex relations between the ASSIST and the TSI were also drawn, but the interpretation was more difficult due to the lack of consistent replications of patterns.

The relations between styles and academic indicators were useful in considering expression of behaviours and their respective impact on performance. The results were largely in line with expectations, with the ASSIST (particularly the strategic approach) correlating moderately with participation, attendance and punctuality, but overall it was expected a stronger relations between styles and behaviours. As we will explore in the next chapter, online usage is another form of expressed behaviour, therefore this might impact on later evaluations.

The third key insight was the fact that data reduction, using data mining techniques was effective. Clustering (even the simple) technique applied with the Two-steps algorithm, produced groups which are *useful* and *relevant* to discriminate performance. The statement was true for the clusters obtained from three out of the four measures of styles used in year 1, and the clusters for the ASSIST were consistent also year 2.

This is an important result which will be carried on into the next chapter: what is expected is that the same classification will be useful and effective in differentiating between online usage.

Chapter 8. Online usage: a window on students' behaviour

"(…) no matter what tool you use, the best that all this data [referring to web usage] will help you understand is what happened. It cannot, no matter how much you torture the data, tell you why something happened"

(Kaushik, 2007)

In the last chapter we were able to identify a number of interesting correlations between cognitive/learning styles and how they relate to students' performance. As well as identifying differences based on demographic groups, we were also able to isolate patterns from styles metrics, which allowed clustering of students based on their stylistic preferences. Such clusters, emergent from the styles metrics, allowed us to find significant differences in AP. This promising result confirmed the utility of clustering techniques to reduce the complexity of the data to demonstrate key differences between groups. Similar outcomes are expected in the analysis of the behavioural expression from studying and learning using the online material.

The data from the online usage will be at the core of this chapter as a window into the *behaviours* and *intentions* of the students.

Although the use of online resources is only one, relatively small component of students' lives at university, and one of many learning tools available to them, usage patterns will be examined and related to performance outcomes in order to identify interesting variations in AP directly related to the usage of e-learning.

We will also consider the question of how the concept of styles can be used in this context. The first obvious step is to use the same styles-based clusters discovered in the last chapter to look at possible differences in the use of the material online. If these groups were good enough to highlight differences in AP it is expected that similar differences will be found in the online usage. However, as we will quickly discover in the next section, it will be necessary to make a step backward, because the concept of usage needs an adequate characterization before it can be used in the analysis. By taking a data-driven approach we will identify three dimensions of usage in *extent*, *efficiency* and *richness* of browsing behaviours. Next, emerging patterns from usage will be sought, leading to *styles* of browsing. Two examples could be efficient paths or *typical* sequences of behaviour, both of which might lead to specific learning outcomes.

Contrary to some of the literature in intelligent tutoring systems, this approach is taking a step away from the context and the domain specific content. The learning trajectories of students interacting with the system over a protracted period of time, and in a natural context, will provide a wealth of information not usually available to teachers.

It is therefore necessary to start from a more detailed overview of the data available and the steps leading to its interpretation of behaviours.

8.1. Making sense of online behaviours

8.1.1. From data to interpretation: assumptions and terms

Online activity, or usage, is intended as the log of all access to the VLE and additional e-packs. The basic level of usage is characterised by the **clickstream**. This is simply a sequence of clicks which a user performs when viewing content (i.e. a number of online pages). Every time a click is actuated, the system records a list of items in the form of a line of text. This list is dependent on the software residing on the server there and are different conventions (see appendix 1). The line of text is appended to a ‘traffic log’ file on the server. In our case this information is presented in the table 8.1. In bold we highlighted the most important items.

From the timestamps shown at the bottom of Figure 8.1, a first core assumption is that the time spent on a page (computed as the difference between the timestamp for page B minus the timestamp for page A in the example) is considered equivalent to ‘active engagement’. This means that we believe that students are interacting with the tool by reading content or engaging with the activities presented on screen. However, as in psychology experiments, in which very long reaction times are *trimmed* to avoid skewing the results (see Ratcliff, 1993), we adopted a similar approach for very long dwell times on a single web page. In web usage mining (WUM), more than 3 minutes on a single page is considered an index of disengagement (Srivastava et al. 2000, Berendt 2001, Cooley 1997). However, from the outset, and based on the system specifications, as well as the nature of the web resources which require longer times (i.e. a quiz or test) we increased this limit. The natural dwell on each page is measured in seconds (~6 seconds, with an average duration of each session of about 40 seconds –session is intended as a continuous list of clicks). It should also be noted that the system used has a timeout limit (i.e. the time of 'inactivity' after which the user is automatically logged off from the system). This interval was set by the system administrator to 30 minutes. Instances of such occurrences are usually at the end of the session, clearly identifying examples of disengagement. However when a student takes a test, although the sequence lapses for longer than 30 minutes, clicks on the answers allow the system to know that the user is still active, leading to extremely long dwell times. For the evaluation of the time spent we will use the real times reported in the system, however we will use a 30 minutes cap for the creation of sessions.

Ip	(unique address of the computer from which the page was requested)
Identity	(identifier of the user -if logged in from a private network)
user	**(username associated with the user –if the user is authenticated)**
date	
time	**(date and time are used to compute the dwell times)**
timezone	
method	(type of request, i.e. GET is used to download a page)
path	**(the full URI of the resource)**
protocol	(definition of the type of exchange, most web-based requests are HTTP)
status	(response code from the server)
bytes	(size of the resource/page)
referer	**(full URI of the resource from which the path was requested, i.e. parent page)**
agent	(information about the users' browser and operating system, i.e. Mozilla on Windows XP)

127.0.0.1 - **lorenzo** [10/Oct/2000:13:55:36 -0700] "GET **/pageA.html** HTTP/1.0" 200 2326 "**http://www.example.com/start.html**" "Mozilla/4.08 [en] (Win98; I ;Nav)"

127.0.0.1 - **lorenzo** [10/Oct/2000:13:55:36 -0700] "GET **/apache_pb.gif** HTTP/1.0" 200 2326 "**http://www.example.com/pageA.html**" "Mozilla/4.08 [en] (Win98; I ;Nav)"

127.0.0.1 - **lorenzo** [10/Oct/2000:13:57:36 -0700] "GET **/pageB.htm** HTTP/1.0" 200 2326 "**http://www.example.com/pageA.html**" "Mozilla/4.08 [en] (Win98; I ;Nav)"

127.0.0.1- **lorenzo** [10/Oct/2000:13:57:36 -0700] "GET **/imagebanner.gif** HTTP/1.0" 200 2326 "**http://www.example.com/pageB.html**" "Mozilla/4.08 [en] (Win98; I ;Nav)"

Figure 8.1. The structure of a web log entry.

At the top the list of chucks of information available in web logs with a focus on the information used. In the bottom frame an example extract from the log highlighting the core elements.

In the last paragraph we used the term *sessions*: the basic assumption is that a user is in a session when there are a number of actions performed in a continuous sequence, and that this sequence doesn't exceed the timeout limits. For the purposes of the analysis of log activity, whenever a timeout occurs the session is considered over, even though the following one might continue from the last page after a certain amount of time (i.e. after the session timeout). To clarify, if a user visits page A, B and C, and then goes for a coffee allowing the system to lapse for over 30 minutes, then comes back and starts again from C, D and E, these two streams will be considered effectively as two separate sessions.

Sessions for each user have to be built from the log file using an algorithm to identify the key nodes and the links between them. In the example (figure 8.1, using the paths) we have a simple example:

Start.html -> *pageA.html* -> *pageB.html*

In this instance only the content was filtered out to identify the sequence, however in the example there are two types of files: documents with extension .html are the pages or containers, the two .gif files are pictures embedded in the pages. Normally the media pages ‘belonging’ to a container page are removed from the database to identify only the core content.

Sessionising the log is an essential step to find all possible link pairs and nodes which were actually used (i.e. visited). A limitation of this approach is that only pages which were visited at some point are noted, which makes it difficult to reconstruct the whole structure of the website without a prior knowledge of the overall organization. However in chapter 5 we produced a notional tree structure to explain how the websites used look like, focusing around a small set of semantic containers.

The information from the log is rich, but limited, in the sense that no information is available about what pageA and pageB actually contain, nor what their function is in the overall site. Knowledge of the site structure and the function of each element must be added before an interpretation about behaviours can be done and in chapter 5 we detailed how the meta-information about elements was categorised before it was integrated with the log information.

8.1.2. Data reduction, aggregation, enrichment and visualization

One of the most striking problems emerging from the web logs is the huge amount of data present when the records cover an extended period of time. As we have explained earlier, it is reasonably easy to individuate a single session and represent a graph which shows the nodes and links as well as the time distributions of clicks (i.e. Figure 5.5). However, things start to look messy when one wants to represent multiple sessions over time, both for a single user or aggregating the data for a group.

Web analytics is a fairly new trend in e-commerce and its purpose is to make the data intelligible from the business point of view. So far, however, web analytics has been useful in diagnosing unwanted behaviour from users and the lesson learnt are applied to maximise traffic through certain pages or resources. A typical example is when a customer visits a popular commercial websites: how many browsing customers go to the checkout page and purchase items? From the business point of view this last page is the *conversion page*, the single resource which can deliver sales. It is obvious how important it is for the business to ensure that as many customer as possible get to this stage in the click stream as quickly and as simply as possible.

This is quite different for educational resources: to start with, there is no single page which can be singled out as essential. Competing goals might be in action within a single session: the browsing intentions are also more difficult to categorize as one student might simply want to obtain a specific document associated with a specific lecture and then leave. Another student might want to get information about the course and follow the path leading to course admin documents, and then pursue this further by reading the discussion forum. Yet another might want to specifically test the knowledge acquired after reading the textbook and following a lecture, by attempting a self-test with multiple choice quiz.

The real *conversion*, in these three examples is the knowledge/information that the students have fulfilled their goal during the session, but this cannot be specified with certainty, it can only be inferred from the sequence of actions performed and only if metadata about each element is interpreted alongside the log data, enriching the database.

In the last chapter we have considered how, in principle, data mining techniques could be used to provide a reasonably simple solution to make sense of a large database. In the following sections we will look at different *views* to provide a better insight and determine what type of information best conveys a useful characterization of behaviours. It is important to remember that this is a very dynamic and young field of research, which is gaining great impetus mainly from e-commerce. Often the techniques used are very much driven by the economic needs rather than the academic interest, therefore they need some adaptations for the educational context.

8.1.3. Static View (snapshot)

A static view of the data is offered by common web analytics reports. There are a number of free and commercial solutions which can be used to produce a report of usage. One option is to use offline software to read the log files and produce a variety of summaries, another is to use a 'live' tracking system which adopts snippets of code embedded within the content (i.e. Google Analytics is the most popular free solution since 2007). WebCT had its own administrative tools to provide reports of activity, but these were simply too rudimentary to be useful for a more in-depth analysis. In fact the STEER project mentioned in chapter 5 (Hardy et al. 2006), and precursor of the work conducted for this thesis, offered a solution with live tracking, which highlighted the shortcomings of this method of analysis. For this reason we decided to go back to the original server logs which, although they presented a number of practical challenges, are more accurate.

Based on the weblogs, one can derive and categorise a number of descriptors for the activity based on **time spent** and **frequency** of access of particular pages. A step further is achieved by the ability of identifying the content; however these metrics are aggregated from the log without any further manipulation. For example Figure 8.2 shows a static snapshot of the frequency and patterns of access for a specific page (Lecture Notes hub) in the Y1 course (academic year 2007-08). The figure highlights the fact that students tend to visit the resource more often earlier in the week, coinciding with the days in which the lecture is given (held on Monday, Wednesday and Friday between 11 and 12) with a peak in the hours following the lecture. An exception is Friday in which tutorials were available after the lecture and most students had other tutorials in the afternoon (this information available from the tutorial signup preferences).

The last paragraph is interesting for a number of reasons: on one hand it provides a good insight into what the students do as a group with a *specific* resource, on the other hand, it would be impossible to explain the patterns without the contextual information about their schedule and preferences which is not coming from the logs.
In other words the static view, typical of web analytics in e-commerce, can provide good answers for *what* and *when* questions, but little or no information about *why* student behave in a certain way, even less how to interpret their actions.

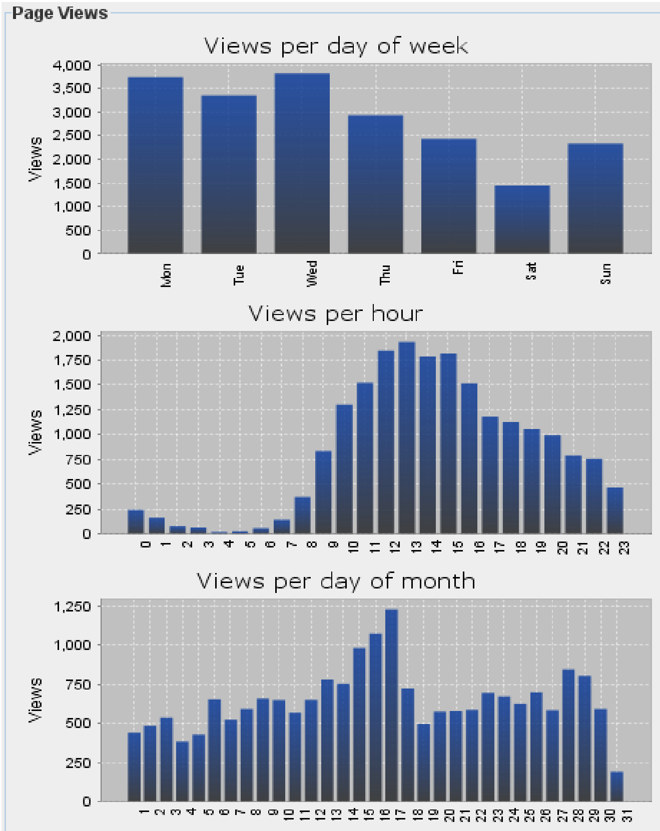


Figure 8.2. Number of page views for the 'Lecture notes' hub page for the PS0001_11 course (academic year 2007-08).

The focus is on the aggregation of data specifically looking at the content rather than the individual use. This can provide an idea of typical activity for each of the courses and it is certainly informative to evaluate how e-learning is used, but it is also extremely limited in directing pedagogy.

measure	unit	metrics
Visits	semester	total
Time spent	weeks	average
	sessions	skew of distr. over time
		kurtosis of distr. over time
		ratios time per unit
		ratios visits per unit

Table 8.1. Metrics characterising the extent of use of e-learning. The combination of the elements provide the indicators (i.e. total number of visits per semester)

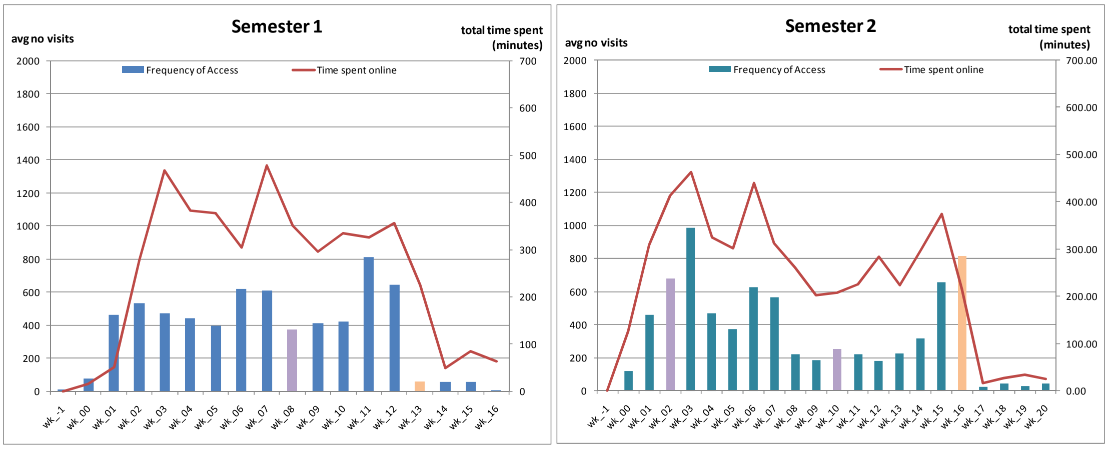
The first level of analysis, already labelled a *static* view of usage, offers a number of useful insights about the *extent of use* of e-learning. In fact, using three basic units as aggregators (i.e. semester, week and session) of two core measures (i.e. time spent and number of visits), a number of metrics were generated and can be summarised in table 8.1.

Based on these metrics it is fairly straightforward to produce an overall picture of activity. In the next few pages this is exemplified by presenting a parallel of two courses in year 1 and year two (Figures 8.3 to 8.6). The choice of the courses is not incidental as these are offering a better insight on the variation of behaviour moving from the psychology 1 to the psychology 2 course, but the patterns demonstrated are similar for the other courses at the same level.

In Figure 8.3 and 8.5 we plotted the activity over the two semesters for each week highlighting the time points in which a form assessment was due (i.e. coursework or exams). The metrics used are the average number of visits (per week) and the total time spent (per week). We specifically use total time and average number of visits as intuitively one can work out the average time per session in each week. The semester windows were slightly different in different years, therefore, to maintain a constant frame we used 17 weeks in the first semester and a 21 weeks for the second. The -1 is the week before the term which, as evident in figure 8.3 in semester 2 is the continuation from the previous graph. Although in the second semester the Easter holidays leave more time to prepare for the exams than in semester 1, the idea was to offer a roughly equivalent 20-weeks span for both semesters. Figures 8.4 and 8.6 give an overview of the average time per page and number of unique pages visited each week. As we will see in the next section the number of unique pages is obtained after overlaying a structural interpretation framework to the log data.

The two types of graphs offer a valuable insight in the patterns of visitation expanding the understanding of how the system is used and hinting at the relation between the forms of assessment and the pattern of use. For example in both Y1 and Y2 activity tends to spike before a piece of coursework is due or in the weeks before the exams. We can also observe a relation between time spent and number of visits overall (possibly hiding the fact that some might prefer longer less frequent visits rather than short and frequent). The second type of graphs shows the trends of the relation between the average amount of time spent per page and the extent of usage (i.e. number of unique pages, rather than the total pages, hinting at issues of efficiency which will be considered in the next section). Although these representations are informative, they raise more questions than answers and stress the limitations of this descriptive 'surface' approach to analysis common of web analytics.

Y1 starting in 2007



Y1 starting in 2008

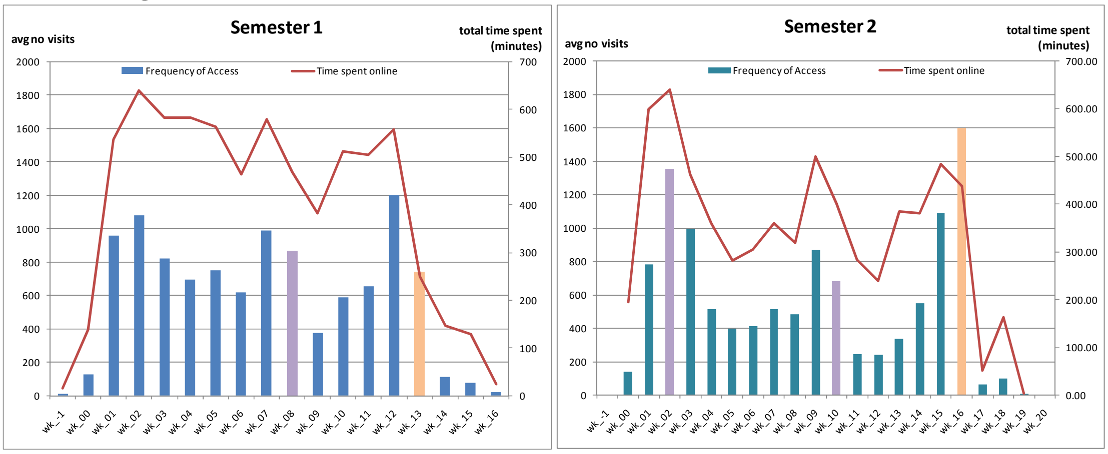
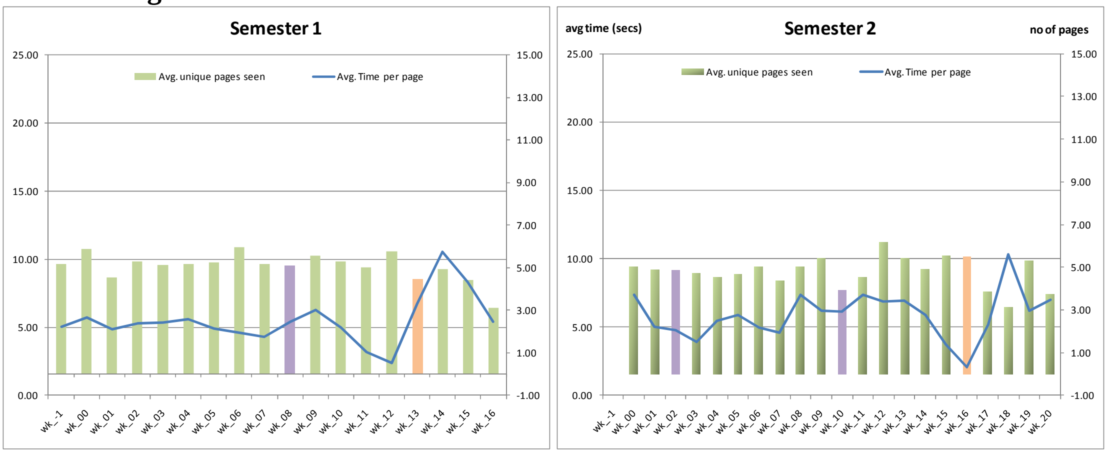


Figure 8.3. Parallels between two first year courses (starting in 2007 and 2008) across the two semesters looking at the average number of visits and time spent online each week.

We have also highlighted the weeks in which coursework was due (purple) and the exam week (orange).

Y1 starting in 2007



Y1 starting in 2008

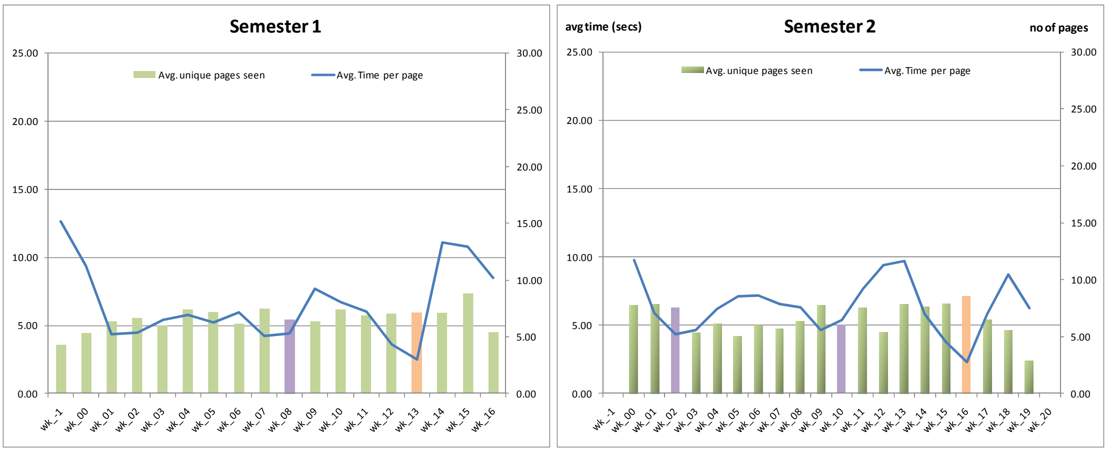
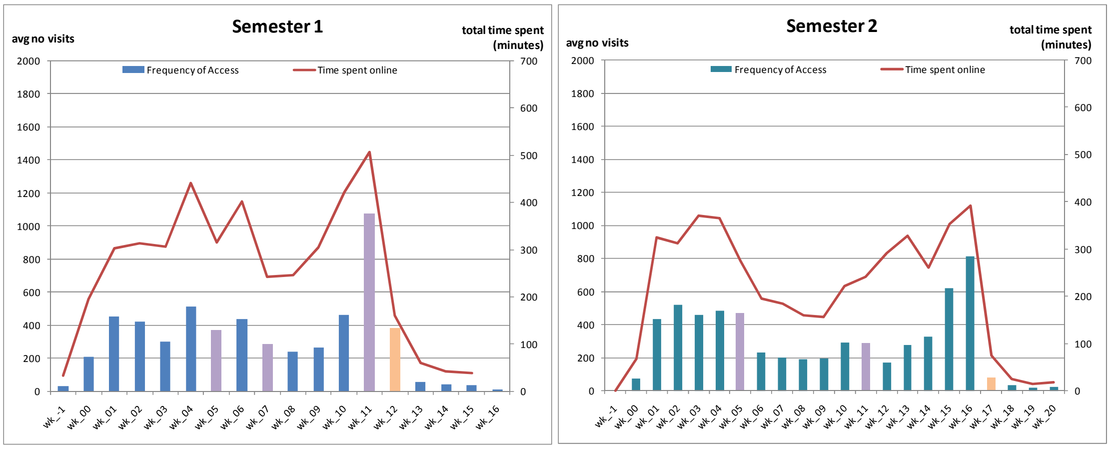


Figure 8.4. Parallels between the same first year courses (starting in 2007 and 2008) looking at the richness of visits (i.e. average number of unique pages seen in each session during the week, and average time per page).

Y2 starting in 2008



Y2 starting in 2009

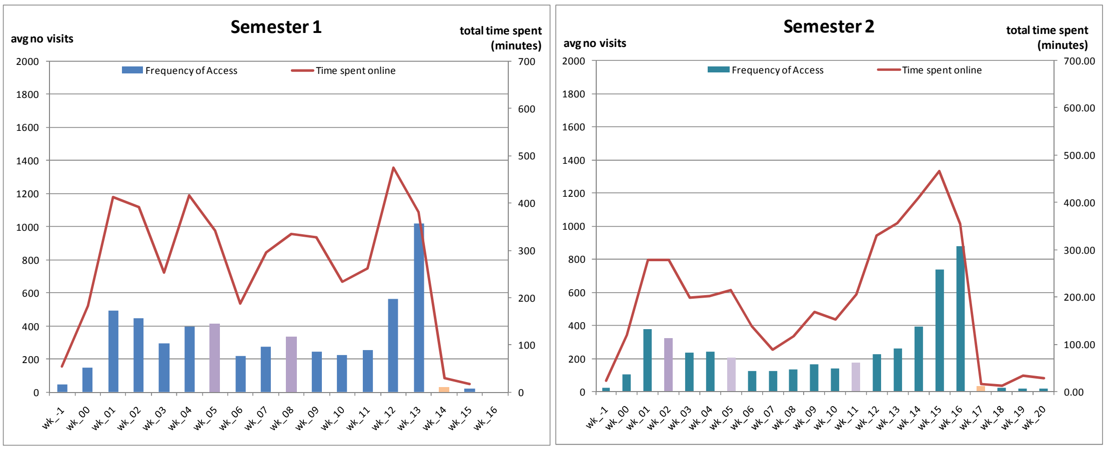


Figure 8.5. Parallels between two second year courses (starting in 2008 and 2009) across the two semesters looking at the average number of visits and time spent online each week.

We have also highlighted the weeks in which coursework was due (purple) and the exam week (orange). Note that the selection was done for courses offering an insight into longitudinal behaviours; therefore those in the top graph are mostly the same cohort who started their degrees in 2007 shown in Figure 8.3.

Y2 starting in 2008

Y2 starting in 2009

Figure 8.6. Parallels between the two corresponding second year courses (starting in 2008 and 2009) across the two semesters looking at the average number of visits and time spent online each week.

8.1.4. A Dynamic view of usage

The static approach to the analysis of usage is a good way to characterise the *extent of use* of e-learning. However, after gaining an overall picture of the access to the course, the focus must shift to the basic unit of activity: the session. The session is a single episode of interaction with the e-learning system which can be derived from the manipulation of the log data. Basically, sessions are temporal units; however, combined with the metadata and construed as episodes of activity for *individual* users whose aims are to achieve some goals, these convey a richer understanding that a mere description of behaviour.
In both cognitive psychology and human-computer studies (i.e. Anderson 1981) a task can be defined a priori in a rational way, even if the agent performing the task does not have a rational model of the task in their mind.

When working on web logs, interpreting the rationality of the goal-directed activity can be tricky; an example can be the difference between the appearances of rational behaviour in automata when the driving intention is a simple mechanistic application of programmatic rules. In the case of students' activity, a layered interpretation can help to understand the behaviour. In a lab experiment it is possible to control the task in such a way that a user has to explore a set of online documents to prepare for a test or search for specific information. In a VLE, the purpose of each visit doesn't necessarily imply that the user is goal-driven. As for many other activities online (i.e. checking email, searching for something or reading a blog) the new generation of students might simply check the VLE for anything novel or changed and are then happy to leave after landing on the entry/portal page which gives an overview of recent changes[16]. If this is the case a pattern of visitation with shallow branching might be the most typical example of a session.

In the last section, by using the frequency and timing, it was possible to derive weekly visitation patterns, average time spent online per week, and to find the length of each session. The most interesting characterization of activity, however, comes from the overlay of an interpretational frame over the data. Two were added: the knowledge of the structure and architecture of the system and a psychological one, looking at cognitive efficiency and intentionality.

---

[16] Note that since 2006 the portal/entry page could be customized by users, however we don't have any data available about the extent to which students personalized it.

To achieve this aim we used the meta-information associated with each document to categorise activities and constructed sessions applying the knowledge of the system and its structure. As this step is increasing the complexity of the data, data reduction techniques (i.e. clustering used in the last chapter) become necessary to abstract meaningful patterns. The key purpose is to transform patterns into knowledge and this is where the psychological framework becomes most useful.

Up to this point our understanding of the *extent of engagement*, could be easily represented by the static view of usage.
We already mentioned in chapter 5 how the structure of the website is important because it creates a closed-world environment in which only a certain type of behaviours can be expressed. The knowledge emerging from the inclusion of *structural* and *topological* details about the site will open up an avenue to understand *efficiency in browsing behaviours*.
A *semantic* layer, which uses information about the *function* of the content used, provides an insight into intentions via expressed sequences of actions. This, in turn, allows for a better interpretation for the *richness of use* in each visit.
To make these observations about usage more systematic, it is useful to adopt a framework based on three different level of analysis which will consider in turn in the following sections:

- A topological view (i.e. based on the structural features/organization of the website)
- A semantic view (i.e. based on the functional organization and nature of content)
- A user-centred view.(i.e. based on the intention of the user)

To support the analysis, new tools and techniques are necessary: the visual representation of the metrics about the extent of usage in the last section was an example of how visual presentation can help to summarise complex interactions. Chapter 4 detailed how a vast literature can be represented using visual maps. However, activity is dynamic by definition, adding an extra layer of complexity to the problem: how to represent the flux of activities meaningfully is a complex problem which requires the aid of different tools and/or disciplines. Data mining tools and information visualization techniques (InfoVis) provided a useful way to meaningfully represent the large amount of data.
In the following sections we approach the analysis specifically referring to *views of the data* and we will introduce tools and techniques which helped in the analysis.

8.1.5. A topological view of usage

When a website is made available online, the interface is an essential tool for the user to infer its underlying structure. Some websites have explicit site maps, but more often users learn the structure by exploring the links available.

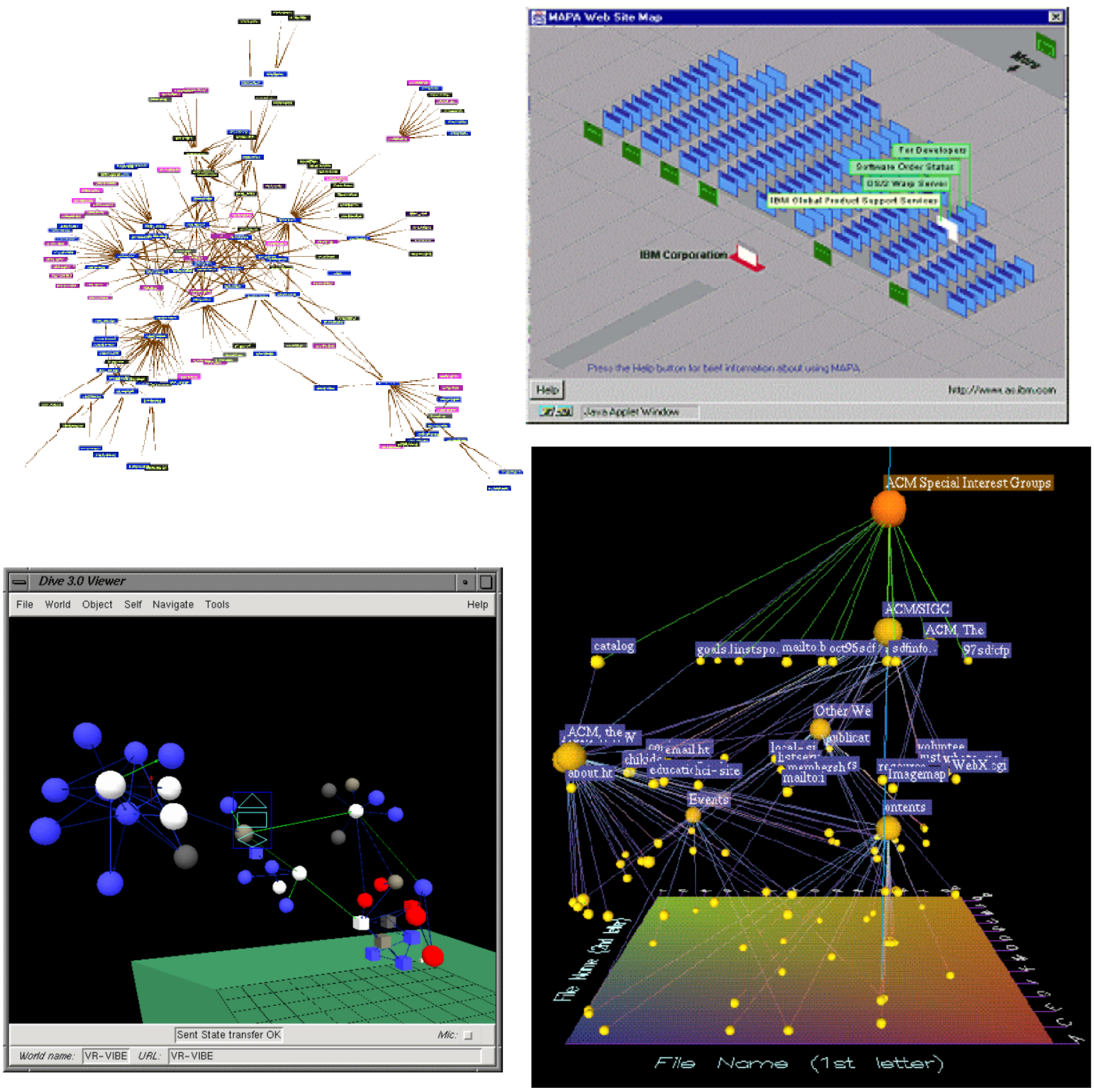


Figure 8.7. Four examples of website visualization.

From the top left in clockwise direction: using a radial tree (Cugini & Scholtz, 1999), hierarchical 3d formats (Durand & Kahn, 1998), a graph-like structure in 3d space (Shiozawa, Okada, & Matsushita, 1999) and a 3d 'molecular' structure (Ingram & Benford, 1996) .

A core problem already been identified in deducing the site structure from the logs as the structure which is generated has fewer pages than the real site (i.e. some links don't feature in the list of *used* pages). However, if one starts from the real site, this would ignore dynamic content which is not initially available (i.e. results from the search pages or threads of discussion in the forum). What is not used is less important in the study of expressed behaviour than if the website was the aim of the evaluation. It was therefore decided to focus on the structure emerging from the log from content and, where suitable, use the knowledge of the structure for tests and activities. This provides a balanced approach to include as much as possible of the website content.

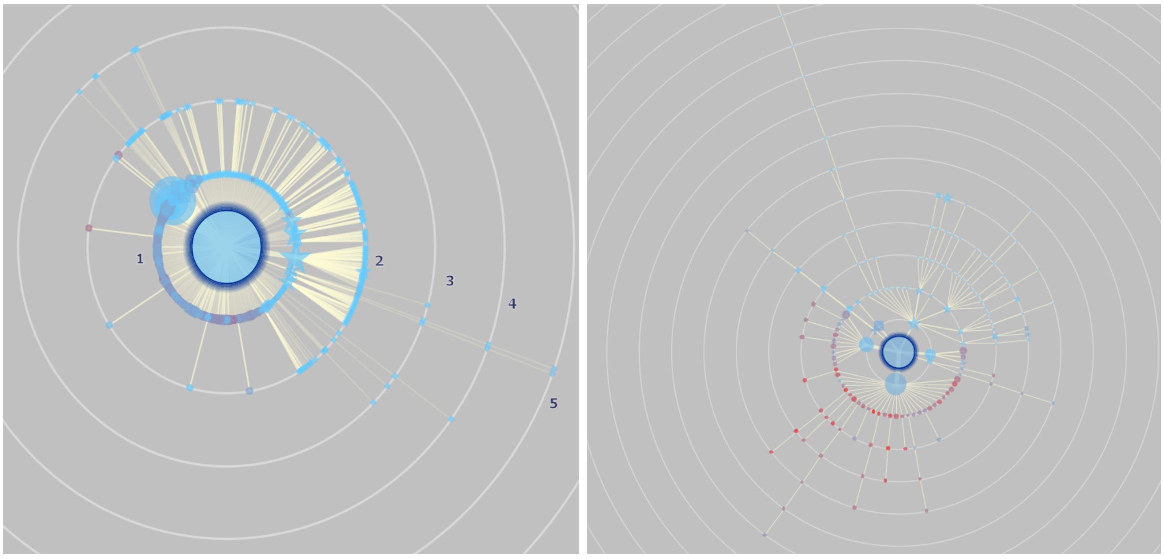


Figure 8.8. A radial representation of a web site structure.

(left) A radial representation of the home page and the minimum distance (number of clicks) to resources in the website. (Right) The home page and real paths taken by users to reach a resource. This is the same site on the left, but it shows how users rarely pursue the shortest path to resource X with up to 20 clicks in a session. The visualization on the right is also less cluttered organising the location according to the type of resources visited.

A website can be represented in a variety of ways and the literature on web analysis is rich in examples showing a variety of techniques to represent either the structure or the paths taken by users (see Figure 8.7). A typical way of representing a website structure is the tree view. However, the hierarchical tree becomes impractical when there are numerous nodes at lower levels, therefore a radial variant is often used to maximise the visibility of items in 3d space. Figure 8.8 displays a different approach to solve the same problem: there are 3 major problems which Information Visualization (InfoVis) tools are facing with the particular case of websites: 1) the amount of data which is available becomes cluttered in space when there is too much data; 2) the way in which meta-data related to each node is difficult to represent (i.e. labels with page names); 3) the links to be shown become particularly complex with the complexity of the website.

Recently, a solution was proposed to simplify the problem of visualising complex web structures in 2d space with WET (Web Exploration Tool, Pascual-Cid, 2008; Pascual-Cid, Baeza-Yates, Dürsteler, Minguez, & Middleton, 2009). The authors used a radial structure like the one in the figure 8.3 in which, as well as website structure, other useful information could be represented dynamically. This is a simple yet powerful solution to create interactive meaningful visualizations of the structure of the website, and was used successfully in supporting the analysis of educational websites (Estrada and Quixcal 2009, Pascual-Cid, Vigentini and Quixcal 2010).

The example in figure 8.8 shows a partial view of one of our courses: the structural representation (left) shows the maximum distance of the pages from the home page (node in the middle), the links between resources at different levels, the 'type' of resource marked with a particular symbol, the number of times it has been used during a given period (size of each node) and the number of times a page was an 'exit page' (i.e. the last item viewed in a session) represented by the intensity of the edge (from red to dark blue). The figure on the right shows the real usage, or common paths to reach content.

Although these are snapshots of the activity in the system, WET provides an exploratory interactive environment in Java, which is dynamically updated, allowing the analyst to render visually insights which the traditional metrics from web analytic tools cannot easily provide. This tool was used extensively in the exploration of activity. It will provide some material in this chapter as the result of an extensive collaboration between me and Victor Pascual-Cid to adapt both the tool and the methodology to represent particular features within the tool. More practical details about further data processing required to go from the web logs to the graphs are provided in Appendix 4.

To demonstrate the effectiveness of WET in exploring the data giving the visual abstractions of the site structure and organization, figure 8.9 to 8.12 provide the visualization of some individual courses (note that it was not possible to convert the _7 and _8 courses, i.e. years 2005/2006) to a suitable format for this representation) and the corresponding usage pattern highlighting two core features: on one hand it provides a useful demonstration of the similarity in structure across the courses, on the other it demonstrates differences in visitation patterns between the first and second year courses, which could be considered as the result of a 'path to expertise' in using the material.

In fact, it is evident from the comparison of the Y1 and Y2 graphs that students in Y1, novices in using the system, present much longer paths than Y2 students and their patterns of visitation are generally less efficient. The graphs presented in the previous section have already pointed out the temporal pattern of students who tend to cram usage toward the end of the term. This type of graphs shows another difference between Y1 and Y2 students: in Y1 it is more typical to find longer visits to obtain all information needed. It is possible that in Y2 students simply become more effective and strategic in their visits, knowing what to look for and visit the site more regularly.

Learner control is expressed in the confidence of usage: figure 8.10 shows how students in Y1 and Y2 use the search function. Whilst in Y1 students seem to struggle in finding material and invoke the search at deeper levels in their visitation sessions, Y2 seem to be more efficient and perform searches earlier in the sessions (usually within 3 clicks) and use it less (number of branches). The comparison from students of the same cohorts is quite revealing in this case and provided the evidence for further analysis of efficiency in browsing.

In the next few pages a number of graphs are shown to represent the website structure and patterns of visitation. The type of shape is clearly associated with the abstractions described in chapter 5. Please note that it was not possible to maintain a fixed colour scheme. Most notably the reference to social activities which is shown in pink or blue in some courses, however the shapes are consistent.

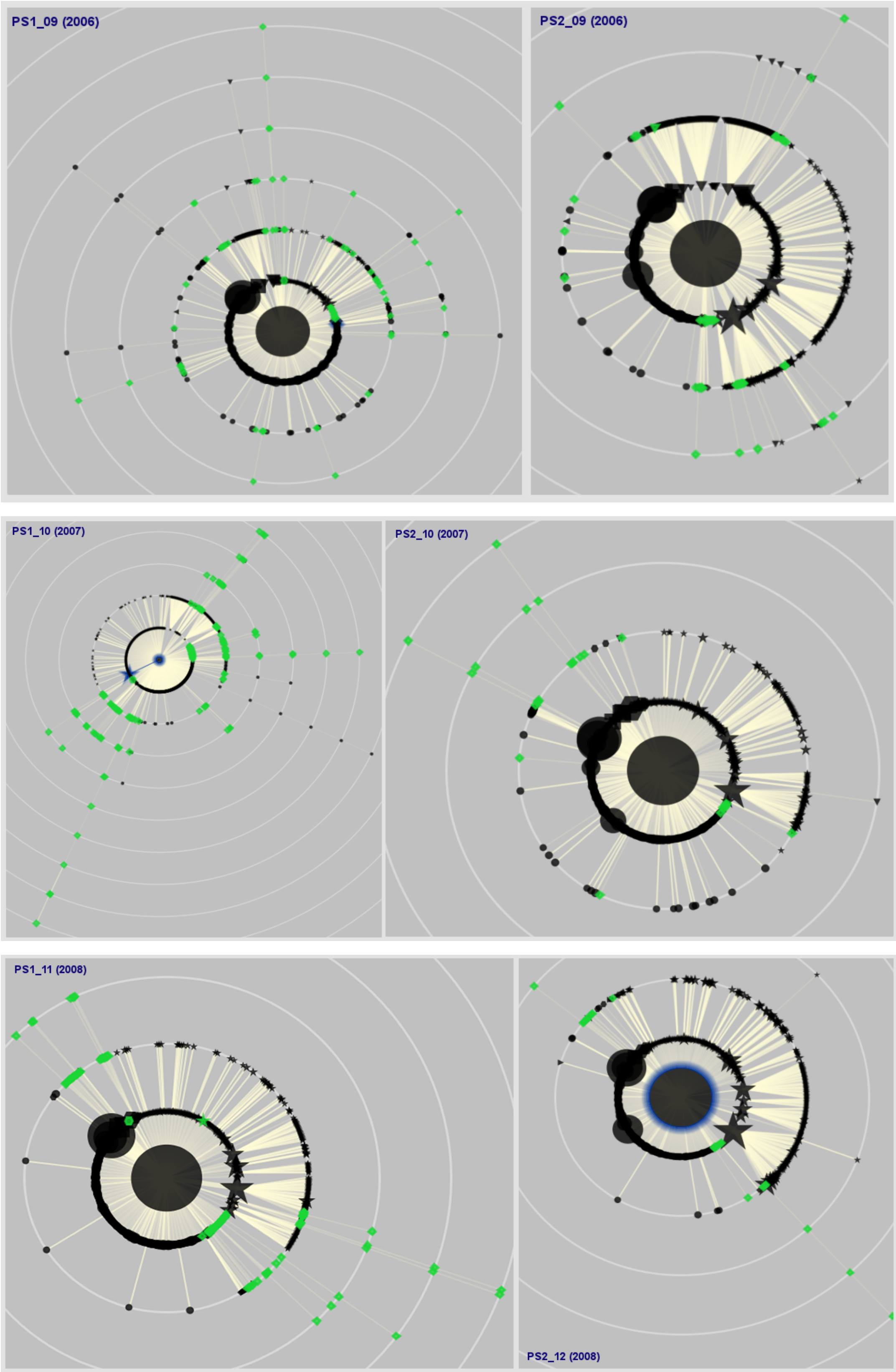


Figure 8.9. A radial tree of three parallel courses in Y1 and Y2. Each circle is a step down in the site hierarchy. Y2 courses are generally more compact, with shorter minimum paths to reach any content. Highlighted in green the dynamically generated search results pages at different levels showing the use of the search function at deeper levels, size represents number of visits.

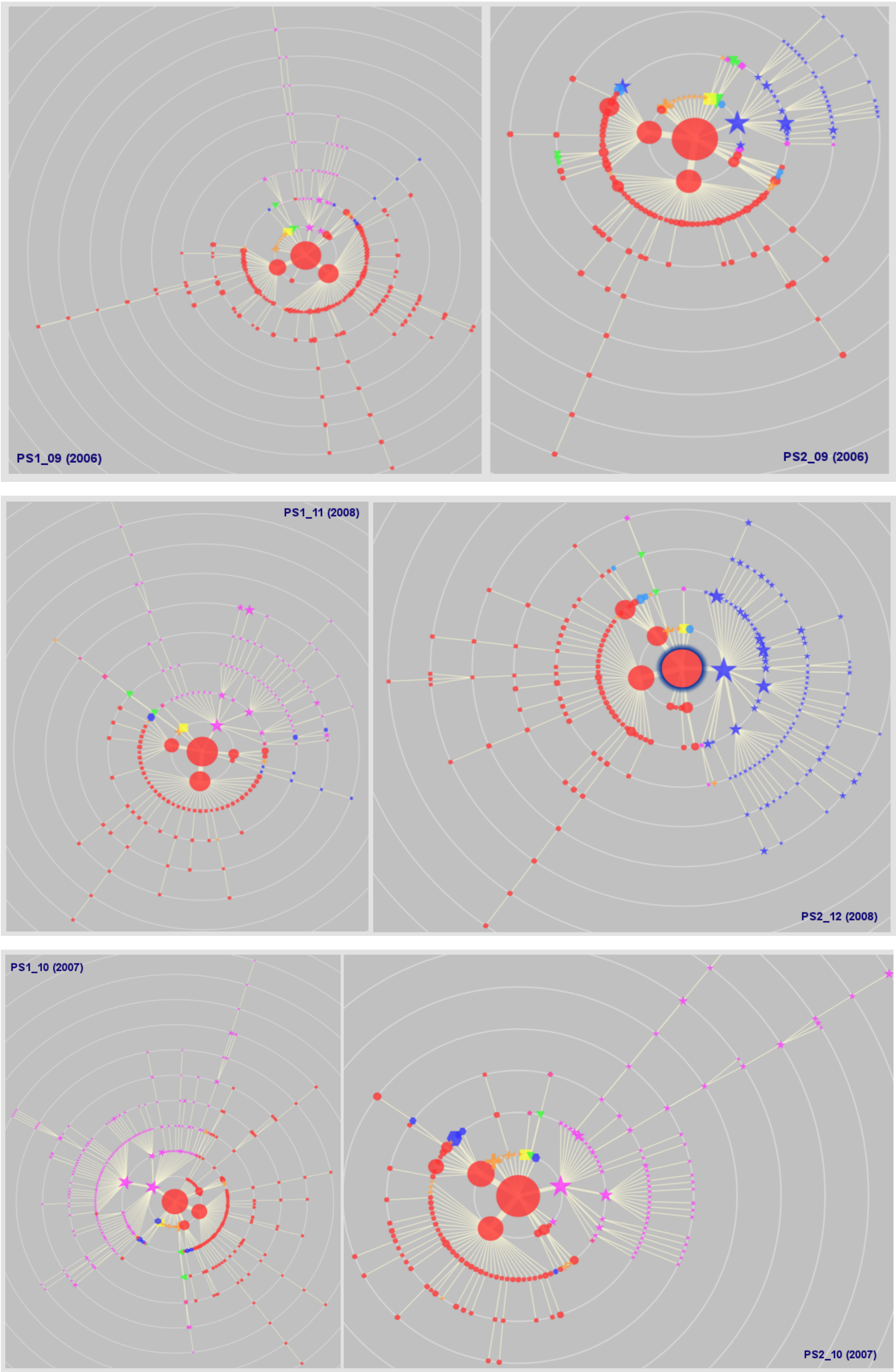


Figure 8.10. Overlay of usage (number of visits using the size of shapes to enrich the common paths representation in WET. The strength of this representation compared to figure 8.10 is that it demonstrates the difference between the ideal and real paths taken by the students. This provided the basis to introduce the efficiency measures to characterise individuals.

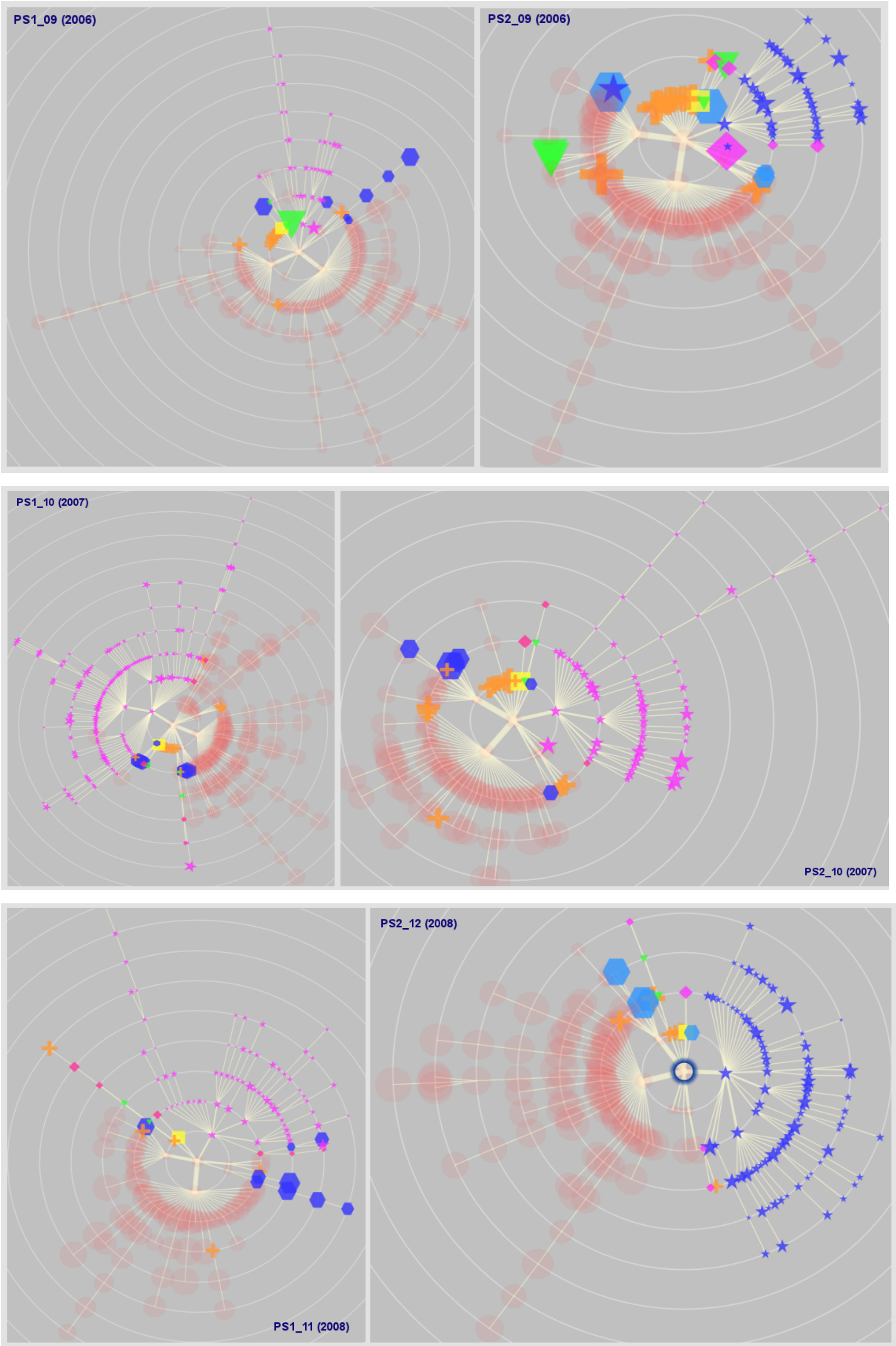


Figure 8.11. Overlay of usage (time per resource, using the size of shapes). In this type of graph we shaded the prominence of the time spent on core content to demonstrate how the other resources also play an important role, often presenting similar usage to the core content. Particular attention should be given to the social aspect of the recorded usage. This provides an excellent foundation to the concept of richness of use.

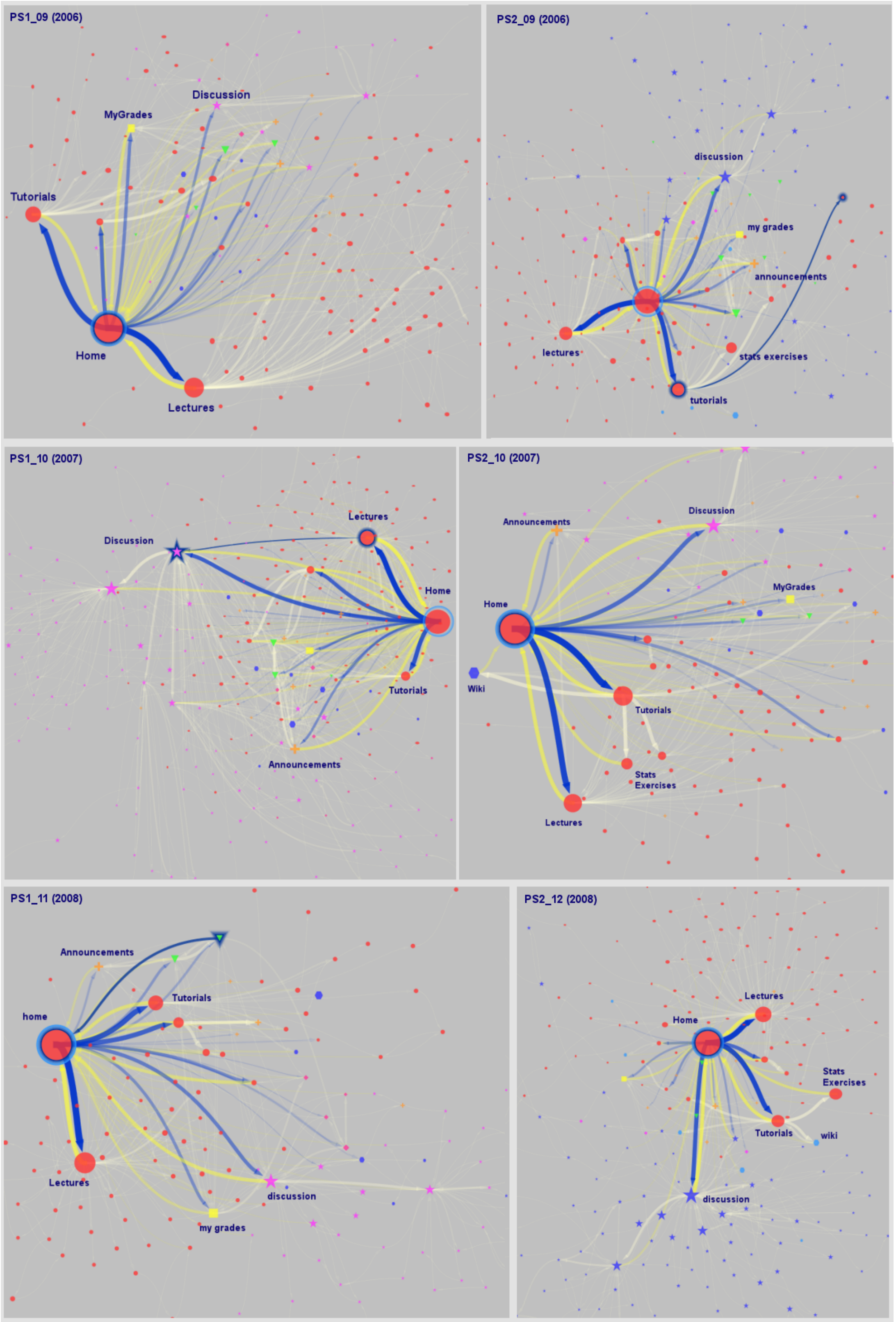


Figure 8.12. The site galaxy is possibly the most powerful representation of usage as the size of the arrows linking the various resources clearly illustrates the most common paths with the strength of the lines directly related to the frequency of use. The size of the items shows the frequency of access and the combination of colour and shape is used to identify resources. More than for the other graphs this is showing the value of the semantic overlay on usage.

The observation of the shortest paths (structure of the site 8.10) and the usage (8.11 and 8.12) provide even more evidence that the way in which resources are used in the courses *is* different. For example the depth of the branches and the comparison of the extent of usage (i.e. number of visits in 8.10 and time in 8.11 equating to the proportional size of the shapes) for different types of content is quite striking. In Y1 all branches tend to be longer, in Y2 the branches representing content (circles) and social activities (stars) seem to indicate that the discussion forums are read in longer episodes all together, whilst content is accessed more frequently in short and efficient operations.

The final type of graph, (8.12) displays the site galaxy: as the heat map in chapter 4 this unusual representation of the site traffic offers an immediate sense of the organization of the site and provides a vivid impression. In figure 8.13 we highlighted the core items with the size of the shapes proportional to the number of visits over the entire year and show outgoing (blue) and incoming (yellow) links in which the size and intensity of the arrows represent the frequency of usage. The number of items displayed and intricacy of the network are quite useful in demonstrating the differences in the modes of access which require a more in-depth characterization.

The representations offered with WET provide a simple and powerful way of representing visitation patterns in a meaningful way. Although the topological representation of the site structure provided useful details about how a user might go about using a web site, these representations have limited use in understanding *why* a user takes a specific path.
To a certain extent, the website structure *forces* the user to take certain steps: Pirolli and colleagues talked about the *scent* of certain nodes and Graff and colleagues demonstrated that different site structures could affect navigation based on the style of the user. In the representation of figure 8.12 the typicality of paths showed that the use of abstractions suggested in chapter 6 (shapes and colours of nodes) could provide the means to discriminate between different episodes of usage and between different types of users.

From the graphs it seems even more plausible that it would be possible to identify styles of use. In fact, in the interactive process between individual preferences and site structure, Herder and Juvina (2004) extrapolated navigation styles (flimsy and laborious): from the graphs we can think about a deep and surface approach to browsing implemented for different types of content. Such patterns however, need to be quantified and without a meaningful exploration of the data it becomes difficult to grasp which metrics might be useful.

One clear feature emerging from the graphs generated with WET is that certain structural features of a dynamic website actually provide information about behaviours. For example, the use of the search function in WebCT is particularly useful to explain the concept. The search box, like in many other sites, is an omnipresent feature of the user interface. The graphs (8.11) act as a reference to the search results pages. At any point in the click stream the user could decide to use the search. The following trace in the log would be the dynamically generated results page. Now, if a user needed 5-7 clicks to reach a particular resource and such resource attracted a lot of traffic, this would be interpreted as a case a bad usability of the site. However, in the case of dynamic content, this is an expression of behaviours. Three different examples will help to clarify this interpretation: in one instance, a user logs into the system and as first action points to the search and follows the appropriate result. In a second case the user might click through some links after log in, and then decides to search for what they were looking for in the first place. Yet a third example could be, like in the graph, cases of persistent searches in which users literally hop from search to search refining the key terms in long sessions. These three cases could be labelled as ‘confident’, ‘explorer’ and ‘clueless’ users, and as will be explained in the next section metrics could be computed to represent these patterns.

8.1.6. A semantic analysis

Neither a topological analysis nor the click-stream can provide a sensible insight on users’ intentions regarding a single visit.

Psychologists, however, are very much used to observe behaviours and this can provide a different approach to abstract intentions. In chapter 6 a semantic organization of the content was provided (table 6.11). By recoding sessions into *episodes* of interaction it is possible to determine the presence or absence (with a binary coding 0 and 1) of a specific observed behaviour and compute the frequency of occurrences. Table 6.11 provided a distribution of the different tools and actions in the whole database for the different courses. However there is an obvious disparity between certain types of resources which, as a consequence, produce some noise in the data caused by the combinatorial explosion of possible paths that one might take in their visits to the website. As a matter of fact, the long tail of the distribution of items accounts for no more than 5 to 10% of the variance in the usage activity.

In a theoretical model, using the 10 abstractions proposed earlier (table 5.13, p 199), the combinations of possible binary sequences are (2^{10}=1024) which is a big, but reasonably sized set. However, using the 53 actions would generate a set of 9007199254740992 possible sequences, making the categorization of each unique session more problematic. Another observation is that in a similar way to the length of sequence of pages visited, most of the session are short: this is also reflected in the number of intentions represented by the abstractions proposed in which up to 70% of the episodes of interaction can be categorised as having a unique intention or 2 goals (i.e. enter -> examine core content -> check information updates).
The statistical analysis of groups based on such differences becomes meaningless unless enough episodes contain sufficient occurrences of such sequences.

As specified in the extent of use in the last section, aided by the knowledge of the structure and value of the content, sessions were detailed using the metrics of richness of browsing and efficiency mentioned in the last section and listed in table 8.2 and 8.3.

measure	unit	metrics
depth	sessions	total
time	pages	average
occurrence		ratio time per page
re-currence		ratio time per resources
type of category		unique content

Table 8.2. Richness of browsing from the combination of measures and units.

The combination of the measures with the different units listed in table 8.2, allowed the generation of a number of metrics which provide a detailed characterization of each session (for example the average depth per session).

The categorization of sessions is necessary to create a dataset defining the nature of each episode based on intentions, length, richness and depth of the visit and a specific localization in time. Efficiency can be used both at the level of sessions and for individuals. For example, using the case of search mentioned earlier, the number of times a user interrogates the system and at which point in the session the action is performed.

measure	unit	metrics
depth	sessions	unique content
time	pages	efficiency ratio for pages
occurrence	sessions' clusters	efficiency ratio for abstractions
re-currence	individual student	use of search
type of category		

Table 8.3. Efficiency of browsing from the combination of measures and units. Note that units for this schema apply to both sessions and individual in slightly different ways..

To circumvent the problem of multiple categorical variables leading to meaningless statistical analysis due to the lack of sufficient cases in each group, we used again the Two-steps cluster algorithm introduced in the last chapter on the sessions.

Two different solutions were explored from the result of the clustering on the richness and efficiency metrics. The first considered all the metrics described in table 8.4. The other included an extra metric for social activities (i.e. the proportion of type of resources used in the session with a social- rather than content-based purpose) in the algorithm.
Although the addition of this variable reduced the number of cluster from 6 to 4, none of the following analysis seemed to be affected; therefore we opted to use the first type of analysis as social metrics will be used to define the users.

At this point a practical issue emerges in explaining the clusters. Already in the last chapter we showed that when the number of variables fed in the clustering algorithm is small, it was fairly straightforward to identify a cluster and define it. The simple example of the CSI clusters with the two dimensions and the frequency of genders in the different groups was clear enough. It was more difficult to precisely define the clusters of styles for the ASSIST.
A similar problem arises with the definition of the clusters for the sessions in which 64 categorical (including presence/absence of abstractions, tools and the week of term) and 11 numerical variables are contributing to the solution. Table 8.4 shows the variables which contribute to the classification of each session

Metrics used	1	2	3	4	5	6
Duration of session (time)	55.48	107.15	65.98	42.36	27.11	28.99
Depth of visitation (length of sequence)	12.04	43.97	19.14	10.38	2.57	5.73
Number of tools	3.92	3.55	5.78	2.73	1.35	1.76
Number of abstractions	2.92	1.31	4.19	2.30	1.20	1
Average time per page	4.82	17.69	4.99	5.58	18.89	5.88
Progressive session count (per user)	56.64	42.92	36.77	62.86	49.48	53.38
Efficiency ratio for pages (Number of clicks/unique pages)	0.81	0.68	0.67	0.72	0.94	0.79
Efficiency ratio for tools (Number of clicks/unique tools)	0.71	0.50	0.75	0.80	0.97	0.89
Efficiency ratio for abstractions (Number of clicks/unique abstractions)	0.73	0.92	0.71	0.80	0.99	1
Time per Intentions ratio (ratio of duration per abstraction divided by time per page)	4.20	30.23	4.34	4.40	2.10	5.31
Time per tool ratio (ratio of duration per tool divided by time per page)	2.91	13.67	3.15	3.57	1.89	2.92
Week of term (from -1 to 20)						
Abstractions string (i.e binary sytring of 10 items)						
Tools string (i.e. binary string of 53 items)						
Prevalence of content/social (tot social features)	0.2	12.49	7.23	5.21	0.18	0.47
Total session in each cluster and percentage	17429	9907	11519	37191	29837	43357
	11.68%	6.64%	7.72%	24.92%	19.99%	29.05%

Table 8.4. Metrics used to characterise each session. The final algorithm contained 75 variables making it difficult to verbally define each cluster precisely. The centroids (means) of the numerical variables for each cluster are listed.

Looking at the centroids of each numerical variable included provides a better understanding of the features of each cluster. For example, it is easy to see the difference between cluster 2 and all the rest in the duration of the sessions. Smaller variations distinguish some of the other variables. In combination with table 8.5, which provides a breakdown of the distribution of each type of session, these details are useful to gain an overall picture of the different types of sessions and the courses in which they feature most. A reminder of what was said earlier about the inability to reduce all usage to a common denominator in full is evident in table 8.5 in which some types of session do not feature at all.

		TwoStep Cluster Number (core metrics)						Total
		1	2	3	4	5	6	
Course	**PS0001_07**	2301	3967	0	0	4778	0	11046
	PS0001_08	3846	564	5	45	12133	4	16597
	PS0001_09	39	123	1634	3763	662	7737	13958
	PS0001_10	6	233	2865	9393	1458	8139	22094
	PS0001_11	3	148	2538	6556	1751	13451	24447
	PS0002_07	213	1613	0	0	781	0	2607
	PS0002_08	11008	2611	0	0	6629	16	20264
	PS0002_09	11	258	3241	6734	454	5322	16020
	PS0002_10	2	114	706	6177	730	4072	11801
	PS0002_12	0	276	530	4522	461	4616	10405
Total		17429	9907	11519	37191	29837	43357	149239

Table 8.5. Distribution of the clusters' types in the different courses

The next step in the analysis was to shift the focus from the episodes of activity to the individual users and how their activity patterns could be characterised.

8.1.7. A user-centred view

In addition to the metrics of extent and richness of usage, we mentioned in the last section the use of metrics of efficiency of browsing. Some metrics like the *efficiency ratio* for pages and abstraction make sense for the individual sessions, but, when looking at a number of episodes over a long period of time, these measures provide a valuable insight into cognitive efficiency. For example, in a very different domain, McGonigle and colleagues (McGonigle and Chalmers 2001), studying long episodes of learning tasks over time, observed that efficiency of learning of new sequences could be quantified, and this was an excellent indicator of executive control as well as a measure allowing them to assess cognitive growth and compare task performance between children, capuchin monkey and human cases affected by cognitive deficiencies (autism and dementia).

Two new metrics were introduced at this point. The efficiency ratios for pages and abstractions are quite simple: the ratio is calculated by dividing the number of unique items (pages or abstractions) found in a sequence and then dividing by the number of clicks. The values generated have a maximum of 1 in which maximum efficiency is achieved (i.e. a student who targets a particular resource and obtains it in the minimum possible path). Although this might seem trivial, in longer sequences, if a student moves back and forth numerous times, the sequences contain a number of repetitions: whilst it cannot be explained from the sequence why a student moved back on his/her steps, the efficiency ratios allow a quantification of such behaviours.

The second type of index is slightly more complex and it is used to quantify the average unit of time dedicated to different categories of activity (i.e. tools and abstractions). This measure is computed by calculating a ratio of the duration of a session by the category and divided by the time per page. As the ratio is proportional to the amount of time per page an individual is dedicating in each session, effectively the value obtained is an indicator of how many ‘time units’ are allocated for different activities.
Furthermore, a set of measures to evaluate the specific use of social activities (like the discussion forum) was introduced.

A problem realised at this stage was the fact that not all students had data recorded for all classes, nor did all students who used the system have measure of styles recorded. For these reasons 2 separate analyses based on the year class were conducted.

As was the case for the sessions, a number of metrics were used to characterise students. These are listed in tables 8.6 and 8.7 and demonstrate the wide-ranging scope of all the three interpretational schemas adopted so far. The extent of use is summarised in aggregated measures of individual use for different units (i.e. page, session, week and semester). The richness of use features both at session level (already clustered in the previous step), but also as overall metrics (i.e. the use of unique number of items) as well as additional information about the use of the search function. Finally the efficiency of browsing is accounted for by aggregated efficiency ratio measures and the skewness and kurtosis indicators for the distribution of the frequencies of time and number of visits. These are providing metrics for the shape of the distributions of time and visits over the two semesters (i.e. we have seen the visual representation in figures 8.3 to 8.6).

Analyses were performed using both the totals/averages for the year and the split semesters-based metrics with no differences between the solutions. For the Y1 data, the Two-steps cluster analysis produced a 3-clusters solution. Centroids of most of the metrics are reported in table 8.6.

Metrics for clustering users	1	2	3
Total number of visits spent over the semester	52.66	98.65	56.55
Skewness of the frequency of visitation	0.91	1.22	1.14
Kurtosis of the frequency of visitation	0.45	1.58	1.36
Skewness of the frequency of time spent online	1.07	1.26	1.16
Kurtosis of the frequency of time spent online	1.11	1.75	1.25
Average time per week (minutes)			
Average time per session (seconds)	53.23	43.57	42.29
Average depth of session	26.63	14.02	10.47
Average Time per page	3.57	9.18	4.7
PER - Average Efficiency ratio (pages)	0.74	0.79	0.65
AER - Average Efficiency ratio (abstractions)	0.86	0.87	0.91
Number of searches	3.9	2251.96	34.3
Depth of search	2.3	7	7.2
Total Social content	19.02	691.77	3.55
Total time on social	142.88	2216.34	57.72
Average time for social per session	10.08	5.2	0.32
Average Depth of social	20.4	12.36	3.5
Priority of social	0.04	0.42	0.47
Unique abstractions (overall)	2.16	1.92	2.2
Unique tools (overall)	1.95	2.31	2.55
Unique actions (overall)	8.88	3	2.88
Unique pages (overall)	3.85	15.19	5.09
presence of session cluster			
Total session in each cluster and percentage	85	26	247
	23.70%	7.30%	69.00%

Table 8.6. Metrics used to cluster students' based on their usage patterns in year 1

For the Y2 4-clusters solution was obtained which provided a good distribution of students over the four groups. Centroids of most metrics are reported in table 8.7.

It should be noted that variances were obviously quite large (the procedure tries to maximise the differences) therefore only mean values were provided to give an idea of the features characterising each cluster in the solutions.

Metrics for clustering users	**1**	**2**	**3**	**4**
Total number of visits spent over the semester	96.96	65.58	86.77	146.5
Skewness of the frequency of visitation	0.9	1.49	1.23	1.07
Kurtosis of the frequency of visitation	0.92	2.53	2.02	1.33
Skewness of the frequency of time spent online	1.29	1.29	1.25	1.04
Kurtosis of the frequency of time spent online	2.27	2.03	1.69	0.96
Average time per week (minutes)	3.65	2.95	3.48	4.81
Average time per session (seconds)	47.1	46.61	46.46	48.64
Average depth of session	2.4	13.02	8.84	20.9
Average Time per page	4.07	5.78	6.81	5.61
PER - Average Efficiency ratio (pages)	0.86	0.71	0.81	0.7
AER - Average Efficiency ratio (abstractions)	0.85	0.88	0.87	0.84
Number of searches	9.65	113.98	14.08	427.83
Depth of search	2.4	13.02	5.36	16.5
Total Social content	18.54	234.37	709.64	957.68
Total time on social	5.57	883.59	946.75	3772.43
Average time for social per session	1	7.35	0.79	7.41
Average Depth of social	7.92	14.4	5.74	9.77
Priority of social	0.01	0.11	0.99	0.33
Unique abstractions (overall)	2.49	2.41	2.29	2.24
Unique tools (overall)	3.1	2.92	2.62	2.46
Unique actions (overall)	2.49	3.27	3.32	3.29
Unique pages (overall)	6.13	7.25	7.58	6.16
presence of session cluster				
Total session in each cluster and percentage	209	179	126	96
	34.10%	29.20%	20.60%	15.70%

Table 8.7. Metrics used to cluster students' based on their usage patterns in year 2

We decided not to conduct a separate analysis for the longitudinal sample because of the limitation of the algorithm: with twice as many parameters (for both Y1 and Y2 usage) and a limited number of participants, it doesn't make sense to force a cluster solution in which attributes are more than the participants.

A final piece of the characterization of students was the use of a questionnaire at the end of each year looking at students' opinions about e-learning. Some preliminary results were reported in Hardy et al. (2006, 2007). It would be interesting to know that most students found the resources offered useful and they liked them. However it was not possible to include the results of the surveys here because they were collected anonymously and independently from the other metrics.

8.1.8. Summary of findings

Usage of online material produced a vast amount of data which needed to be cleaned, organised and aggregated appropriately before it could be used for any analysis. Up to this point it showed how using a mix of visual and statistical approaches and the overlay of three interpretational schemas (extent, richness and efficiency of browsing) allowed a meaningful representation of the online activity. Starting from an overview of activities in year-classes we turned to the smallest unit of behaviours, the clicks. Then we explained how sequences of clicks grouped into sessions, or episodes of interaction, and described them in details adding depth to the log data with the abstraction to the different interpretational layers. Using clustering techniques, similar to the ones presented in the last chapter, we also provided a way of grouping sessions with similar features. Together with a number of other metrics, the characterisation of the sessions allowed us to describe the users' activity and, in turn, apply clustering techniques to group together students expressing similar patterns of behaviour in their online usage. As was the case for the styles metrics, the next step will be to evaluate if online usage as a window on students' behaviour and intentions, formulated in this representation is actually useful in identifying variation in AP.

8.2. Online behaviours & performance

In the last chapter we were able to identify some relations between patterns of behaviours like attendance and participation with AP. Although partly expected, the relations were weak and inconsistent. This could be partly due to the fact that the students in the cohorts considered were high achievers in prior performance, with little within group variation, and possibly could compensate with higher exams' performance to achieve the desired grades.

In certain respects, the fruition of online material could be considered in parallel to attendance as an expressed intention to participate: why would a student use the online content when they do not attend lectures or tutorials? There is an argument according to which the e-learning material, rather than complementing traditional content in a blended course could actually become a substitute to traditional delivery methods. Newer generations of students, becoming more reliant on high speed and omnipresent communication, and devices allowing them to be always *connected*, certainly welcomed the introduction of a system which could give them access anywhere, anytime. In fact, from the data about usage of the WebCT courses offered to those taking psychology, an average of 7% (declining from 15.3% in 2004 to 1.2%) of students over the 5 years span used the system less than once

every two weeks and about 18% (decreasing from 38.5% to 6.5) used it less than once a week.
Slightly different is the pattern of access of additional material (i.e. the publishers' e-packs) which is largely depending on the way in which the material is offered. Data showed that over the 4 years period considered between 35% and 45% used the additional tests and activities in the external offering.
Whilst this demonstrates overwhelmingly that students *do* use the e-learning content often, the question is how useful this is and how does this usage relate to their studying and academic performance.

8.2.1. A broad, but simplistic view

One way of assessing whether e-learning is effective is to look at the simple correlations between usage and grades. It has already been shown in Vigentini (2008) how there seemed to be a reasonable correlation between the amount of time spent online and the extent of usage with the grades achieved, particularly for the multiple choice section of the exam in Y1. Here we look into more details of such relations after we have provided a more in-depth characterization of online usage.

The extent of access, both in terms of frequency of access and time spent does correlate with the grades obtained across the board (see tables 8.8 and 8.9) in both year 1 and year 2. The size of the correlations is not very big, but given the size of the sample is a reasonable result. A couple of interesting observations emerge from the negative correlations between the average time spent per week and the various pieces of coursework and the average depth of visits. In other words more use (time spent) in deeper sessions does not lead to better performance.
Vigentini (2008) argued that greater use is not necessarily better. This statement can be further supported by showing how students access the system counts. In particular it is possible to infer that students spending more time in the system might just be struggling to find what they need, especially in the year 1. In fact, if we look at the efficiency of visitations between the semesters and years for the longitudinal sample, it is evident that students become more efficient and the difference between data points tested with a repeated measures ANOVA on the efficiency ratio for abstractions produced a significant difference for both year ($F_{(1,551)}$=63.5, p<.001) and semesters ($F_{(1,551)}$=108.3, p<.001) as well as an interaction

between the two factors ($F_{(1,550)}$=58.6, p<.001). The test on the efficiency ratio for pages provided similar results.

In the year two table (8.9) it seems that regularity and efficiency are the parameters which correlate significantly with the performance in the various forms of assessment.
This is not surprising given the intensity of the workload in this course and those who are able to organise themselves effectively are the ones who succeed and obtain higher grades.

A repeated measures ANOVA was performed on the metrics for each semester, particularly on the total visits and on the average visits per week (parameters of extent and regularity) which produced significant differences. For the total number of visits the difference between semester 1 and semester 2 is up to a third more in year 1 and 10% in year 2. For what concerns the average visits per week the difference is from 2.7 times per week in year 1 semester 1 to 3.6 in year 2 semester 1 (in year 2 semester 2 the traffic shifts to the wiki supporting the projects and this activity is not included here).

The metrics indicating the extent of use of social resources were excluded from these summaries (table 8.8 and 8.9) for both year 1 and year 2 because they were either not significantly correlated with aspects of assessment or, where the relation was significant the size of the correlation was less than .1.

8.2.2. The importance of 'how'

An alternative approach is to focus on the actual patterns of usage. We have already produced clusters for particular types of students, like we have done for types of students according to styles, therefore we can look at possible differences in performance.
In the analysis of the first half of this chapter, what type of material and how material is used have been at the core of what makes a particular type of session and how a student behaves.

To minimize the noise in the data the dataset was split into three samples according to the year-class. The effect that the clusters emerging from usage are producing (if any) in AP was then assessed.

		Essay1	Essay2	Essay3	Coursework	Exam 1 (essay quest.)	Exam 1 (MCQs)	Exam 1 (essay quest.)	Exam 2 (MCQs)	Psychology 1 final grade
Total Visits	Pearson Correlation	.169**	.229**	.240**	.229**	.318**	.209**	.276**	.288**	.147**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	1340	1340	1340	1332	1346	1346	1346	1338	1286
Average number of visits per week	Pearson Correlation	.172**	.232**	.224**	.232**	.302**	.196**	.257**	.277**	.134**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	1340	1340	1340	1332	1346	1346	1346	1338	1286
Average time spent per week	Pearson Correlation	-.071**	-.091**	-.062*		-.243**	-.068*	-.117**	-.070*	
	Sig. (2-tailed)	.009	.001	.022		.000	.012	.000	.010	
	N	1340	1340	1340		1346	1346	1346	1338	
Average time per session	Pearson Correlation					-.195**		-.078**		
	Sig. (2-tailed)					.000		.004		
	N					1346		1346		
Average depth of visit	Pearson Correlation					-.534**	.166**	-.251**	-.295**	
	Sig. (2-tailed)					.000	.000	.000	.000	
	N					1346	1346	1346	1338	
Average time per page	Pearson Correlation	-.056*			.072**	.147**	-.163**	.154**	.334**	
	Sig. (2-tailed)	.039			.008	.000	.000	.000	.000	
	N	1340			1332	1346	1346	1346	1338	
Average depth of search	Pearson Correlation	-.058*	-.066*	-.071**	-.140**		-.131**		-.078**	-.074**
	Sig. (2-tailed)	.034	.015	.009	.000		.000		.004	.008
	N	1340	1340	1340	1332		1346		1338	1286
Efficiency ratio (pages)	Pearson Correlation				.199**	.136**		.073**	.331**	
	Sig. (2-tailed)				.000	.000		.007	.000	
	N				1332	1346		1346	1338	
Efficiency ratio (abstractions)	Pearson Correlation	.101**	.078**	.094**	.084**	.164**		.188**	.221**	.068*
	Sig. (2-tailed)	.000	.004	.001	.002	.000		.000	.000	.015
	N	1340	1340	1340	1332	1346		1346	1338	1286
Unique Abstractions	Pearson Correlation								-.061*	
	Sig. (2-tailed)								.025	
	N								1338	
Unique tools	Pearson Correlation		-.059*	-.086**	-.095**	.107**	-.090**			
	Sig. (2-tailed)		.032	.002	.001	.000	.001			
	N		1340	1340	1332	1346	1346			
Unique actions	Pearson Correlation					-.251**	.065*	-.200**	-.191**	.097**
	Sig. (2-tailed)					.000	.018	.000	.000	.001
	N					1346	1346	1346	1338	1286
Unique pages	Pearson Correlation					.127**		.069*	.069*	-.059*
	Sig. (2-tailed)					.000		.011	.012	.035
	N					1346		1346	1338	1286

Table 8.8. Significant correlations between AP in year 1 and the corresponding usage metrics.

		Report 1	Report 2	Report 3	Report 4	Coursework	Exam 1	Exam 2	Statistics Exam avg.	Psychology 2 final grade
Total Visits	Pearson Correlation	.174**	.170**	.136**	.168**	.205**	.222**	.201**	.290**	.116**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.002
	N	728	724	726	723	742	742	742	742	743
Average number of visits per week	Pearson Correlation	.149**			.138**	.174**	.234**			
	Sig. (2-tailed)	.000	.000	.005	.000	.000	.000	.000	.000	.021
	N	728	724	726	723	742	742	742	742	743
Average time spent per week	Pearson Correlation							-.086*		
	Sig. (2-tailed)							.019	.003	.038
	N							742	742	743
Average time per session	Pearson Correlation									
	Sig. (2-tailed)			.016						.022
	N			726						743
Average depth of visit	Pearson Correlation									
	Sig. (2-tailed)								.014	
	N								742	
Average time per page	Pearson Correlation									
	Sig. (2-tailed)								.001	
	N								742	
Average depth of search	Pearson Correlation							.143**		
	Sig. (2-tailed)						.000	.000	.000	
	N						742	742	742	
Efficiency ratio (pages)	Pearson Correlation								.184**	
	Sig. (2-tailed)	.009			.043		.007		.000	
	N	728			723		742		742	
Efficiency ratio (abstractions)	Pearson Correlation			.158**	.208**	.222**	.151**			
	Sig. (2-tailed)	.046	.000	.000	.000	.000	.000	.000	.002	
	N	728	724	726	723	742	742	742	742	
Unique Abstractions	Pearson Correlation									
	Sig. (2-tailed)									
	N									
Unique tools	Pearson Correlation									
	Sig. (2-tailed)				.035	.012	.039	.007		
	N				723	742	742	742		
Unique actions	Pearson Correlation									
	Sig. (2-tailed)					.026				
	N					742				
Unique pages	Pearson Correlation							-.102**		
	Sig. (2-tailed)							.005		
	N							742		

Table 8.9. Significant correlations between AP in year 2 and the corresponding usage metrics.

The year 1 sample

The three clusters emerging from the procedure were used to test whether usage patterns could account for some of the variance in the grades obtained in the psychology 1 course. Table 8.10 presents the summary of the means and significant differences and relative post-hoc t-tests (only the significance values are reported here). Significant differences between the grades obtained by students in virtually all clusters were found, this is a very useful result demonstrating the effectiveness of the clustering in accounting for variability in AP.

	mean C1	sd C1	mean C2	sd C2	mean C3	sd C3
Essay 1	55.78	10.76	53.24	13.86	55.66	12.45
Coursework avg.	60.35	11.35	56.30	7.26	50.14	12.19
Exam1 (essay)	0.00	0.00	60.04	7.65	57.51	12.62
Exam1 (MCQs)	60.03	15.17	46.94	13.26	40.62	11.52
Exam2 (essay)	21.40	28.04	55.92	15.18	58.02	14.43
Exam2 (MCQs)	20.33	27.56	51.45	11.19	41.02	13.54
Final Grade	61.81	8.67	55.12	6.64	56.86	7.23
Attendance	8.88	1.31	6.88	1.98	6.85	2.35

	df	F	Sig.	post-hoc T C1-C2	post-hoc T C1-C3	post-hoc T C1-C4	post-hoc T C2-C3	post-hoc T C2-C4	post-hoc T C3-C4
Coursework avg.	3, 345	24.06	.000	*ns*	**.000**	**.000**	*ns*	**.002**	**.000**
Exam1 (essay)	3, 347	898.10	.000	**.000**	**.000**	**.000**	*ns*	*ns*	*ns*
Exam1 (MCQs)	3, 347	72.75	.000	**.000**	**.000**	**.000**	*ns*	**.000**	**.000**
Exam2 (essay)	3, 345	120.81	.000	**.000**	**.000**	**.000**	*ns*	*ns*	**.013**
Exam2 (MCQs)	3, 345	50.29	.000	**.000**	**.000**	**.000**	*ns*	**.000**	**.000**
Final Grade	3, 336	14.67	.000	**.000**	**.001**	**.000**	*ns*	*ns*	**.000**
Attendance	3, 345	28.15	.000	**.001**	**.000**	**.000**	*ns*	*ns*	*ns*

Table 8.10. Descriptive statistics and one-way ANOVA using the 3 users' clusters in Year 1. Post-hoc t-tests of the various academic indicators for each cluster are also reported.

The year 2 sample

For the second year sample 4 groups emerged from the analysis. Similar to the patterns discovered in year 1, the clusters were reasonably good in accounting for the variation in grades. Table 8.12 presents the relevant statistics with significant differences in the one-way ANOVA.

The sample remaining is much smaller than the sample available for the first year, but the clusters created are sensitive enough. Given the small differences between cluster 2 and 3, it could be argued that a 3-clusters model might be more suitable (and matching the model obtained in Y1), however there seems to be a great difference in the online usage between the two clusters making it difficult to justify a merge (i.e. avg. visits in C2 is 78, but it is 102 in

C3, the distribution of access is also very different with C2 very skewed toward the end of term and C3 more evenly distributed).

Parameter	mean C1	mean C2	mean C3	mean C4	sd C1	sd C2	sd C3	sd C4
Y2 Report 1	50.12	46.78	50.21	50.54	16.85	14.85	9.45	13.65
Y2 Report 2	54.81	50.60	54.24	54.25	16.78	15.86	9.34	12.89
Y2 Report 3	53.14	48.56	56.50	53.71	19.12	19.09	9.87	16.10
Y2 Report 4	58.31	53.51	59.83	58.85	20.04	16.85	9.70	11.77
Coursework avg.	55.38	50.48	51.54	56.01	16.96	16.23	16.44	11.22
Exam1 (essay)	55.24	57.30	61.03	59.95	12.02	15.54	18.15	8.69
Exam2 (essay)	51.33	55.31	50.07	55.22	18.46	15.21	16.60	11.40
Statistics	60.43	60.12	62.55	63.56	16.61	20.30	20.03	15.63
Final Grade	57.96	57.53	58.06	58.06	8.69	8.18	7.68	8.38
Attendance	83.33	87.58	89.75	89.43	.	17.19	14.25	15.86

	df	F	Sig.	post-hoc T C1-C2	post-hoc T C1-C3	post-hoc T C1-C4	post-hoc T C2-C3	post-hoc T C2-C4	post-hoc T C3-C4
Y2 Report 1	3, 524	2.144	ns	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
Y2 Report 4	3, 522	4.291	.005	**.044**	*ns*	*ns*	***.011***	*ns*	*ns*
Coursework avg.	3, 534	4.032	.007	**.031**	*ns*	*ns*	*ns*	**.049**	*ns*
Exam1 (essay)	3, 534	4.635	.003	*ns*	**.004**	**.066**	*ns*	*ns*	*ns*
Exam2 (essay)	3, 534	3.577	.014	*ns*	*ns*	*ns*	***.045***	*ns*	*ns*
Statistics	3, 534	0.983	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
Final Grade	3, 538	0.126	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
Attendance	3, 246	0.396	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*

Table 8.11. Descriptive statistics and one-way ANOVA using the 3 users' clusters in Year 2. Post-hoc t-tests of the various academic indicators for each cluster are also reported.

8.2.4. Summary of findings

This section showed two key results: we found significant correlations between AP and the patterns of usage and we showed that the clusters obtained from usage patterns lead to significant differences in performance. By looking at the patterns of the relations between types of assessment and different ways of using e-learning, we were able to deduce that a regular and efficient way of accessing the online content could improve significantly AP. It is yet to be established how types of usage relate to types of stylistic preferences, which will be explored in the next section

8.3. Stylistic differences in usage

8.3.1. Styles and behaviours

In the same ways that stylistic differences were observed in relation to AP, this section examines how styles relate to behaviours.

A number of correlations were identified between measures of styles and the patterns of usage. Tables 8.13, 8.14 and 8.15 summarise the significant relations. As in the previous chapter, the TSI provided fairly weak and inconsistent results. The VICS-WA and CSI highlighted only some relations which seem quite intuitive in the direction.

The ASSIST, however, as it was the case for the relations with AP, offered a range of significant relations with online usage, particularly in the subscales of the surface and strategic approach, demonstrating the relevance of the instrument in the academic domain.

		Total visits (Year 1)	Average visits per week (Year 1)	Efficiency ratio (abstractions)	Total visits (Year 2)	Average depth of visits (Year 2)
V/I Ratio	Pearson Correlation	.131*	.146*	-.138*	.020	-.163*
	Sig. (2-tailed)	.041	.022	.031	.783	.022
	N	246	246	246	197	197
W/A Ratio	Pearson Correlation	.176**	.127*	-.049	.079	-.104
	Sig. (2-tailed)	.006	.046	.448	.269	.146
	N	246	246	246	197	197
CSI (Analysis)	Pearson Correlation	.250**	.193**	.069	.085	-.015
	Sig. (2-tailed)	.000	.000	.208	.335	.861
	N	333	333	333	132	132
CSI (Intuition)	Pearson Correlation	-.228**	-.188**	-.053	-.227**	.165
	Sig. (2-tailed)	.000	.001	.330	.009	.059
	N	333	333	333	132	132

Table 8.12. Significant correlation between the usage metrics and the CSI and VICS-WA. In red p significant at .01 level and in green at the .05 level.

Year 1

		Total Visits	Average number of visits per week	Average time spent per week	Average time per session	Average depth of visit	Average time per page	Average depth of search	Efficiency ratio (pages)	Efficiency ratio (abstractions)	Unique Abstractions	Unique tools	Unique actions	Unique pages
DEEP Approach														
Seeking meaning	Pearson Correlation			.102*	.113*					-.102*				
	Sig. (2-tailed)			.030	.017					.030				
	N			449	449					449				
Relating ideas	Pearson Correlation		-.096*	.098*	.106*					-.081			.110*	
	Sig. (2-tailed)		.041	.039	.025					.086			.019	
	N		449	449	449					449			449	
Use of evidence	Pearson Correlation									-.093*				
	Sig. (2-tailed)									.049				
	N									449				
Interest in ideas	Pearson Correlation			.167**	.158**				-.115*					
	Sig. (2-tailed)			.000	.001				.015					
	N			449	449				449					
STRATEGIC Approach														
Organised studying	Pearson Correlation	.175**	.137**						-.099*					
	Sig. (2-tailed)	.000	.004						.037					
	N	449	449						449					
Time management	Pearson Correlation	.151**	.123**											
	Sig. (2-tailed)	.001	.009											
	N	449	449											
Alertness to assessment demands	Pearson Correlation	.095*	.093*											
	Sig. (2-tailed)	.044	.048											
	N	449	449											
Achieving	Pearson Correlation	.195**	.141**				.098*				.106*			
	Sig. (2-tailed)	.000	.003				.039				.025			
	N	449	449				449				449			
Monitoring effectivness	Pearson Correlation	.108*												
	Sig. (2-tailed)	.022												
	N	449												
SURFACE/APATHETIC Approach														
Lack of purpose	Pearson Correlation					.121*	-.141**			-.106*				
	Sig. (2-tailed)					.010	.003			.025				
	N					449	449			449				
Unrelated memorising	Pearson Correlation	.109*												
	Sig. (2-tailed)	.021												
	N	449												
Syllabus boundness	Pearson Correlation			-.121*	-.136**								-.094*	
	Sig. (2-tailed)			.010	.004								.047	
	N			449	449								449	
Fear of failure	Pearson Correlation	.121*		.114*	.095*							-.098*	-.094*	
	Sig. (2-tailed)	.010		.016	.044							.039	.046	
	N	449		449	449							449	449	
Preferences for learning environments														
Transmitting information (SURFACE)	Pearson Correlation		.094*						.168**					
	Sig. (2-tailed)		.047						.000					
	N		445						445					
Supporting understanding(DEEP)	Pearson Correlation												.130**	
	Sig. (2-tailed)												.006	
	N												445	

Year 2

		Total Visits	Average number of visits per week	Average time spent per week	Average time per session	Average depth of visit	Average time per page	Average depth of search	Efficiency ratio (pages)	Efficiency ratio (abstractions)	Unique Abstractions	Unique tools	Unique actions	Unique pages
DEEP Approach														
Seeking meaning	Pearson Correlation			.161**	.167**		.156**							
	Sig. (2-tailed)			.004	.003		.006							
	N			310	310		310							
Relating ideas	Pearson Correlation			.125*	.194**					-.121*			.128*	
	Sig. (2-tailed)			.028	.001					.033			.024	
	N			310	310					310			310	
Use of evidence	Pearson Correlation			.129*	.182**									
	Sig. (2-tailed)			.023	.001									
	N			310	310									
Interest in ideas	Pearson Correlation	.125*			.165**									
	Sig. (2-tailed)	.027			.004									
	N	310			310									
STRATEGIC Approach														
Organised studying	Pearson Correlation	.237**	.193**											
	Sig. (2-tailed)	.000	.001											
	N	310	310											
Time management	Pearson Correlation	.235**	.201**											
	Sig. (2-tailed)	.000	.000											
	N	310	310											
Alertness to assessment demands	Pearson Correlation	.119*	.128*											
	Sig. (2-tailed)	.036	.024											
	N	310	310											
Achieving	Pearson Correlation	.276**	.230**											
	Sig. (2-tailed)	.000	.000											
	N	310	310											
Monitoring effectivness	Pearson Correlation	.100	.079		.115*								-.128*	
	Sig. (2-tailed)				.042								.025	
	N				310								310	
SURFACE/APATHETIC Approach														
Lack of purpose	Pearson Correlation	-.179**	-.138*			.129*	-.122*							
	Sig. (2-tailed)	.002	.015			.023	.032							
	N	310	310			310	310							
Unrelated memorising	Pearson Correlation				-.122*							-.115*		
	Sig. (2-tailed)				.032							.044		
	N				310							310		
Syllabus boundness	Pearson Correlation	-.154**												
	Sig. (2-tailed)	.007												
	N	310												
Fear of failure	Pearson Correlation								.119*			-.126*		
	Sig. (2-tailed)								.036			.026		
	N								310			310		
Preferences for learning environments														
Transmitting information (SURFACE)	Pearson Correlation													
	Sig. (2-tailed)													
	N													
Supporting understanding(DEEP)	Pearson Correlation				.140*									
	Sig. (2-tailed)				.014									
	N				308									

Table 8.13. Correlations between the ASSIST and usage metrics. Only significant correlations are reported in the two years.

			Year 1								Year 2								
			Total Visits	Average number of visits per week	Average time spent per week	Average depth of visit	Efficiency ratio (pages)	Efficiency ratio (abstractions)	Unique Abstractions	Unique tools	Total Visits	Average number of visits per week	Average time spent per week	Average depth of visit	Efficiency ratio (pages)	Efficiency ratio (abstractions)	Unique Abstractions	Unique tools	
Functions	LEGL	Pearson Correlation						.163**						-.171*	.176*				
		Sig. (2-tailed)						.000						.013	.010				
		N						459						211	211				
	EXEC	Pearson Correlation							.116*	.103*									
		Sig. (2-tailed)							.013	.027									
		N							459	459									
	JUDI	Pearson Correlation	-.096*	-.106*			-.129**											.143*	
		Sig. (2-tailed)	.040	.023			.006											.038	
		N	459	459			459											211	
Levels	GLOB	Pearson Correlation					-.138**												
		Sig. (2-tailed)					.003												
		N					459												
	LOCL	Pearson Correlation																	
		Sig. (2-tailed)																	
		N																	
Leanings	PROG	Pearson Correlation					-.141**												
		Sig. (2-tailed)					.002												
		N					459												
	CONS	Pearson Correlation					-.106*												
		Sig. (2-tailed)					.023												
		N					459												
Forms	HIER	Pearson Correlation							.096*								.153*	.174*	.180**
		Sig. (2-tailed)							.041								.027	.011	.009
		N							459								211	211	211
	MONA	Pearson Correlation	-.099*				-.099*												
		Sig. (2-tailed)	.033				.033												
		N	459				459												
	OLIG	Pearson Correlation	.108*	.095*							.179**	.171*	.161*						
		Sig. (2-tailed)	.021	.042							.009	.013	.019						
		N	459	459							211	211	211						
	ANAR	Pearson Correlation					-.100*							.145*	-.148*	-.179**			
		Sig. (2-tailed)					.032							.035	.032	.009			
		N					459							211	211	211			
	DEMO	Pearson Correlation			-.095*	-.159**	.123**	.121**			-.153*	-.166*							
		Sig. (2-tailed)			.042	.001	.008	.010			.026	.016							
		N			459	459	459	459			211	211							
Scopes	INTR	Pearson Correlation						.110*				.150*			.155*				
		Sig. (2-tailed)						.018				.029			.024				
		N						459				211			211				
	EXTR	Pearson Correlation					-.101*												
		Sig. (2-tailed)					.031												
		N					459												

Table 8.14. The significant correlations between the TSI and the usage metrics in year 1 and year 2.

8.3.2. Emergent styles

After establishing that there are some interesting correlations between usage of the online material and styles, put the users' clusters to the test, in order to identify potential differences in the metrics of styles.

For the year 1 sample, the usage clusters produced only a couple of significant differences (see table 8.16). In year 2 some of the subscales of the ASSIST and the Deep approach were found to be significantly different using the users' clusters. This was quite unexpected.

		df	F	p
Year 1	**ASSIST - Lack of Purpose**	1, 182	11.32	0.001
	VICS-WA - W/A Ratio	1, 124	12.3	0.001
Year 2	**ASSIST - Relating ideas**	3, 291	3.38	0.019
	ASSIST - Lack of Purpose	3, 291	2.61	0.05
	ASSIST - Deep Approach		2.72	0.045
	TSI - Democratic Scale	3, 182	4.465	0.005

Table 8.15. One-way ANOVA using the 3 users clusters on the metrics of styles.

At the opposite starting point, when looking at the styles clusters to test whether there might be usage differences between types of people as defined by styles, the picture is more varied showing that different instruments produced a number of significant differences in the expressed behaviours (table 8.16).

In the year 1 sample, the ASSIST subscales, as was the case for AP, seem to be quite sensitive to the differences across the three interpretational domains and differences were found in extent, richness and efficiency. The TSI, quite weak in detecting differences in AP, found some differences (some weaker options could be accepted at the 10% level of significance) in the extent (both time spent and frequency), as well as richness (avg. depth of sessions) and efficiency (only abstractions). The CSI clustering was able to detect differences in time spent and number of visits. Finally the VICS also detected differences in the time spent and number of visits. In the year 2 the patterns are less clear-cut, but it is possible that this is a side effect of the way in which constructed the procedure to cluster sessions.

		ASSIST			TSI			VICS-WA			CSI		
		df	F	sig.	df	F	sig.	df	F	sig.	df	F	sig.
Year 1	Total number of visits spent over the semesters	5, 439	4.573	.000	3, 455	4.027	.008	3, 242	4.818	.003	3, 329	8.695	.000
	Average number of visits per week	5, 439	3.695	.003	3, 455	2.562	.054	3, 242	3.652	.013	3, 329	6.861	.000
	Average time spent per semester	5, 439	5.815	.000	3, 455	3.213	.023	3, 242	2.045	*ns*	3, 329	1.568	*ns*
	Avearage time per session	5, 439	4.690	.000	3, 455	2.354	.071	3, 242	2.142	.096	3, 329	1.362	*ns*
	Average Depth per session	5, 439	2.059	.070	3, 455	2.811	.039	3, 242	.263	*ns*	3, 329	4.803	.003
	Average time per page	5, 439	6.298	.000	3, 455	2.580	.053	3, 242	.457	*ns*	3, 329	2.037	*ns*
	Number of searches	5, 439	1.130	*ns*	3, 455	.845	*ns*	3, 242	.280	*ns*	3, 329	2.752	.043
	Page Efficiency ratio	5, 439	7.331	.000	3, 455	1.058	*ns*	3, 242	2.393	.069	3, 329	.350	*ns*
	Abstraction Efficiency ratio	5, 439	3.122	.009	3, 455	3.247	.022	3, 242	2.083	*ns*	3, 329	1.168	*ns*
	Total Social items over the semesters	5, 439	1.359	*ns*	3, 455	1.665	*ns*	3, 242	.575	*ns*	3, 329	3.750	.011
	Average Depth of use of social resources	5, 410	.495	*ns*	3, 428	2.625	.050	3, 229	.102	*ns*	3, 306	2.434	.065
	Priority ratio of social resources	5, 438	13.250	.000	3, 452	.568	*ns*	3, 238	.978	*ns*	3, 327	.707	*ns*
	Proportion of social resources per session	5, 438	4.034	.001	3, 452	.261	*ns*	3, 238	.491	*ns*	3, 327	1.452	*ns*
	Total time on social resources	5, 439	2.646	.023	3, 455	3.109	.026	3, 242	3.226	.023	3, 329	3.543	.015
	Average time on social resources per session	5, 410	4.768	.000	3, 430	2.022	*ns*	3, 232	4.395	.005	3, 308	1.180	*ns*
	Proportion of session time on social resources	5, 439	1.615	*ns*	3, 455	3.200	.023	3, 242	.818	*ns*	3, 329	1.710	*ns*
Year 2	Total number of visits spent over the semesters	4, 207	1.415	*ns*	3, 207	1.659	*ns*	3, 193	1.848	*ns*	3, 128	2.320	.078
	Average number of visits per week	4, 207	1.523	*ns*	3, 207	1.496	*ns*	3, 193	1.529	*ns*	3, 128	1.749	*ns*
	Average time spent per semester	4, 207	.853	*ns*	3, 207	2.041	*ns*	3, 193	.522	*ns*	3, 128	1.055	*ns*
	Avearage time per session	4, 207	1.161	*ns*	3, 207	1.327	*ns*	3, 193	.369	*ns*	3, 128	.633	*ns*
	Average Depth per session	4, 207	1.133	*ns*	3, 207	2.631	.051	3, 193	.299	*ns*	3, 128	1.664	*ns*
	Average time per page	4, 207	1.995	*ns*	3, 207	3.940	.009	3, 193	.669	*ns*	3, 128	.331	*ns*
	Number of searches	4, 207	1.499	*ns*	3, 207	1.335	*ns*	3, 193	.568	*ns*	3, 128	1.144	*ns*
	Page Efficiency ratio	4, 207	3.317	.012	3, 207	1.851	*ns*	3, 193	.126	*ns*	3, 128	1.954	*ns*
	Abstraction Efficiency ratio	4, 207	1.069	*ns*	3, 207	.959	*ns*	3, 193	.447	*ns*	3, 128	1.191	*ns*
	Total Social items over the semesters	4, 207	.658	*ns*	3, 207	.837	*ns*	3, 193	.187	*ns*	3, 128	.391	*ns*
	Average Depth of use of social resources	4, 189	.671	*ns*	3, 196	1.637	*ns*	3, 183	.814	*ns*	3, 124	2.006	*ns*
	Priority ratio of social resources	4, 207	.466	*ns*	3, 207	1.460	*ns*	3, 193	1.460	*ns*	3, 128	2.720	.047
	Proportion of social resources per session	4, 207	1.120	*ns*	3, 207	1.015	*ns*	3, 193	.500	*ns*	3, 128	1.613	*ns*
	Total time on social resources	4, 207	1.039	*ns*	3, 207	1.145	*ns*	3, 193	2.332	.076	3, 128	1.278	*ns*
	Average time on social resources per session	4, 193	.615	*ns*	3, 199	1.107	*ns*	3, 185	3.301	.022	3, 125	.217	*ns*
	Proportion of session time on social resources	4, 207	.163	*ns*	3, 207	2.634	.051	3, 193	1.463	*ns*	3, 128	1.163	*ns*

Table 8.16. One-way ANOVA statistics on the usage metrics using the clusters emerged from each separate instrument measuring styles.

Although we were able to reveal correlations between styles and usage, the results presented in this section seem to demonstrate that whilst the clustering of students based on usage is not very good at discriminate AP or styles, the classification of students using styles stratifies the sample in such a way that the variation in the patterns of usage can be detected, with the ASSIST as the most powerful instrument.

What was disappointing from this analysis was the lack of relations between the behavioural patterns (i.e. types of sessions or users' preferred styles of browsing) with the more cognitive subscales of the various instruments, particularly the VICS-WA.

8.3.3. Summary of findings

In this section the characterization of students, using usage patterns was related to styles with some interesting correlations between the two. In particular the ASSIST showed a number of significant correlations between the subscales and usage.
Then, we turned the question around and examined whether emergent usage patterns were able to account for the variation in styles. This was shown not to be the case, as student clusters by usage are not sensitive enough to discriminate levels of styles. Clusters based on styles, however were much better in differentiating patterns of activity with all instruments performing reasonably well in year 1 and less well in year 2. Sample size and the way in which clusters were generated might be one of the many factors affecting the analysis.

8.4. Overall summary of findings

This chapter was rich in valuable results which allowed an in-depth analysis of online usage, its effects on AP and its relations with cognitive and learning styles.

With a systematic approached to usage it was possible to move from the sequences of clicks to the interpretations of behaviours and intentions banking on three interpretational schemas and clustering techniques to precisely define activity (episodes) and users' patterns according to extent, richness and efficiency of online usage. The different year-classes produced slightly different solutions (3 or 4 clusters), but the analysis of the variation of AP based on the patterns of usage lead to similar significant differences in AP for a number of assessment points (i.e. both coursework and exams).

By correlating the metrics of usage with AP it was also determined that there are a number of useful patterns, especially with regards to the frequency, regularity and efficiency of use.
Next, the relations between patterns of use and measures of styles were examined: results were not dissimilar from those hinted in the last chapter by the behavioural expressions of participation (i.e. attendance and punctuality).
In this analysis, parallel to the one in the last chapter, the ASSIST emerged as a very valuable tool, correlating well with the usage metrics and able to identify variation in usage patterns which the other instruments are not able to detect. On the other hand, clusters of students based on usage were not good in accounting for the variability in styles. This could suggest a one-way relation between styles and usage, which will be considered in the next chapter

From observations drawn in the last two chapters, it is evident that: 1) emergent types of usage do provide invaluable information about potential performance in the courses taken into consideration; 2) measures of styles like the ASSIST, provides a powerful tool to *diagnose* potentially problematic students, and 3) behaviour online could complement the measures of styles into a profile of the students which could provide empirical evidence and a strong basis to design pedagogical interventions.

Chapter 9. A possible approach to evaluating e-learning: a model for online learning

The first part of this thesis identified the context in which learning technology and e-learning was to be analysed. Then, a case was made for studying individual differences and provided a detailed characterisation of students both from a psychological perspective and the variation of online usage of material supporting learning and studying in psychology courses. As indicated in chapters 1 and 2, this provided a wealth of empirical data in support of the evaluation of e-learning as well as a set of descriptive, rather than prescriptive pointers, in order to improve the implementation of learning technology (tools and methods) and the interactive process with which students engage as part of their daily activities.

Chapters 3 and 4 considered what aspects of differentiation could be most useful to inform practitioners and technologists and made a connection between academic performance as an objective measure of achievement and individual differences. We considered in turn a number of metrics. In the literature, prior performance (to university) was found to be a fairly good predictor of university grades, and it was suggested that given the consistent and high correlations between IQ test scores and academic achievement, this could be used also as a proxy of ability in general.

Personality differences were also shown to have a certain effect on the variability of AP, However the interaction between abilities and personality portrays students' profiles which are overcomplicated for practical applications. Furthermore, it is possible that both ability and personality are too deterministic in nature for the purposes of e-learning implementations, especially when taking into account the interactive nature of the learning process and studying at university.

For example, if someone is very introverted and s/he is immersed in an educational context in which it is expected to work and produce in small groups, depending on the type of assessment, even if prior performance indicated a great potential for this student, s/he might not perform as well as expected. In this case neither excellent prior performance, nor personality could predict with confidence how the student would react to the assessment requirements and the demands of the specific context in a university course.

For this reason we turned our attention to *softer* metrics to assess individual differences. Measure of styles, specifically because of the controversy behind their theoretical formulation offered a space between the proximal and distal assessment of *types* and the variation became an appealing option. Another argument in support of styles is that because they are measuring preferences, unlike traditional IQ and traits measures, they might be malleable, offering the possibility to design specific interventions in case particular combinations proved to be detrimental for learning and studying.
However, because of the potential weaknesses of the instruments elected, it was specifically decided to use a number of different instruments to assess styles, leading to a more comprehensive overview of the relations between the tools and their theoretical assumptions.

9.1. A review of the findings

In the last three chapters we presented a wide set of results based on academic performance, styles and usage. Chapter 6 looked at the effect of prior performance on the university grades. Chapter 7 considered measures of styles and their relations with AP and chapter 8 looked at the online usage in relations to both AP and styles.

The most important finding was a number of meaningful patterns of relations between the metrics from the three domains (AP, styles and usage) in the database studied. To provide a logical structure to the discussion, these patterns are explained in more detail in the following sections.

9.1.1. Academic performance and prior performance

In chapter 6 we determined that even though there are high entry level requirements for entrance to our psychology courses, therefore narrowing the range of entry grades to high achieving students, we found a certain variation in AP at university level.

In the first hypothesis (H1, p. 166) it was postulated that there would be a significant positive relation between previous performance and successive university grades. A detailed analysis of the impact of prior performance was found to be a good predictor for university grades, particularly in the pre-honours courses and overall, accounting for up to 13% of the variance. Paradoxically, prior performance in our sample was too narrow to be a good predictor for low AP at university. Most students admitted for these particular courses have a history of high achievement, therefore one would expect them to perform at university as well. This however, didn't seem the case: excluding those students who had recorded specific issues in their private domain, some students still seem to underachieve.

Re-examining the H1 and questioning whether the number of As, the types of As and the overall progress have an impact on university performance, we found that the number of As at high school level alone can account for up to about 9% of the variance in university performance. Although excellent performance in English and Maths are the stronger predictors, when comparing the sciences and social sciences subjects, at least for the psychology students, the trends were very similar.

The second interesting result was that even if the cohort considered presented evidence of high prior performance, there seem to be a trend according to which students with higher number of As tend to improve very little over time whilst those with Bs and Cs could improve considerably.

This is reasonable as 'A' students might have reached their maximal performance before university level whilst 'B' and 'C' students could -and one would hope- that they take advantage of university to develop their potential.

The regression analysis was somewhat limited in its ability to determine what type of grades might be more or less likely to affect performance at university. Therefore we used data mining (Two-steps clustering) to identify meaningful patterns. Three main groups could be identified: two groups with an average of 3 As and 2 Bs separated by the fact that one of the two groups had taken psychology before, and a third group accounting for about 25% of the sample considered of which marks were either not recorded or less than the other two groups.

This finding is largely in line with expectations and adds further evidence to the fact that prior performance is a reasonably good indicator of academic achievement.

9.1.2. Overview of the measures of styles

With the availability of a large sample, and the longitudinal sub-samples, as well as offering more data in support of the validity of each style instrument, we were able to present a unique insight in the relations between different measures of styles.

The reliability of the ASSIST was the most striking feature from all subsamples. We identified some differences in the scores of the subscales of the ASSIST in the comparison of genders and country of origin. There were also small differences in the longitudinal sample (particularly between the strategic and surface approach) between year 1 and year 2. However, there was no difference in students taking different degree paths, suggesting that the effect of the specific course might be more pronounced than the fact that the course was taken by students as a core or optional module in their curriculum.
Whilst it was expected that some changes in approaches to learning in year 1 and 2 are determined by the courses which differ in workload and assessment demands, it is unclear why differences were found between genders and country of origin. There is no doubt that females perform better than males at high school (see chapter 5), a pattern which remains valid in our courses (chapter 6). It is also interesting to point out that there could be culturally-related differences between UK, EU and other students who attend university. The data should make us reflect on the impact of an increasingly multicultural student base and the necessity of adapting course design to cater for these differences.

The Thinking styles inventory, based on the Mental self-government theory, was less reliable than reported in previous research. A cause of concern was the inconsistent replication of the results across classes which were not expected to differ greatly. Quite interestingly, however, table 4.6 summarised the relations between thinking styles and methods of instruction, and the summary of the differences between year 1 and year 2 (table 7.13) makes it possible to argue that the nature of the courses, in a similar way to the differences reported for the ASSIST, might *cause* the variations in scores. In particular changes on the judicial scale (related to thought-based questions and discussion of ideas in small group activities) and external scale (cooperative learning and small group activities) could be indicators that intellectual styles are much more reliant on the context than was advocated by Sternberg and colleagues.

We also provided further evidence in support of a two-dimensional factor structure for the CSI. This supports Hodgkinson and Sadler-Smith in their interpretation of intuition and analysis as two separate dimensions. The correlation between the two, however, also seems to indicate that these dimensions might not be orthogonal. Small differences between years and genders were not significant, which seems to suggest a relative stability over time for these dimensions.

As expected the VICS-WA proved to be quite reliable. However, we questioned whether it is measuring *styles* or *abilities*. These doubts were cast mainly because of the lack of correlations with other measures of styles which was unexpected.
Peterson indicated that more research was necessary to explore the relations of this measure of styles with other instruments. From our data, with the particular samples in psychology 1 and psychology 2, whilst we determined that the reliability of the instrument is strong, the construct validity seem to be unfounded leading to the conclusion that this tool is measuring something else than styles. Because no longitudinal sample was available we also cannot determine whether the resulting styles are stable over time, however the small differences in the scores between years suggest that the types are not stable over time.
To maintain methodological consistency we used only two types of algorithms in the analysis. Of course, it is possible that the clustering algorithm selected is inappropriate for the data available, and it is strongly recommended to test a variety

9.1.3. Relations between instruments to measure styles

The interrelation between the various metrics and the emergent grouping patterns are the most interesting findings on styles.
Because of the lack of previous research we specifically selected measures of styles from different theoretical frameworks, therefore any relationship found is a novel piece of evidence. Overall, where we found correlations, these were fairly small (less than .3).

As mentioned above, no significant correlation was found between the VICS-WA and the other instruments, which clearly suggests that the VICS-WA is measuring something other than styles.

The intercorrelations between the other instruments are much more complex to disentangle. This is particularly the case for the relations between the ASSIST and TSI in which patterns are not stable for different year-classes. The CSI and the ASSIST present correlations between the strategic approach and the intuition scale (negative correlation) and analysis scale (positive correlation). This provided further support to the construct and face validity of the CSI as well as an insight in what is required to implement a strategic approach. Table 7.23 effectively highlights the patterns.

More dubious is the pattern of relations between the CSI and the TSI in which the lack of coherent patterns presented in the internal validity of the TSI might actually undermine the nature of the relationships with the CSI. Two dimensions (level and leaning) were an exceptions; the relations between these scales with the analysis and intuition dimensions followed consistent patterns.

A factor analysis was also attempted to determine whether it was possible to reduce the subscales to higher order factors, but this resulted in a complex 15-factors model. Although this was a superficial attempt, it suggests that the lack of consistent correlations might not be casual and supports a more complex representation of the relations between the instruments used.

This could be taken as either good or bad news by researchers in the style field. In fact, on one hand the relative independence of the different subscales and metrics might provide evidence to those who argue against styles as just a ragbag of incoherent metrics with little or no standardization. Unlike in the field of personality, in which a general consensus seems to have brought to the 5-factors model, a lot more work needs to be done with styles.

On the other hand the fact that the instruments considered in this thesis are weakly related could mean that the spectrum of styles is providing a better, richer characterization of individual differences and therefore could provide a wider range of possible practical applications.

9.1.4. Relations between styles and AP

In the second hypothesis (H2, p. 166) we asked whether there would be a correlation between styles and academic performance: the second major finding of this research was the surprising effectiveness of emergent stylistic groups (or types) in detecting differences in AP.

For each measure of styles we used data mining techniques (two-steps clustering) to identify groups. This provided a data-driven way of looking for stylistic types rather than score-based methods or norms. For all but the VICS-WA we found that the clusters were sensitive to differences in AP in year 1, making this methodology a feasible option to provide empirical data to inform pedagogical intervention strategies and/or automated systems.

This is the most appealing of all the findings so far, particularly for educational purposes as it clearly demonstrates that there is a link between certain *types* and AP. This finding is important for two reasons: if the aim of many researchers in styles is to demonstrate the practical utility of styles, this shows evidence that not all styles are good for the purpose of learning and studying.

Furthermore, we demonstrated that using a data-driven approach to the classification of *types*, could actually be effective in identifying useful characterizations (and have done so with three different instruments). To make this result useful, a core assumption must be that styles are intended as preferences and can be changed.
We gave some evidence that styles change over time, at least for the two year courses considered, suggesting that this assumption might be founded. However, a lot more work needs to be carried out to determine the effectiveness and strength of each *type*, and therefore lend support to the instrument used, in determining these differences in AP.

Looking back at hypothesis 2, the relations between styles and AP seem to be a lot more complex than expected with patterns not necessarily intuitive.
In fact, in H2.1 we expected a deep or strategic approach to be positively related to performance. However we found that the deep approach (full scale) is actually negatively correlated to performance, but the strategic approach is very highly correlated to grades at different points in both Y1 and Y2 with moderate to high strength and a peak in the correlation values between the year 2 final grade and the strategic scale ($r=.93$).
The lack of correlations between the VICS-WA and performance forced the rejection of H2.2 stating that a verbal style would be more related with the types of assessment in year 1 and a more analytic style could favour students taking MCQs and project reports types of assessment.

The relations between the TSI and performance are less coherent, but seem to follow the patterns highlighted above in respect of a particular form of assessment and specific

subscales. This partially supports H2.3 in which we expected some support of the relations between forms of assessment and styles.
H2.4 was also supported showing that intuition is negatively correlated with MCQ-type of assessment and analysis is positively correlated to the project reports.
Overall these correlation show that styles, particularly the subscales, are more related to specific forms of assessment than final grades, demonstrating that the interrelations between styles and AP are more complex than anticipated, however the clustering of *types* could be a useful diagnostic tool to tailor pedagogical interventions and detect at an early stages if students might need further help.

These finding are in line with the literature explored in chapters 3 and 4 and, as indicated by Cools (2008) and Coffield et al (2004), if we used a single instrument we could only have added some evidence in support of individual measures of styles.

By deliberately selecting different instruments to assess styles, the results could be placed more clearly in the matrix presented in figure 4.10 (p. 137), making progress in the theory, measurement and practical relevance of styles, particularly because it allowed to verify the identity of fundamental dimensions (not supported by our results in which it was difficult to reduce the dimensions), situate cognitive styles and link with other observable behaviours (i.e. longitudinal nature of the research conducted and relations to other behaviours such as attendance and punctuality).

9.1.5. Online usage as a window on behaviours and intentions

In chapter 2 we considered the impact of e-learning in the context of higher education and suggested that usage of online resources in a blended course could be useful to enhance students' experience and improve their learning. Although a direct link is difficult to single out because the use of online material is only a small part of students' activities, we set out to determine if such relation existed at all and to identify possible patterns which facilitate learning to the point of affecting grades.

In chapter 4 we identified some published research in which styles were observed in conjunction to the use of websites and e-learning and wanted to explore further the possible relations between measures of style and online usage.

Hypothesis 3 questioned whether an *appropriate* use of e-learning could improve student performance. A first superficial analysis of activity in relation to AP demonstrated that there are correlations between usage and AP. However, to test this hypothesis in full, a number of steps were necessary to provide a definition of *appropriate*.

This involved a procedure to reduce data complexity by defining dimensions of usage (extent, richness and efficiency), which allowed us to first give a meaningful representation of sessions (smallest units or episodes of online behaviour), then to characterise each students according to their behaviours.

The process led to the categorization of different *types* of sessions which provide a richer profile of students' behaviours. The profiles of the users were then used to consider differences in performance.
The findings show that the clusters originated from the three core dimensions of usage are related to AP in different types of assessment and that the emergent groupings are sensitive to differences in AP.

The results are encouraging for a number of reasons. First of all, we demonstrated how data mining techniques can be used to provide meaningful representations of usage, enriching understanding of individual students and relating these emergent features to AP.
Secondly, we provided a large amount of evidence in support that additional online resources *are* adding value to instruction in a blended form of teaching.
Although it was not possible to compare directly the effect of the implementation (i.e. course with or without it), we clearly showed that the explicit choice the students are able to make in using/not using the tools available, does have an impact on their performance.
From the data available it is not possible to say whether the choice is a cause or an effect of the fact that students have different degrees of motivation, but it might emerge from a more in-depth analysis of the first section of the ASSIST inventory and from the analysis of the relations between styles and e-learning which provided some clues about intentions.

The relation between styles and usage is not as strong as expected, but there seems to be consistency in the number of visits and efficiency of visits with a number of grades for different forms of assessment. This result was expected, as similar findings were pointed out in Vigentini (2008).

Most importantly, the in-depth characterisation of usage emerging in the clusters helped to explain performance. We were able to verify that the clusters, as it was the case for those emerging from styles, are sensitive enough to find differences in AP.
A consequence of this results is that we can argue that *how* students use the resources is important, which bring us back to the initial point of this section about *appropriate* use.

The details about the nature of the clusters (tables 8.4 to 8.7) helped us to explore this concept further and stress the value of the dimensions of usage (*extent*, *richness* and *efficiency*) as fundamental to provide support to Hypothesis 5, which questioned whether there are emergent styles in the usage of e-learning.

The last aspect explored was the relation between measures of styles and patterns of usage related to Hypothesis 4. Looking at the interrelations between the metrics of styles and usage some interesting correlations were found between the ASSIST and online usage with measures of extent of usage correlating with the subscales of the strategic and deep approach in particular.
Because patterns in year 1 and year 2 were dissimilar, it is not possible to argue that the relations are consistent, however, as we have indicated earlier for AP, the courses are different in demands, and it is perfectly reasonable to believe that if a student adjusts the styles to the course demands, the behaviour (and hence the online usage) will also change accordingly.
Similar patterns were found in the TSI in which there are differences in the courses. However, and most interestingly, we found correlations between the VICS-WA and the amount of usage and the V/I ratio and a negative correlation between efficiency and the W/A ratio.

Using a similar method to the one implemented for AP and styles, we looked at the possible differences in styles detected by the clustering of usage and this analysis did not produce significant differences.
Using styles as grouping factor, however, led to some significant differences in usage patterns, suggesting that styles have an impact on the expression of behaviours measured by online usage.

Based on these results it is quite difficult to portray a coherent picture: there are relations between styles and usage and these have been explained. It also seems plausible to argue that

typologies of styles are useful to differentiate patterns of usage as they were in differentiating AP.
Although the methods to reach this conclusion were novel, it is still consistent with previous research presented at the end of chapter 4 in which some examples of application of styles to adjust the behaviour of learning technology were reported.
The fact that the relations between styles and usage in our work are not particularly strong or consistent makes a strong case for further research in which more details are needed to understand how usage is affected by different stylistic preferences.

When considering the practical applications for adaptive systems, our findings suggest that the early attempts already developed and in development, might be based on limited evidence which hinders the effectiveness of the tools. In the successful cases presented, we can see the best potential. Like in the use of different variables in the organization of clusters presented, a single variable (percentage of use of social activities in a session) caused a reorganization of the clusters from 5 to 8. The extra variable, allowed to gain a richer and more differentiated set of groups which in turn provided a better foundation for an automated system to decide what to do to enhance the students' experience.

9.2. A model to understand academic performance

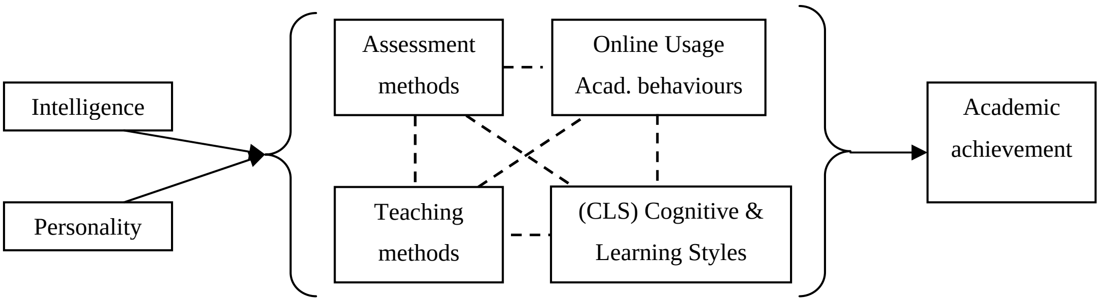
In chapters 3 and 4, we looked at the possible relations between prior performance, individual differences and academic performance at university.
We considered Chamorro-Premuzic summary of the correlations between the impact of prior scholastic achievement on university performance. He showed that the predicting power of prior performance decreased with age/level of education. (Figure 3.1)
Our data demonstrated that prior performance is a relatively good indicator of university grades.

As expected, the great majority of top performing students tend to improve very little, possibly reaching the peak of their abilities. However, there is a small group which perform better, 'maturing' at university and fulfilling their potential. Another small group was found, in which students clearly underperform in the university setting. These two groups are a testament of the fact that there is a complex interaction between teaching methods, individual differences in ability, personality and styles, and assessment demands, which can make a difference in terms of AP.

In chapter 4 we considered Furnham's model (figure 4.1) as a possible way of representing the interactions between the different aspects.
This model put styles as mediators of both intelligence and personality for AP. From the research conducted and the results obtained, the Furnham's model seems to be plausible: this research did not directly investigate personality or intelligence with traditional methods, but was able to shed more light on the middle part of the model, specifically looking at the interaction between cognitive styles, teaching methods (in our case with the support of e-learning) and assessment methods (with the breakdown of different components.

As demonstrable from the various results, the interactions seem to be more complicated than represented in Furnham's model and we try to provide an explanation in figure 9.1.

Option 1



Option 2

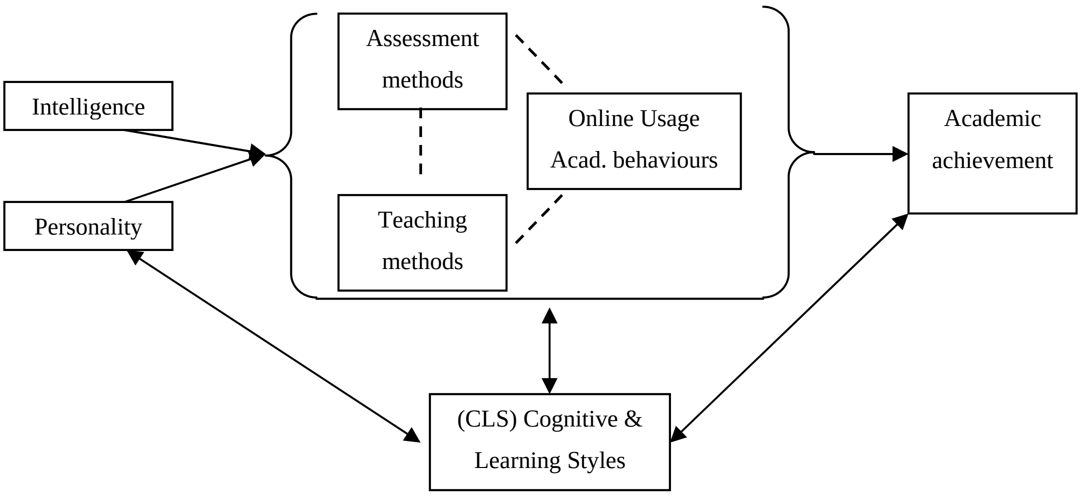


Figure 9.1. Possible extensions of Furnham's model which include behavioural expression (i.e. online usage).

One possibility (option 1) could be that CLS are an integral part of a feedback loop and a precursor of AP. Styles are largely malleable and the interaction with the course's teaching and assessment make the student adapt their styles and approaches based on the demands. From the data presented we could confirm that that ability (prior performance) has an impact on AP (the relation is both direct and mediated like in Furnham's model).

A slightly different model (option 2) shows that students' CLS come into the loop as mediators, partially as a consequence of course teaching and assessment. In this sense, CLS are a response, like an adaptive mechanism, which allow students to regulate their interaction with the courses based on previous experiences, performance and motivations. If a student is more flexible in adjusting their styles they can be more successful.

Whilst we brought some evidence of the former (both styles and usage correlate and can discriminate AP), the latter can only be speculated based on the fact that the correlations between styles and AP are stronger than usage and AP. Furthermore, whilst styles significantly differentiate patterns of usage, the opposite is not true. Although we found evidence of small changes in styles in our longitudinal sub-samples, more investigation is needed on the stylistic changes over time to determine whether their relations with the contexts is a cause or an effect of behaviours and ultimately AP.

9.3. A model to interpret the effectiveness and utility of e-learning

In chapter 2, after considering a variety of methods to evaluate the effectiveness of e-learning in education, we presented Johnson et al. (2008) model. Which was conceptualized on three aspects: course performance, course instrumentality and course satisfaction. Course satisfaction was superficially considered in this thesis, only at the group level via end of the year feedback surveys. Instead the main focus was on the other two aspects as they provide hard evidence, rather than opinions, about the impact of learning technology on a course. Six core factors affecting the evaluation of effectiveness were listed:

- human dimension (computer self-efficacy, motivation and individual skills & characteristics)
- design dimension (user satisfaction, usability)

- perceived usefulness (what students think about the system)
- interaction (behavioural proxies of usage and communication)
- social aspects (engagement & communication)
- learning (performance measures as learning outcomes)

Because we were satisfied of the preliminary results emerged from the STEER project about usability and satisfaction (i.e. Hardy et al. 2006), we focused on individual characteristics to explore the human dimensions (Johnson et al. (2008) and on interaction, social aspects and learning emerging from the long-term use of the online material offered.

Throughout chapter 7 and 8 we provided a wealth of evidence about the value of these metrics for AP. The question one ought to ask is whether the same information can be used to evaluate the effectiveness and utility of learning technology.
We believe that the answer is emphatically yes, they can be used to evaluate the system, and we would like to add that the value of the metrics obtained goes beyond the traditional methods of evaluation of the system, allowing practitioners to contextualise individual profiles for a more tailored intervention, when students become less effective or underperform.

We will exemplify by taking into consideration the analysis of sub-groups of performance. After dividing the longitudinal subsamples in 4 groups named according to their grades only, we obtained: the 'poorer' students (with average grades below 55), the underachievers' (who perform well in Y1 above 60-B grade- and drop below 55 -middle C grade-, in Y2), the 'improvers' (opposite of underachievers, moving from below 55 to over 60) and the 'better' students (with averages over 65 –middle B and A grades-, in both Y1 and Y2).

We then looked at significant differences in styles and usage to understand whether any of the metrics considered could be used as an indicator. A one-way ANOVA on the usage parameters using these four performance groups indicated that there are significant differences in the total number of visit, average visits per week, particularly in the first semester in Y1 (extent), and in the average depth of sessions (richness).
A similar analysis on the styles metrics indicated that there are differences in the deep and strategic approaches to learning.

A quick evaluation of this finding led to the formulation of a simple decision tree: in which 3 questions were asked:

- Does the student use what's available (particularly at the start of the year)?
- Is the student pattern of behaviour a regular one?
- Is the student simply confused/overwhelmed or is there a tendency to adopt an approach to study and learning which might be detrimental?

This crude method just presented is only one of the many possible practical uses in a university course and allows the portrayal of a simple yet powerful way of intervening early when students fall into patterns which might have detrimental effects on their university performance.

Considering that this system seems good enough, even with the fairly narrow, high-achieving group of students taking psychology at this university, it is expected that the patterns will be even more striking in a wider spectrum of achievement in other universities.

9.4. Purposes of student/user models

In chapter 2, following a very detailed analysis conducted by Brusilowsky (2001) we explored some of the reasons why a richer knowledge of students/users might be useful, particularly focusing on the utility of such profiles for learning technology and pedagogy.

One could argue that what we have done in this thesis is to provide detailed profiles of *stereotypes* and proposed that these could be effectively used for practical purposes to inform educators or automated systems.

As a matter of fact, already back in the 1970s Elaine Rich (1979, 1983), talked about stereotypes in user modelling. One example given was the use of descriptions of themselves to deduce the characteristics of books people might enjoy. A lot of ground has been covered and Brusilowsky (2001) observed that features-based models (stereotypes) were heavily implemented in the early nineties. Interestingly, the core of most systems was that content was the element modelled/adapted to the features of the class or group of users defined by the stereotypes. However, the content is what makes a system highly specialised in a domain and useless outside its original design.

Two problems are evident from the literature: one of the key issues of such systems was that the profiles were often too narrow to be useful. The second is that the specificity of the content limits the applications of the system.

Looking at the process we used in this thesis to construct the profiles, one key feature was that we actually *did not* focus on the content, but the description and role of the content, which is a meta-level, common across systems and applications. We extracted the patterns of features which could, in principle, provide the foundation for a more universal approach to adaptivity.

In fact sessions of online usage, or episodes were simply defined by the three core dimensions (extent, richness and efficiency) in which actions and intentions (abstracted from the sequences of usage) were the building blocks of such descriptions.

Earlier we demonstrated how a simple decision tree could be useful for an academic advisor to identify possible cases of students needing extra help or advice on how to tackle their studies; however, a similar approach can be automated, providing useful early feedback to students who increasingly receive less feedback in their courses.

In the next chapter, this research will be used to provide an excellent starting point to explore further the complex interrelations between individual differences and academic performance and to implement learning technology which ultimately fulfils the goal of a greater personalization of instruction, and an enhanced learning experience.

Chapter 10. Conclusions and future directions

The work done for this thesis was specifically placed at the boundaries between psychology, education and IT (specifically e-learning, and data mining). The principal aim of exploring the *utility* and *effectiveness* of learning technology in higher education offered a unique opportunity for contributions going beyond the discipline-specific area of psychology for which the thesis is submitted.

To conceptualise the utility of e-learning, it was necessary to provide a framework to understand students' learning in higher education, characterise the pedagogical role of technology and describe which aspects affect students' performance at university.
We suggested that, beside prior performance, students' individual differences (personality and behaviours online) could be an effective means to 1) understand what contributes to academic performance and 2) provide evidence in support of the evaluation of the effectiveness of e-learning.

The work presented allows advancements under four major headings:

- Methodology
- Psychology
- Education
- IT

10.1.1 Analysis methods in a novel setting

Despite the fact that data mining has previously been used in a variety of settings and disciplines, to our knowledge, this is the first example in which data mining techniques have been applied to this particular combination of complex data encompassing both personality and behaviours.

Three different application were demonstrated in this thesis: 1) making sense of the literature on style, where a meta-analytic approach was not appropriate because of the lack of a coherent system of measurement or methodological application and where a number of systematic reviews provided only a biased selection; 2) clustering applied to metrics of styles to find groups with specific combinations of preferences; 3) techniques applied to web usage mining (WUM) to identify suitable patterns in students' online behaviours.

10.1.2. Individual differences with styles

In the field of psychology this thesis contributed to enhance our understanding of online behaviours providing a rich characterization of usage and of individual students.
It also presented further evidence is support of the psychometric validity of four measures of styles. These measures can be used as an alternative way to look at individual differences other than personality or intelligence and provided a practical solution to differentiate between stylistic *types*.
The systematic exploration of the interrelations between different measures of styles, and the complex interactions with academic performance on one side and behaviours in a natural closed-world online system on the other, offered an original perspective in the study of the effects of individual preferences (measured by styles) on particular learning behaviours and approaches to studying and how these impact achievement at university
.

10.1.3. E-learning and pedagogy

In the field of education, after exploring the pedagogical principles in support of e-learning and personalization of instruction, it reaffirmed the reliability and validity of the ASSIST. This tool, which was thought to be useful in the self-appraisal of approaches to learning, turned out to be a powerful instrument able to account for a large amount of the variance in AP.

We also showed evidence of the positive impact of e-learning in the modern curriculum and related the effects to measures of styles. This helped in understanding in more detail how individual differences could be combined with AP in the attempt to deliver a personalised experience and hinted at the potential to use these variables together to produce a diagnostic profile, which allows targeted pedagogical interventions.
However, it is apparent that metrics of individual differences (styles and behaviours) can be used by practitioners to inform their pedagogy and feed back in the loop of course design, using empirical evidence, rather than surveys or other sources.

10.1.4. E-learning evaluation

In IT, because the implementation of the courses was an essential precondition to run this project, we demonstrated a case of good practice in which the evaluation of the courses, both from students' feedback as well as behavioural analysis, led to a continuous cycle of development and optimization which used a variety of technologies and systems without affecting the overall pedagogical goals of the courses.
We presented how metrics taken from psychology, encompassing both behaviour and individual differences, can be used effectively to inform system design, enhancing the understanding of users and their needs in a natural context over a prolonged period of time. The ecological validity of the profiles generated offers a wealth of information which convoluted experimental scenarios might not ever be able to replicate. We demonstrated that these sources of information could (and should) be considered in the evaluation of the implementations of e-learning.
We borrowed heavily from theories and techniques in informatics, data mining and information visualization and applied them in a different domain: all these proved to be highly effective in discovering *useful* and *usable* patterns in the large database, suggesting that more research should be conducted in applying similar techniques outside the domain they were originally thought for.

10.2. Limitations

Although this research presents unique features normally unavailable for the timescale of a doctoral thesis, such as the 5-year span which included a number of longitudinal sub-samples, there are two core criticisms which could be put forward: one is theoretical, the other is practical.

At the theoretical level, it could be argued that the choice of styles might not be appropriate, and the selection of the four instruments is not ideal. The selection of styles over personality and intelligence is justified because of the strong theoretical entrenchment of the two and because of a fundamental factor behind both aspects: these are stable features of an individual. The practical implication is that knowledge of the relations with e-learning would be useful to inform system design, but would not be able to provide an insight into what can be changed for the better. Styles, instead, because are malleable and affected by the context, offered a suitable alternative.

A second theoretical criticism concerns the abstraction to intentions from the behavioural click stream: this has been carefully considered and the conclusion that an abstraction was necessary seems obvious to reduce the complexity of behaviours over prolonged period of time. An impressive leap from the descriptive view of data to the interpretations of intentions and behaviours was clearly demonstrated.

A third theoretical criticism could be directed at the choice of the clustering algorithms used. This obviously affected the way in which groups were formed and the following analysis of the differences in AP and styles. However, as this was a novel approach applied to the field of psychology, the results are very promising and only leave room for improvement.

From the practical point of view, the biggest limitation was the inability to collect full datasets for all participants limiting the sample sizes considerably. We were also unable to control for the conditions in which teaching and learning occurred, starting from the change in e-learning systems to the small variations in course delivery and assessment requirements. The former made it impossible to reduce all behavioural data to a common denominator which might have lead to some biases in the frequency of the classifications of actions and tools used by students.

The practical limitations, however, could also be seen as a major strength because, to date, we were unable to find examples of research conducted in an ecologically valid setting.

10.3. Directions

Replication of the findings in a different discipline and/or institution would help validate the wide-ranging reach of the conclusions proposed. However, there are numerous ways in which this research can be developed further.

The use of data mining techniques and information visualization offered a reliable starting point and proved to be useful in supporting the research conducted. A first possible improvement would be the applications of other clustering techniques to classify both styles and usage -possibly more suitable for the type of data-.

As in many other cases in which cross-disciplinary insemination occur, the cases demonstrated how similar problems could be tackled using these techniques.
The research could also be expanded to include records of style throughout the full university curriculum. Although this is well beyond the reach of a single doctoral programme, it would provide essential data to enrich our understanding of how styles develop over time.

With the advent of web 2.0 technologies, the rules of interaction in e-learning are changing at very rapid pace. It would be essential to step into this domain to study how the shift from passive users to active contributors is relating to styles and AP. The richer variety of expression of behaviours, still in a closed-world system, can provide a more differentiated set of groupings from the behavioural data.
Another example would be to expand the understanding of individual differences by complementing measures of styles with measures of anxiety and coping. These additional details can highlight how particular students, who might feature in our analysis as similar types, might lead to very different performance outcomes.
The practical advantage of this approach is that a richer individual profile could be used proactively to aid students who fail to cope with the demands of higher education. Such an exploration might provide sufficient details about students for an automated warning system to identify potentially problematic cases. In fact, the implementation of an automated system based on the simple rules and relations suggested in the last chapter is one of the most interesting applications of the findings. Questions can be raised about the practicality of producing such detailed psychometric profiles as well as the potential misuse of personal data, but the potential to improve students' performance clearly justifies the effort.

References

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, *22*(2), 227-257.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131,* 30-60.

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

Ackerman, P. L., & Rolfhus, E. L. (1999). The locus of adult intelligence: Knowledge, abilities, and nonability traits. *Psychology and Aging, 14,* 314-330.

Ackerman, P. L., & Woltz, D. J. (1994). Determinants of Learning and Performance in an Associative Memory/Substitution Task: Task Constraints, Individual Differences, Volition, and Motivation. *Journal of Educational Psychology, 86*(4), 487–515.

Alavi, M., & Leidner, D. E. (2001). Research commentary: Technology-mediated learning--A call for greater depth and breadth of research. *Information Systems Research*, *12*(1), 1.

Alavi, M., Yoo, Y., & Vogel, D. R. (1997). Using information technology to add value to management education. *Academy of management Journal*, *40*(6), 1310-1333.

Alderman, B., & Fletcher, S. (2005). The role of interaction in enhancing achievement and student satisfaction in an online course: A rubric analysis. In *Conference of the Open and Distance Learning Association of Australia (ODLAA) 2005*.

Alexander, W. P. (1935). Intelligence, concrete and abstract. *British Journal of Psychology Monograph Supplement*, *19*(177), 935.

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

Allinson, C. W., & Hayes, J. (1996). The cognitive style index: A measure of intuition-analysis for organizational research. *Journal of Management Studies*, *33*(1), 119-135.

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, *47*(1), 171–220.

Almeda, M. B., & Rose, K. (2000). Instructor satisfaction in University of California Extension's on-line writing curriculum. In *Online Education: Proceedings of the Sloan Summer Workshop on Asynchronous Learning Networks* (p. 175). Olin College-Sloan-C.

Anderson, J. R. (1981). *Cognitive skills and their acquisition*. Lawrence Erlbaum.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological review*, *89*(4), 369-406.

Austin, E. J., Deary, I. J., & Gibson, G. J. (1997). Relationships between ability and personality: Three hypotheses tested. *Intelligence*, *25*(1), 49-70.

Austin, E. J., Deary, I. J., Whiteman, M. C., Fowkes, F. G. R., Pedersen, N. L., Rabbitt, P., Bent, N., et al. (2002). Relationships between ability and personality: does intelligence contribute positively to personal and social adjustment? *Personality and Individual Differences*, *32*(8), 1391-1411.

Austin, E. J., Hofer, S. M., Deary, I. J., & Eber, H. W. (2000). Interactions between intelligence and personality: Results from two large samples. *Personality and Individual Differences*, *29*(3), 405-427.

Ausubel, D. P., Novak, J. D., & Hanesian, H. (1968). *Educational psychology: A cognitive view*. Holt, Rinehart and Winston New York.

Ausubel, D. P., & Youssef, M. (1963). Role of discriminability in meaningful paralleled learning. *Journal of Educational Psychology*, *54*(6), 331-336.

Bacher, J., Wenzig, K., & Vogler, M. (2004). SPSS TwoStep Cluster–A First Evaluation. *Arbeits-und Diskussionspapiere*, *2*.

Baldi, M., Baralis, E., Risso, F., & e Inf, D. A. (2005). Data mining techniques for effective and scalable traffic analysis. In *Integrated Network Management, 2005. IM 2005. 2005 9th IFIP/IEEE International Symposium on* (pp. 105-118).

Barton, K., Dielman, T. E., & Cattell, R. B. (1972). Personality and IQ measures as predictors of school achievement. *Journal of Educational Psychology*, *63*(4), 398.

Bauer, K. W., & Liang, Q. (2003). The effect of personality and precollege characteristics on first-year activities and academic performance. *Journal of College Student Development*, *44*(3), 277–290.

Berendt, B., Hotho, A., & Stumme, G. (2002). Towards semantic web mining. *The Semantic Web—ISWC 2002*, 264-278.

Berendt, B., Mobasher, B., Spiliopoulou, M., & Wiltshire, J. (2001). Measuring the accuracy of sessionizers for web usage analysis. *Proceedings of the Web Mining Workshop at the First SIAM International Conference on Data Mining*, 7–14.

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., Wallet, P. A., et al. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research*, *74*(3), 379.

Biggs, J. (1979). Individual differences in study processes and the quality of learning outcomes. *Higher Education*, *8*(4), 381-394.

Biggs, J. B. (1987). *Student Approaches to Learning and Studying. Research Monograph.* Australian Council for Educational Research Ltd., Radford House, Frederick St., Hawthorn 3122, Australia.

Biggs, J. B., & (Guildford, G. S. F. R. I. H. E. (1999). *Teaching for quality learning at university*.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, *32*(3), 347–364.

Binet, A., & Simon, T. (1905). New methods for the diagnosis of the intellectual level of subnormals. *L'annee Psychologique, 12*, 191–244.

Black, P. J. (1993). Formative and summative assessment by teachers. *Studies in Science Education, 21*(1), 49-97.

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Educ.: Principles, Policy & Practice*, *5*(1), 7–74. doi:10.1080/0969595980050102

Blackwell, L., Trzesniewski, K., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*(1), 246-263.

Blickle, G. (1996). Personality traits, learning stratigies, and performance. *European Journal of Personality*, *10*(5), 337-352. doi:10.1002/(SICI)1099-0984(199612)10:5<337::AID-PER258>3.0.CO;2-7

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*(2), 187.

Bloom, B. S. (1968). Learning for mastery. *Evaluation comment*, *1*(2), 1-5.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Lawrence Erlbaum.

Bonk, C. J., Graham, C. R., Cross, J., & Moore, M. G. (2005). *The handbook of blended learning: Global perspectives, local designs*. Pfeiffer & Company.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text* 1. *Cognitive psychology*, *11*(2), 177-220.

Brand, C., Egan, V., & Deary, I. J. (1994). Intelligence, personality, and society: Constructivist versus essentialist possibilities. *Current topics in human intelligence*, *4*, 29–42.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall1. *Journal of verbal learning and verbal behavior*, *11*(6), 717-726.

Bruner, J. (1985). Models of the learner. *Educational researcher*, *14*(6), 5.

Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User modeling and user-adapted interaction*, *6*(2), 87-129.

Brusilovsky, P. (2001). Adaptive hypermedia. *User modeling and user-adapted interaction*, *11*(1), 87-110.

Brusilovsky, P., & Millán, E. (2007). User models for adaptive hypermedia and adaptive educational systems. *The adaptive web*, 3-53.

Burton, L. J., & Nelson, L. (2006). The relationships between personality, approaches to learning, and academic success in first-year psychology distance education students. In *Proceedings of the 2006 Higher Education Research and Development Society of Australasia Annual Conference* (Vol. 29, pp. 64-72). Higher Education Research and Development Society of Australasia (HERDSA).

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*(1), 129–140.

Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences, 29*(6), 1057–1068.

Buzzetto-More, N. A., & Alade, A. J. (2006). Best practices in e-assessment. *Journal of Information Technology Education*, *5*(5), 251-269.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*(2), 197–253.

Carroll, J. B. (1993a). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge Univ Pr.

Carroll, J. B. (1993b). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. *Contemporary intellectual assessment: Theories, tests, and issues*, *16*, 122–130.

Cashion, J., & Palmieri, P. (2002). The secret is the teacher. *The learner's view of online learning*.

Cashion, J., & Palmieri, P. (2009). *'The secret is the teacher': the learner's view of online learning*. National Centre for Vocational Education Research (NCVER).

Catledge, L. D., & Pitkow, J. E. (1995). Characterizing browsing strategies in the World-Wide Web. *Computer Networks and ISDN systems*, *27*(6), 1065–1073.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, *38*(592), 10.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: a critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22.

Cattell, R. B. (1979). *Personality and learning theory. 1. The structure of personality in its environment*. Springer.

Cattell, R. B., & Bernard, R. (1946). Description and measurement of personality.

Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement*, 139–164.

Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality*, *37*(4), 319-338.

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

Chamorro-Premuzic, T., & Furnham, A. (2005). *Personality and intellectual competence*. Lawrence Erlbaum Associates.

Chamorro-Premuzic, T., & Furhnam, A. (2003). Personality traits and academic examination performance. *European journal of Personality*, *17*(3), 237–250. doi:10.1002/per.473

Chamorro-Premuzic, T., & Furhnam, A. (2006). Intellectual competence and the intelligent personality: A third way in differential psychology. *Review of General Psychology*, *10*(3), 251–267. doi:10.1037/1089-2680.10.3.251

Chamorro-Premuzic, T., & Furhnam, A. (2008). Personality, intelligence and approaches to learning as predictors of academic performance. *Personality and Individual Differences*, *44*(7), 1596–1603.

Cheminais, R. (2002). *Inclusion and school improvement: a practical guide*. David Fulton Publish.

Chen, G., Gully, S. M., Whiteman, J. A., & Kilcullen, R. N. (2000). Examination of relationships among trait-like individual differences, state-like individual differences, and learning performance. *The Journal of applied psychology*, *85*(6), 835.

Chen, S., & Liu, X. (2008). An integrated approach for modeling learning patterns of students in Web-based instruction: \ldots. *portal.acm.org*. Retrieved from http://portal.acm.org/citation.cfm?id=1352782.1352783

Chen, S. Y., & Liu, X. (2004). The contribution of data mining to information science. *Journal of Information Science*, *30*(6), 550.

Chen, S. Y., & Macredie, R. D. (2004). Cognitive modeling of student learning in web-based instructional programs. *International Journal of Human-Computer Interaction*, *17*(3), 375-402.

Chen, S. Y., & Macredie, R. D. (2005). The assessment of usability of electronic shopping: a heuristic evaluation. *International Journal of Information Management*, *25*(6), 516-532.

Chi, E. H., Pirolli, P., Chen, K., & Pitkow, J. E. (2001). Using information scent to model user information needs and actions and the Web. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 490–497.

Chi, E. H., Pirolli, P., & Pitkow, J. E. (2000). The scent of a site: A system for analyzing and predicting information scent, usage, and usability of a web site. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 161–168.

Clancey, W. J. (1997). *Situated cognition: On human knowledge and computer representations*. Cambridge Univ Pr.

Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004a). Should we be using learning styles?: what research has to say to practice/Frank Coffield...[et al.]. *voced.edu.au*. Retrieved from http://www.voced.edu.au/td/tnc_79.72

Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004b). Learning styles and pedagogy in post-16 learning: a systematic and critical review/Frank Coffield... \ldots. *voced.edu.au*. Retrieved from http://www.voced.edu.au/td/tnc_79.71

Cohen, P. A. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. *Research in Higher Education*, *13*(4), 321-341.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, *51*(3), 281.

Cohen, S. (1987). Instructional alignment: Searching for a magic bullet. *Educational Researcher*, *16*(8), 16.

Cohen, S. (1984). Implications of instructional psychological research on mastery learning. *Outcomes, A quarterly newsletter of the Network of Outcome-based schools*, *2*, 18-25.

Conole, G., & Oliver, M. (2007). Contemporary perspectives in e-learning research: themes, methods and impact on practice.

Conole, G., Dyke, M., Oliver, M., & Seale, J. (2004). Mapping pedagogy and tools for effective learning design. *Computers and Education*, *43*(1), 17–33.

Conole, G., Laat, M. D., Dillon, T., & Darby, J. (2008). `Disruptive technologies',`pedagogical innovation': What's new? Findings from an in-depth study of students' use and perception of technology. *Computers & Education*, *50*(2), 511–524.

Cook, D. (1999). Behavior Bucks: A Unique Motivational Program. *Intervention in School and Clinic*, *34*(5), 307.

Cook, D. A. (2005). Learning and cognitive styles in web-based learning: theory, evidence, and application. *Academic medicine*, *80*(3), 266.

Cook, T. D., Campbell, D. T., & Peracchio, L. (1979). *Quasi-experimentation*. Rand McNally Chicago.

Cooley, R., Mobasher, B., & Srivastava, J. (1997). Web mining: Information and pattern discovery on the world wide web. In *Ninth IEEE International Conference on Tools with Artificial Intelligence, 1997. Proceedings.* (pp. 558-567).

Cooley, R. W. (2000). *Web usage mining: Discovery and application of interesting patterns from Web data*. University of Minnesota.

Cooley, R. W. (2000). Web usage mining: Discovery and application of interesting patterns from Web data.

Cools, E., Evans, C., & Redmond, J. A. (2009). Using styles for more effective learning in multicultural and e-learning environments. *Multicultural Education & Technology Journal*, *3*(1).

Costa, P. T., & McRae, R. R. (2001). Temperament in context.

Cowan, B. R., Vigentini, L., & Jack, M. A. (2009). Exploring the effects of experience on wiki anxiety and wiki usability: an online study. In *Proceedings of the*

2009 British Computer Society Conference on Human-Computer Interaction (pp. 175-183). British Computer Society.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American psychologist*, *12*(11), 671-684.

Cugini, J., & Scholtz, J. (1999). VISVIP: 3D visualization of paths through web sites. *Proceedings of International Workshop on Web-Based Information Visualization (WebVis' 99). Florence, Italy: IEEE Computer Society*.

Curry, L. (1983a). An Organization of Learning Styles Theory and Constructs.

Curry, L. (1983b). An Organization of Learning Styles Theory and Constructs.

Curry, L. (1987). Integrating concepts of cognitive or learning style: A review with attention to psychometric standards. *Ottawa, ON: Canadian College of Health Service Executives*.

David, M., Brennan, J., Broadfoot, P., Brown, A., Davis, P., Entwistle, N., Fuller, M., et al. (2009). Effective learning and teaching in UK higher education. *Imprint*, *67*, 7.

De Barbenza, C. M., & Montoya, O. A. (1974). Academic achievement in relation to personality characteristics in university students. *Revista Latinoamericana de Psicologia*, *6*, 331–340.

De Raad, B. (1998). Personality traits in learning and education. *European Journal of Personality*, *10*(3), 185-200.

De Raad, B., & Schouwenburg, H. C. (1998). Personality in learning and education: A review. *European Journal of Personality*, *10*(5), 303-336.

De Villiers, M. R. (1999). Applying the Hexa-C Metamodel of instructional theory and design to educational web applications. *Proceedings of Web Net*, *99*.

De Villiers, M. R., & Dersley, R. (2003). Emerging technology for emerging peoples: evaluation of an African AIDS-awareness training system. In *Proceedings of Ed-Media 2003 World Conference on Educational Hypermedia & Telecommunications*.

Deary, I. J. (2001). Human intelligence differences: a recent history. *Trends in Cognitive Sciences*, *5*(3), 127-130.

Delone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, *19*(4), 9-30.

Demetriou, A., & Kazi, S. (2001). *Unity and modularity in the mind and the self: Studies on the relationships between self-awareness, personality, and intellectual development from childhood to adolescence*. Routledge.

Dempster, J. (2004). *CAP e-Learning guides: Evaluating E-Learning*. Centre of Academic Excellence, University of Warwick, www. warwick. ac. uk/go/ca p/resources/eguides.

Der, G., & Deary, I. J. (2003). IQ, reaction time and the differentiation hypothesis. *Intelligence*, *31*(5), 491-503.

Desmedt, E., & Valcke, M. (2004). Mapping the Learning Styles "Jungle": An overview of the literature based on citation analysis. *Educ. Psych.*, *24*(4), 445–464. doi:10.1080/0144341042000228843

DeVellis, R. F. (2003). *Scale development: Theory and applications*. Sage Publications, Inc.

Diseth, A. (2003a). Personality and Approaches to Learning as predictors of Academic Achievement. *European journal of Personality*, *17*, 143–155. doi:10.1002/per.469

Diseth, A. (2003b). Approaches to Learning, Cognitive Style, and Motives as Predictors of Academic Achievement. *Educ. Psych.*, *23*(2), 195–207. doi:10.1080/01443410303225

Dollinger, S. J., Matyja, A., & Huber, J. (2008). Which factors best account for academic success: Those which college students can control or those \ldots. *Journal of Research in Personality*, *42*, 872–885.

Dollinger, S. J., & Orf, L. A. (1991). Personality and performance in personality conscientiousness and openness. *Journal of research in personality(Print)*, *25*(3), 276–284.

Driscoll, M. P. (1984). Alternative paradigms for research in instructional systems. *Journal of Instructional Development*, *7*(4), 2-5.

D'Silva, R., & Reeder, K. (2005). Factors that influence faculty members' uptake and continued use of course management systems. *jep*, *36*(6), 1071–1073.

Duff, A., Boyle, E., Dunleavy, K., & Ferguson, J. (2004). The relationship between personality, approach to learning and academic performance. *Personality and Individual Differences*, *36*(8), 1907-1920. doi:10.1016/j.paid.2003.08.020

Duffy, T. M., & Jonassen, D. H. (1992). *Constructivism and the technology of instruction: A conversation*. Lawrence Erlbaum.

Dufresne, A., & Turcotte, S. (1997). Cognitive style and its implications for navigation strategies. *Artificial intelligence in education knowledge and \ldots*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.46.6441&rep=rep1&type=pdf

Dunn, R. S., & Dunn, K. J. (1975). *Educator's self-teaching guide to individualizing instructional programs*. Parker Publishing Company.

Durand, D., & Kahn, P. (1998). MAPA: a system for inducing and visualizing hierarchy in Websites. In *Proceedings of the ninth ACM conference on Hypertext and hypermedia: links, objects, time and space---structure in hypermedia systems: links, objects, time and space---structure in hypermedia systems* (pp. 66-76). ACM.

Dweck, C. (2005). Self-theories. *Handbook of Competence and Motivation*, 122.

Eisenberg, M. B. (2008). Information literacy: essential skills for the information age. *DESIDOC Journal of Library & Information Technology*, *28*(2), 39–47.

Ely, D. P., & Plomp, T. (1996). *Classic writings on instructional technology*.

Entwistle, N. (1987). *Understanding classroom learning*. Hodder and Stoughton London.

Entwistle, N. (1988). Motivational factors in students' approaches to learning. *Learning strategies and learning styles,* 21-51.

Entwistle, N., & Tait, H. (1990). Approaches to learning, evaluations of teaching, and preferences for contrasting academic environments. *Higher Education*, *19*(2), 169-194.

Entwistle, N. J. (1972). Personality and Academic Attainment. *British Journal of Educational Psychology*.

Entwistle, N. J. (1981). *Styles of learning and teaching: an integrated outline of educational psychology for students, teachers and lecturers*. Wiley Hoboken, NJ.

Entwistle, N. J., & Ramsden, P. (1983). Understanding Student Learning (London, Croom Helm).

Entwistle, N. (1982). Study skills and independent learning. *Studies in Higher Education*, *7*(1), 65–73.

Entwistle, N., & Entwistle, D. (2003). Preparing for examinations: The interplay of memorising and understanding, and the development of knowledge objects. *Higher Education Research & Development*, *22*(1), 19–41.

Entwistle, N., & McCune, V. (2004). The conceptual bases of study strategy inventories. *Educational Psychology Review*, *16*(4), 325–345.

Epstein, S. (1998). Cognitive-experiential self-theory. *Advanced personality*, 211–238.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of personality and social psychology*, *71*, 390-405.

Ertmer, P. A., & Newby, T. J. (1996). The expert learner: Strategic, self-regulated, and reflective. *Instructional Science*, *24*(1), 1-24.

Etzioni, O. (1996). The World-Wide Web: quagmire or gold mine? *Communications of the ACM*, *39*(11), 65-68.

Eysenck, H. J. (1967). *The dynamics of anxiety and hysteria*. Routledge & Kegan Paul London.

Eysenck, H. J. (1992). Four ways five factors are not basic. *Personality and Individual Differences*, *13*(6), 667–673.

Eysenck, H. J. (1993). Creativity and personality: Suggestions for a theory. *Psychological Inquiry*, 147-178.

Eysenck, H. J. (1994). Personality and intelligence: Psychometric and experimental approaches. *Personality and intelligence*, 3-31.

Eysenck, H. J. (1997). Creativity and personality: word association, origence and psychoticism. *Eminent creativity, everyday creativity and health*, 107–118.

Eysenck, H. J., & White, P. O. (1964). Personality and the measurement of intelligence. *British Journal of Educational Psychology, 34*(2), 197–202.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion (Washington, DC), 7*(2), 336.

Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and individual differences*, *6*(1), 21–29.

Farsides, T., & Woodfield, R. (2003). Individual differences and undergraduate academic success: The roles of personality, intelligence, and application. *Personality and Individual differences*, *34*(7), 1225–1243.

Fernandez, H., Kommers, P., & Asensio, M. (2004). Conceptual representations for in-depth learning. *Cognitive Support for Learning: Imagining the Unknown*, 229.

Flynn, J. R. (2007). *What is intelligence?: beyond the Flynn effect*. Cambridge University Press.

Fowell, S. P., & Levy, P. (1995). Computer-mediated communication in the information curriculum: an initiative in computer-supported collaborative learning. *Education for Information*, *13*(3), 193-210.

Fruyt, F. D., & Mervielde, I. (1996). Personality and interests as predictors of educational streaming and achievement. *European Journal of Personality*, *10*(5).

Fu, W. T., & Pirolli, P. (2007). SNIF-ACT: A cognitive model of user navigation on the World Wide Web. *Human-Computer Interaction*, *22*(4), 355–412.

Furnham, A. (1992). Personality and learning style: a study of three instruments. *Personality and Individual Differences*, *13*(4), 429-438.

Furnham, A., Forde, L., & Cotter, T. (1998). Personality and intelligence. *Personality and individual Differences, 24*(2), 187-192.

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., & Mitchell, J. (1991). Personality, needs, social skills and academic achievement: A longitudinal study. *Personality and Individual Differences*, *12*(10), 1067-1073.

Gagne, R. M. (1965). *The conditions of learning*. Holt, Rinehart and Winston.

Gagné, R. M., Wager, W. W., Golas, K. C., & Keller, J. M. (2005). *Principles of instructional design*. Thomson/Wadsworth.

Galton, F. (1883). *Inquiries Into Human Faculty and Its Development: sir Francis Galton*. Macmillan.

Gardner, H. (1993a). *The unschooled mind: How children think and how schools should teach*. Basic books.

Gardner, H. (1993b). *Frames of mind: The theory of multiple intelligences*. Basic Books.

Garnham, N. (2000). ‘INFORMATION SOCIETY’AS THEORY OR IDEOLOGY: A CRITICAL PERSPECTIVE IN TECHNOLOGY, EDUCATION AND EMPLOYMENT IN THE INFORMATION AGE. *Information, Communication & Society*, *3*(2), 139–152.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Psychology Press.

Gibbs, G., & Simpson, C. (1999). Conditions under which assessment supports students’ learning. *Editors’ Introductio n*, 3.

Gilb, T., & Finzi, S. (1988). *Principles of software engineering management*. Addison-Wesley Wokingham, UK.

Gipps, C. (1999). Sociocultural perspectives on assessment. *Review of Research in Education*, *24*(1), 355-392.

Glaser, R. (1991). The maturing of the relationship between the science of learning and cognition and educational practice. *Learning and Instruction*, *1*(2), 129-144. doi:doi: DOI: 10.1016/0959-4752(91)90023-2

Goddard, H. H. (1908). The Binet and Simon tests of intellectual capacity. *The Training School*, *5*(3).

Goff, M., & Ackerman, P. L. (1992). Personality-intelligence relations: Assessment of typical intellectual engagement. *Journal of Educational Psychology*, *84*(4), 537–552.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe*, *7*, 7–28.

Goleman, D. (1995). Emotional intelligence. *New York*.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*(1), 79–132.

Graff, M., Davies, J., & McNorton, M. (2004). Cognitive style and cross cultural differences in internet use and computer attitudes. *European Journal of Open, Distance and E-Learning*, *2004*, *2*, 1–8.

Graff, M. (2003). Learning from web-based instructional systems and cognitive style. *British Journal of Educational Technology*, *34*(4), 407–418.

Graff, M. (2005). Individual differences in hypertext browsing strategies. *Behaviour & Information Technology*, *24*(2), 93–99. doi:10.1080/01449290512331321848

Graff, M. (2006). Constructing and maintaining an effective hypertext-based learning environment: Web-based learning and cognitive style. *Education + Training*, *48*(2/3), 143–155. doi:10.1108/00400910610651773

Green, A., Peters, T. J., & Webster, D. J. T. (2009). An assessment of academic performance and personality. *Medical Education*, *25*(4), 343-348.

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. *Handbook of educational psychology*, 15-46.

Greeno, J. (1980). Psychology of Learning, 1960-1980: One Participant.

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin*, *124*(1), 75-111.

Halamandaris, K. F., & Power, K. G. (1999). Individual differences, social support and coping with the examination stress: A study of the psychosocial and academic adjustment of first year home students. *Personality and individual differences*, *26*(4), 665-685.

Hall, E., & Moseley, D. (2005). Is there a role for learning styles in personalised education and training? *Int. J. of Lifelong Educ.*, *24*(3), 243–255. doi:10.1080/02601370500134933

Hand, D. J. (2000). Data mining: new challenges for statisticians. *Social Science Computer Review*, *18*(4), 442.

Hardy, J., Bates, S., Antonioletti, M., & Seed, T. (2005). Integrating e-learning and on-campus teaching II: evaluation of student use. In *Exploring the Frontiers of e-learning: borders, outposts and migration', Research Proceedings of the 12th Association of Learning Technology Conference* (pp. 140-153).

Hardy, J., Bates, S., McKain, D., Murray, K., Paterson, J., McGonigle, B., Vigentini, L., et al. (2006). The modus operandi of the next-generation e-learner: An analysis of tracking usage across the disciplines. *Association for Learning Technology*, 108.

Harris, D. (1931). *The relation to college grades of some factors other than intelligence*. Columbia university.

Harris, D. (1940). Factors affecting college grades: a review of the literature, 1930-1937. *Psychological Bulletin*, *37*(3), 125-66.

Harvey, J. (1998). Evaluation Cookbook, 1–88.

Hayes, J., & Allinson, C. W. (1998). Cognitive style and the theory and practice of individual and collective learning in organizations. *Human Relations*, *51*(7), 847–871.

Heilesen, S. B., & Josephsen, J. (2008). E-learning: Between augmentation and disruption? *Computers & Education*, *50*(2), 525–534.

Heinze, A., & Procter, C. T. (2004). Reflections on the use of blended learning.

Hellweg, S. A. (1984). The Pervasiveness and Impact of Electronic Communication Technologies in Organizations: A Survey of Major American Corporations.

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

Higgins, J. W., Williams, R. L., & McLaughlin, T. F. (2001). The Effects of a Token Economy Employing Instructional Consequences for a Third-Grade Student with Learning Disabilities: A Data-based Case Study. *Education and Treatment of Children*, *24*(1), 99-106.

Hiltz, S. R. (1994). *The virtual classroom: Learning without limits via computer networks*. Ablex Publishing Corporation.

Hiltz, S. R., & Turoff, M. (2002). What makes learning networks effective? *Communications of the ACM*, *45*(4), 56-59.

Hockey, R. (1984). Varieties of attentional state: The effects of environment. *Varieties of attention*, 449–483.

Hodgkinson, G., & Sadler-Smith, E. (2003). Reflections on reflections... on the nature of intuition, analysis and the construct validity of the Cognitive Style Index. *Journal of Occupational and Organizational Psychology*, *76*(2), 279–281.

Hodgkinson, G. P., Sadler-Smith, E., Hayes, J., Allinson, C. W., HUDSON, R., & KEASEY, K. (2003). Complex or unitary? A critique and empirical re-assessment of the Allinson-Hayes Cognitive Style Index. Commentary. Authors' reply. *Journal of Occupational and Organizational Psychology*, *76*, 243–281.

Hodgkinson, G. P., Sadler-Smith, E., Sinclair, M., & Ashkanasy, N. M. (2009). More than meets the eye? Intuition and analysis revisited. *Personality and individual differences*, *47*(4), 342–346. doi:10.1016/j.paid.2009.03.025

Holmes, B., & Gardner, J. (2006). *E-learning: Concepts and practice*. Sage Publications Ltd.

Horton, W. (2001). *Evaluating e-learning*. ASTD.

Hull, C. L. (1945). Principles of behavior. *The Journal of Nervous and Mental Disease*, *101*(4), 396.

Illeris, K. (2002). *The three dimensions of learning*. Roskilde University Press.

Ingram, R., & Benford, S. (1996). The application of legibility techniques to enhance information visualizations. *The Computer Journal*, *39*(10), 819.

Jackson, L. M., Pancer, S. M., Pratt, M. W., & Hunsberger, B. E. (2006). Great Expectations: The Relation Between Expectancies and Adjustment During the Transition to University1. *Journal of Applied Social Psychology*, *30*(10), 2100-2125.

Jarvis, P. (2006). *Towards a comprehensive theory of human learning*. Routledge.

Jensen, A. R. (1972). Genetics and Education.

Johnson, R. D., Hornik, S., & Salas, E. (2008). An empirical examination of factors contributing to the creation of successful e-learning environments. *International journal of human-computer studies*, *66*(5), 356-369.

Jonassen, D. H., & Reeves, T. C. (1996). Learning with technology: Using computers as cognitive tools. *Handbook of research for educational communications and technology*, *1*, 693-719.

Kagan, J. (1964). *The matching familiar figures test*. Kagan.

Kagan, J. (1966). Reflection-impulsivity: The generality and dynamics of conceptual tempo. *Journal of Abnormal Psychology*, *71*(1), 17-24.

Kanfer, F. H. (1990). The scientist-practitioner connection: A bridge in need of constant attention. *Professional Psychology: Research and Practice*, *21*(4), 264-270.

Kaushik, A. (2007). *Web analytics*. John Wiley and Sons.

Keen, A. (2007). *The cult of the amateur: how today's internet is killing our culture*. Broadway Business.

Kennedy, G. (2009). Adam Smith and the invisible hand: from metaphor to myth. *Econ Journal Watch*, *6*(2), 239-263.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge Univ Pr.

Kirshner, D., & Whitson, J. A. (1997). *Situated cognition: Social, semiotic, and psychological perspectives*. Lawrence Erlbaum.

Kirton, M. (1976). Adaptors and innovators: A description and measure. *Journal of Applied Psychology*, *61*(5), 622-629.

Kirton, M. J. (1994). A theory of cognitive style. *Adaptors and innovators: Styles of creativity and problem solving*, 1–33.

Kitajima, M., & Polson, P. G. (1995). A comprehension-based model of correct performance and errors in skilled, display-based, human-computer interaction. *International Journal of Human Computer Studies*, *43*(1), 65-100.

Kitajima, M., & Polson, P. G. (1997). A comprehension-based model of exploration. *Human-Computer Interaction*, *12*(4), 345-389.

Kizzier, L. D., & Pollar, C. (1992). A research agenda for studying technologically mediated instructional strategies in business and education. In *Proceedings of the 1992 DPE National Research Conference* (pp. 117-125).

Klein, G. S. (1951). The personal world through perception. *Perception: an approach to personality. New York: Ronald*, 328-355.

Klein, G. S., & Schlesinger, H. J. (1951). Perceptual attitudes toward instability: I. Prediction of apparent movement experiences from Rorschach responses. *Journal of Personality*, *19*, 289-302.

Kobsa, A. (2001). An empirical comparison of three commercial information visualization systems. In *Proceedings of the IEEE Symposium on Information Visualization* (p. 123). Citeseer.

Kobsa, A., Koenemann, J., & Pohl, W. (2001). Personalised hypermedia presentation techniques for improving online customer relationships. *The Knowledge Engineering Review*, *16*(02), 111-155.

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice Hall.

Kommers, P., & Lanzing, J. (1998). Mapas conceptuales para el diseño de sistemas hipermedia. Navegación por la Web y autoevaluación. *Nuevas tecnologías para el aprendizaje*, 103-127.

de Koning, K., Bredeweg, B., Breuker, J., & Wielinga, B. (2000). Model-based reasoning about learner behaviour. *Artificial Intelligence*, *117*(2), 173-229.

Kosala, R., & Blockeel, H. (2000). Web mining research: A survey. *ACM SIGKDD Explorations Newsletter*, *2*(1), 1-15.

Koutri, M., Avouris, N., & Daskalaki, S. (2004). A survey on web usage mining techniques for web-based adaptive hypermedia systems. *Proceedings of the Adaptable and Adaptive Hypermedia Systems*, 125–149.

Kozhevnikov, M. (2007). Cognitive styles in the context of modern psychology: Toward an integrated framework of cognitive style. *Psychological Bulletin*, *133*(3), 464–481. doi:10.1037/0033-2909.133.3.464

Kozma, R. B. (1991). Learning with media. *Review of educational research*, *61*(2), 179.

Kulik, J. A., & McKeachie, W. J. (1975). 7: The Evaluation of Teachers in Higher Education. *Review of research in education*, *3*(1), 210.

Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences*, *42*(3), 441–451.

Landau, J. D., & Bavaria, A. J. (2003). Does Deliberate Source Monitoring Reduce Students' Misconceptions About Psychology? *Teaching of Psychology*, *30*, 311-314.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, *104*(2), 211-240.

Landauer, T. K., & Nachbar, D. W. (1985). Selection from alphabetic and numeric menu trees using a touch screen: breadth, depth, and width. *ACM SIGCHI Bulletin*, *16*(4), 78.

Larson, K., & Czerwinski, M. (1998). Web page design: implications of memory, structure and scent for information retrieval. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 25-32). ACM Press/Addison-Wesley Publishing Co.

Laurillard, D. (1987). Computers and the emancipation of students: giving control to the learner. *Instructional Science*, *16*(1), 3-18.

Laurillard, D. (2002). Rethinking teaching for the knowledge society. *Educause Review*, *37*, 16-27.

Laurillard, D. (2005). Rethinking university teaching. *Physical Sciences Educational Reviews*, 60.

Lazarus, R. S., DeLongis, A., Folkman, S., & Gruen, R. (1985). Stress and adaptational outcomes: The problem of confounded measures. *American Psychologist*, *40*(7), 770-779.

Lea, S. J., Stephenson, D., & Troy, J. (2003). Higher Education Students Attitudes to Student-centred Learning: beyond educational bulimia ? *Studies in Higher Education*, *28*(3), 321-334.

Lee, E., & MacGregor, J. (1985). Minimizing user search time in menu retrieval systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *27*(2), 157-162.

Lenhart, A., Simon, M., & Graziano, M. (2001). The Internet and Education: Findings of the Pew Internet & American Life Project.

Levy, Y. (2007). Comparing dropouts and persistence in e-learning courses. *Computers & education*, *48*(2), 185-204.

Lieberman, M.D. (2007) Social cognitive neuroscience. A review of the core processes. *Annual review of psychology*, 58, 259-289.

Logie, R. H., Cocchini, G., Delia Sala, S., & Baddeley, A. D. (2004). Is There a Specific Executive Capacity for Dual Task Coordination? Evidence From Alzheimer's Disease. *Neuropsychology*, *18*(3), 504-513.

Lorenzo, G., & Moore, J. (2002). Five pillars of quality online education. *The Sloan Consortium Report to the Nation*, 15-09.

Lounsbury, J. W., Sundstrom, E., Loveland, J. M., & Gibson, L. W. (2003). Intelligence,"Big Five" personality traits, and work drive as predictors of course grade. *Personality and Individual Differences*, *35*(6), 1231–1239.

Makkonen, P. (1998). WWW-based presentations as a complementary part of conventional lectures in the basic of informatics. *ACM SIGCSE Bulletin*, *30*(3), 165.

Makkonen, P. (2003). Is web-based seminar an effective way of learning in adult education? *Proceedings of the 36th Hawaii International Conference on System Sciences*.

Malone, T. W. (1980). What makes things fun to learn? Heuristics for designing instructional computer games. *Proceedings of the 3rd ACM SIGSMALL symposium and the first SIGPC symposium on Small systems*, 162–169.

Mandinach, E. B. (2005). The development of effective evaluation methods for e-learning: A concept paper and action plan. *Teachers College Record*, *107*(8), 1814–1835.

Markham, S., Ceddia, J., Sheard, J., Burvill, C., Weir, J., Field, B., Sterling, L., et al. (2003). Applying agent technology to evaluation tasks in e-learning environments. *Proceedings of the exploring educational technologies conference*, 16–17.

Marton, F., & Saljo, R. (1976). On qualitative differences in learning: I. Outcome and process. *British Journal of Educational Psychology*, *46*(1), 4-11.

Mathieu, J. E., Tannenbaum, S. I., & Salas, E. (1992). Influences of individual and situational characteristics on measures of training effectiveness. *Academy of Management Journal*, *35*(4), 828-847.

Matthews, G. (1992). Mood. *Handbook of human performance*, *3*, 161-194.

Matthews, G. (2000). *Human performance: Cognition, stress, and individual differences*. Psychology Pr.

Matthews, G., & Wells, A. (1999). The cognitive science of attention and emotion. *Handbook of cognition and emotion*, 171–192.

Matthews, G., Campbell, S. E., Falconer, S., Joyner, L. A., Huggins, J., Gilliland, K., Grier, R., et al. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion*, *2*(4), 315–340. doi:10.1037//1528-3542.2.4.315

Mayer, A. (1955). Die gegenwaertige Problematik der Arbeits-und Betriebspsychologie. The present problematic situation of labor and industrial psychology. *Psychologisch Rundschau*, *6*, 6–18.

McDonald, S., & Stevenson, R. J. (1998). Effects of text structure and prior knowledge of the learner on navigation in hypertext. *Human Factors*, *40*(1).

McGinnis, J., Friman, P., & Carltyon, W. (1999). The Effect Of Token Rewards On “Intrinsic Motivation” For Doing Math. *Journal of Applied Behavior Analysis*, *32*(3), 375.

McGonigle, B., & Chalmers, M. (2007). Executive functioning as a window on the evolution and development of cognitive systems. *International Journal of Comparative Psychology, Special Issue*.

McGonigle, B., & Chalmers, M. (2008). Putting Descartes before the horse (again!). *Behavioral and Brain Sciences*, *31*(02), 142-143.

McKeachie, W. J. (1974). The Decline and Fall of the Laws of Learning. *Educational Researcher*, *3*(3), 7-11. doi:10.2307/1174605

McNaught, C. (2003). Identifying the complexity of factors in the sharing and reuse of resources. *Reusing online resources: a sustainable approach to e-Learning*.

Mehta, P., & Kumar, D. (1985). Relationships of academic achievement with intelligence, personality, adjustment, study habits and academic motivation. *Journal of Personality and Clinical Studies*, *1*(1-2), 57-68.

Merrill, M. (2002). First principles of instruction. *Educational Technology Research and Development*, *50*(3), 43-59. doi:10.1007/BF02505024

Messick, S. (1976). Personality consistencies in cognition and creativity. *Individuality in learning*, 4-22.

Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement*, *21*(3), 215-237.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, *23*(2), 13.

Messick, S. (1996). Bridging cognition and personality in education: the role of style in performance and development. *European Journal of Personality*, *10*(5).

Miller, A. (1991). Personality types, learning styles and educational goals. *Educational Psychology*, *11*(3), 217-238.

Miller, A. (1987). Cognitive Styles: an integrated model. *Educ. Psych.*, *7*(4), 251–268. doi:10.1080/0144341870070401

Miller, C. S., & Remington, R. W. (2004). Modeling information navigation: Implications for information architecture. *Human-Computer Interaction*, *19*(3), 225-271.

Miller, G. A. (1984). Informavores. *The study of information: Interdisciplinary messages*, 111-113.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. Holt, Rinehart and Winston New York.

Mobasher, B., Dai, H., Luo, T., & Nakagawa, M. (2002). Discovery and evaluation of aggregate usage profiles for web personalization. *Data Mining and Knowledge Discovery*, *6*(1), 61–82.

Moran, A. (1991). What can learning styles research learn from cognitive psychology? *Educational Psychology*, *11*(3), 239-245.

Mõttus, R., Pullmann, H., & Allik, J. (2006). Toward more readable big five personality inventories. *European Journal of Psychological Assessment*, *22*(3), 149-157.

Myers, I. B., McCaulley, M. H., Quenk, N. L., & Hammer, A. L. (1999). *MBTI manual*. Consulting Psychologists Press.

Naidu, S. (2003a). Designing instruction for e-learning environments. *Handbook of distance education*, 349-365.

Naidu, S. (2003b). *Learning & teaching with technology: principles and practices*. RoutledgeFalmer.

Nguyen, N. T., Allen, L. C., & Fraccastoro, K. (2005). Personality predicts academic performance: Exploring the moderating role of gender. *Journal of Higher Education Policy and Management*, *27*(1), 105-117.

Nicholas, D., Huntington, P., Williams, P., & Dobrowolski, T. (2004). Re-appraising information seeking behaviour in a digital environment. *Journal of Documentation*, *60*(1), 24-43.

Nicol, D. (2006). Increasing success in first year courses: Assessment re-design, self-regulation and learning technologies. *ASCILITE*.

Nicol, D., & Macfarlane-Dick, D. (2006). Rethinking formative assessment in HE: a theoretical model and seven principles of good feedback practice. *Higher Education Academy, Accessed*, *8*.

Nielsen, T. (2006). Mental Self-government: Relationships between Personality Dimensions, Educational History and Status, and Learning Styles, 1–35.

Nielsen, T., Kreiner, S., & Styles, I. (2007). Mental self-government: development of the additional democratic learning style scale using Rasch measurement models. *Journal of applied measurement*, *8*(2), 124.

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology*, *93*(1), 116–130. doi:10.1037/0022-3514.93.1.116

Norman, D. A. (1980). *Errors in human performance*.

Norman, D. A. (1981). Categorization of action slips. *Psychological review*, *88*(1), 1-15.

Oberauer, K., Schulze, R., Wilhelm, O., & Sü\ss, H. (2005). Working Memory and Intelligence–Their Correlation and Their Relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 61–65. doi:10.1037/0033-2909.131.1.61

O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, *43*(5), 971–990.

Oliver, M., MacBean, J., Conole, G., & Harvey, J. (2002). Using a toolkit to support the evaluation of learning. *Journal of Computer Assisted Learning*, *18*(2), 199-208.

Olson, D. R., & Bruner, J. S. (1974). Learning through experience and learning through media. *Media and symbols: The forms of expression, communication, and education. Chicago, IL: National Society for the Study of Education*.

Olson, G. M., & Olson, J. S. (2003). Human-Computer Interaction: Psychological Aspects of the Human Use of Computing. *Annual Review of Psychology*, *54*(1), 491–516. doi:10.1146/annurev.psych.54.101601.145044

O'Neill, K., Singh, G., & O'Donoghue, J. (2004). Implementing elearning programmes for higher education: A review of the literature. *Journal of Information Technology Education*, *3*, 313-322.

Paivio, A. (1971). *Imagery and verbal processes*. Holt, Rinehart and Winston New York.

Palfrey, J. G., & Gasser, U. (2008). *Born digital*. Basic Books.

Pascual-Cid, V. (2008). An information visualisation system for the understanding of web data. In *IEEE Symposium on Visual Analytics Science and Technology, 2008. VAST'08* (pp. 183-184).

Pascual-Cid, V., Baeza-Yates, R., Dürsteler, J. C., Minguez, S., & Middleton, C. (2009). New Techniques for Visualising Web Navigational Data. In *2009 13th International Conference Information Visualisation* (pp. 621-626). IEEE.

Paunonen, S. V. (1998). Hierarchical organization of personality and prediction of behavior. *Journal of Personality and Social Psychology*, *74*(2), 538–556.

Paunonen, S. V., & Ashton, M. C. (2001). Big Five predictors of academic achievement. *Journal of Research in Personality*, *35*(1), 78–90.

Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher*, *18*(1), 16.

Perry, W. G. (1970). Forms of intellectual and ethical development. *New York: Holt, Rinehart and Winston*.

Peterson, E. R., & Deary, I. J. (2006). Examining wholistic-analytic style using preferences in early information processing. *Personality and Individual Differences, 41*(1), 3-14.

Peterson, E. R., Deary, I. J., & Austin, E. J. (2003a). On the assessment of cognitive style: four red herrings. *Personality and Individual Differences*, *34*(5), 899-904.

Peterson, E. R., Deary, I. J., & Austin, E. J. (2003b). The reliability of Riding's cognitive style analysis test. *Personality and Individual Differences, 34*(5), 881-891.

Peterson, E. R., Deary, I. J., & Austin, E. J. (2005). A new measure of verbal-imagery cognitive style: VICS. *Personality and Individual Differences, 38*(6), 1269-1281.

Peterson, E. R., Rayner, S. G., & Armstrong, S. J. (2009). Researching the psychology of cognitive style and learning style: Is there really a future? *Learning and Individual Differences*.

Peterson, E. R. (2002). *Cognitive Style: The Reliability and Validity of Verbal-Imagery and Wholistic-Analytic Cognitive Style*. PhD thesis, .

Phillips, R., Bain, J., McNaught, C., Rice, M., & Tripp, D. (2000). Handbook for learning-centred evaluation of computer-facilitated learning projects in higher education. *Committee for University Teaching and Staff Development Project. Retrieved April*, *9*, 2002.

Piccoli, G., Ahmad, R., & Ives, B. (2001). Web-based virtual learning environments: A research framework and a preliminary assessment of effectiveness in basic IT skills training. *Mis Quarterly*, 401–426.

Pillay, H. (1998). An Investigation of the Effect of Individual Cognitive Preferences on Learning through Computer-based Instruction. *Educ. Psych.*, *18*(2), 171–182. doi:10.1080/0144341980180203

Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information*.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological review, 106*(4), 643-675.

Pitkow, J. E. (1999). Summary of WWW characterizations. *World Wide Web, 2*(1), 3–13.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*(2), 322–338. doi:10.1037/a0014996

Porter, A. (2003). An examination of the reliability and construct validity of the Thinking Styles Inventory. In *7th Annual ELSIN Conference*. Presented at the European Learning Styles Information Network Conference, Ghent.

Porter, L. (2006). *Student behaviour: Theory and practice for teachers*. Allen & Unwin.

Pospisil, P., Iyer, L. K., Adelstein, S. J., & Kassis, A. I. (2006). A combined approach to data mining of textual and structured data to identify cancer-related targets. *BMC bioinformatics*, *7*(1), 354.

Prensky, M. (2006). Don't bother me mom–I'm learning. *Continuum*.

Raad, B. D., & Schouwenburg, H. C. (1996). Personality in learning and education: A review. *European Journal of Personality*, *10*(5).

Ramsden, P. (1992). *Learning to teach in higher education*. Routledge.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510-532.

Rayner, S. (2007). A teaching elixir, learning chimera or just fool's gold? Do learning styles matter? *Support for Learning*, *22*(1), 24–30.

Rayner, S., & Riding, R. (1997). Towards a categorisation of cognitive styles and learning styles. *Educational Psychology*, *17*(1), 5-27.

Reeve, C. L., Meyer, R. D., & Bonaccio, S. (2006). Intelligence-personality associations reconsidered: The importance of distinguishing between general and narrow dimensions of intelligence. *Intelligence*, *34*(4), 387-402. doi:10.1016/j.intell.2005.11.001

Reid, G. (2005). *Learning styles and inclusion*. Sage Publications Ltd.

Resnick, L. B. (1981). Instructional psychology. *Annual Review of Psychology*, *32*(1), 659–704.

Riding, R., & Cheema, I. (1991). Cognitive styles—an overview and integration. *Educational Psychology*, *11*(3), 193-215.

Riding, R., & Rayner, S. (1998). *Cognitive styles and learning strategies: Understanding style differences in learning and behaviour*. David Fulton London, UK.

Riding, R. J. (1997). On the nature of cognitive style. *Educational Psychology*, *17*(1), 29-49.

Riding, R. J., & Rayner, S. (1991). *Learning styles*. Carfax.

Rindermann, H., & Neubauer, A. C. (2001). The influence of personality on three aspects of cognitive performance: Processing speed, intelligence and school performance. *Personality and Individual Differences*, *30*(5), 829-842.

Robertson, I. T. (1985). Human information-processing strategies and style. *Behaviour & Information Technology*, *4*(1), 19-29.

Robins, R. W., & Pals, J. (2002). Implicit self-theories of ability in the academic domain: A test of Dweck's model. *Self and Identity*, *1*, 313–336.

Rogers, D. L. (2000). A paradigm shift: Technology integration for higher education in the new millennium. *Educational Technology Review*, 19-27.

Rolfhus, E. L., & Ackerman, P. L. (1996). Self-report knowledge: At the crossroads of ability, interest, and personality. *Journal of Educational Psychology*, *88*(1), 174-188.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems With Applications*, *33*(1), 135–146.

Rosenfeld, M., & Rosenfeld, S. (2008). Developing effective teacher beliefs about learners: The role of sensitizing teachers to individual learning differences. *Educational Psychology*, *28*(3), 245-272.

Ross, S. M., Morrison, G. R., & O'Dell, J. K. (1989). Uses and effects of learner control of context and instructional support in computer-based instruction. *Educational Technology Research and Development*, *37*(4), 29-39.

Rothstein, M. G., Paunonen, S. V., Rush, J. C., & King, G. A. (1994). Personality and cognitive ability predictors of performance in graduate business school. *Journal of Educational Psychology*, *86*(4), 516-530.

Rothwell, W. J., & Kazanas, H. C. (2008). *Mastering the Instructional Design Process*. John Wiley & Sons.

Rouet, J. (1996). *Hypertext and cognition*. Routledge.

Rumelhart, D. E. (1979). Some problems with the notion of literal meanings. *Metaphor and thought*, 78-90.

Sadler-Smith, E., Spicer, D. and Tsang, F. (2000) 'Validity of the cognitive style index: Replication and extension' *British Journal of Management*, 11 (2) 175-181.

Sadler-Smith, E. (2001). The relationship between learning style and cognitive style. *Personality and individual differences*, *30*(4), 609–616.

Schmeck, R. R., & Grove, E. (1979). Academic achievement and individual differences in learning processes. *Applied Psychological Measurement*, *3*(1), 43.

Schouwenburg, H. C., & Lay, C. H. (1995). Trait procrastination and the big-five factors of personality. *Personality and Individual Differences*, *18*(4), 481-490.

Seddon, P. B. (1997). A respecification and extension of the DeLone and McLean model of IS success. *Information Systems Research*, *8*(3), 240-253.

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety, Stress & Coping*, *4*(1), 27-41.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference.

Sheard, J., Ceddia, J., Hurst, J., & Tuovinen, J. (2003). Inferring student learning behaviour from website interactions: A usage analysis. *Education and Information Technologies*, *8*(3), 245–266.

Sheard, J., & Markham, S. (2005). Web-based learning environments: developing a framework for evaluation. *Assessment & Evaluation in Higher Educ.*, *30*(4), 353–368. doi:10.1080/02602930500099169

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational researcher*, *29*(7), 4.

Shiozawa, H., Okada, K., & Matsushita, Y. (1999). 3D interactive visualization for inter-cell dependencies ofspreadsheets. In *1999 IEEE Symposium on Information Visualization, 1999.(Info Vis' 99) Proceedings* (pp. 79-82).

Shohreh, K., & Garland, K. (2000). «Development of a Web-Based Internet Marketing Course». *Journal of Marketing Education*, *22*(2), 84-89.

Shuell, T. J. (1986). Cognitive conceptions of learning. *Review of Educational Research*, 411–436.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational researcher*, *15*(2), 4.

Silber, K. H. (1998). The cognitive approach to training development: A practitioner's assessment. *Educational Technology Research and Development*, *46*(4), 58-72.

Skinner, B. F. (1963). *Are theories of learning necessary?* Bobbs-Merrill. Retrieved from http://psychclassics.yorku.ca/Skinner/Theories/

Smith, C., Colbourn, C., Cunliffe-Charlesworth, H., & Lever, R. (2006). Creating lifelong e-learners: Meeting the need for flexible and individual approaches to e-learning. In *European Association of Distance Teaching Universities Conference Proceedings*.

Snyder, C. R., Harris, C., Anderson, J. R., Holleran, S. A., Irving, L. M., Sigmon, S. T., Yoshinobu, L., et al. (1991). The will and the ways: Development and validation of an individual-differences measure of hope. *Journal of Personality and Social Psychology*, *60*(4), 570-585.

Spearman, C. (1904). " General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, *16*(2).

Spence, K. W. (1951). Theoretical interpretations of learning. *Handbook of experimental psychology. New York: Wiley*, 690-729.

Spiliopoulou, M. (2000). Web usage mining for web site evaluation. *Communications of the ACM*, *43*(8), 127-134.

Srivastava, J., Cooley, R. W., Deshpande, M., & Tan, P. N. (2000). Web usage mining: Discovery and applications of usage patterns from web data. *SIGKDD explorations*, *1*(2), 12–23.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory*. Princeton Univ Pr.

Sternberg, R. J. (1981). The Evolution of Theories of Intelligence. *Intelligence, 5*(3), 209–30.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge University Press.

Sternberg, R. J. (1996). Personality, pupils, and purple cows: we have the right answers, but do we have the right questions? *European Journal of Personality*, *10*(5), 447–452.

Sternberg, R. J. (2005). The theory of successful intelligence. *R. interam. Psicol*, *39*, 2.

Sternberg, R. J., & Grigorenko, E. L. (1997). Are cognitive styles still in style?. *American Psychologist*, *52*(7), 700-712.

Sternberg, R. J., & Ruzgis, P. (1994). *Personality and intelligence*. Cambridge University Press.

Sternberg, R. J., & Wagner, R. K. (1993). The g-ocentric View of Intelligence and Job Performance Is Wrong. *Current Directions in Psychological Science*, *2*(1), 1–5.

Sternberg, R. J. (1999). *Thinking styles*. Cambridge University Press.

Sternberg, R. J. (1999). Successful intelligence: finding a balance. *Trends in cognitive science*, *3*(11), 436-442.

Stevens, C. (2001). The Conscientiousness of Students in Subject Pools: Implications for “Laboratory” Research. *Journal of Research in Personality, 35*(1), 91–97. doi:10.1006/jrpe.2000.2310

Thorndike, E. L. (1911). Animal intelligence. *New York*, 1-297.

Thorndike, E. L. (1922). *The psychology of arithmetic*. The Macmillan company.

Toohey, S. (1999). *Designing Courses for Higher Education.* Open University Press, 325 Chestnut Street, Philadelphia, PA 19106 ($34.95). E-mail: enquiries@ openup. co. uk; Web site:< http://www. openup. co. uk>.

Tracey, J. B., Hinkin, T. R., Tannenbaum, S., & Mathieu, J. E. (2001). The influence of individual characteristics and the work environment on varying levels of training outcomes. *Human Resource Development Quarterly, 12*(1), 5.

Triantafillou, E., Pomportsis, A., & Demetriadis, S. (2003). The design and the formative evaluation of an adaptive educational system based on cognitive styles* 1. *Computers & Education*, *41*(1), 87-103.

Upton, D., & Adams, S. (2006). Individual differences in online learning. *Psychology Learning and Teaching*, *5*(2), 141-145.

Uruchrutu, E., MacKinnon, L., & Rist, R. (2005). User Cognitive Style and Interface Design for Personal, Adaptive Learning. What to Model? *User Modeling 2005*, 154-163.

Vermunt, J. D. (1995). Process-oriented instruction in learning and thinking strategies. *European Journal of Psychology of Education*, *10*(4), 325-349.

Vermunt, J. D. (1996). Metacognitive, cognitive and affective aspects of learning styles and strategies: A phenomenographic analysis. *Higher education*, *31*(1), 25-50.

Vernon, P. E. (1961). *The structure of human abilities*. Methuen young books.

Vigentini, L. (2008). 'Wikify' teaching: technological interactivity to support teaching and methods to evaluate its effectiveness. Presented at the PLAT 2008, Bath.

Vigentini, L., & McGonigle, B. (2005). E-learning trajectories: supplementing teaching of undergraduate Psychology
foundation courses in year 1 and 2. Presented at the e-learning@Greenwich, London.

Volet, S., & Järvelä, S. (2001). *Motivation in learning contexts: Theoretical and methodological implications*. Emerald Group Pub Ltd.

Vroom, V. H. (1966). Organizational choice: A study of pre-and postdecision processes. *Organizational Behavior and Human Performance*, *1*(2), 212-225.

Vygotski, L. S., Vygotsky, L. S., & Cole, M. (1978). *Mind in society: The development of higher psychological processes*. Harvard Univ Pr.

Watson, J. B. (1924). *Behaviorism New York*. Norton.

Webb, E. (1915). *Character and intelligence: An attempt at an exact study of character*. University Press.

Webster, J., & Hackley, P. (1997). Teaching effectiveness in technology-mediated distance learning. *Academy of Management Journal*, 1282-1309.

Webster, R. (2001). *Interfaces for E-Learning Cognitive Styles and Software Agents for Web-Based Learning Support*.

Wechsler, D. (1943). Non-intellective factors in general intelligence. *Journal of Abnormal and Social Psychology*, *38*(1), 101-103.

Wenger, E. (1999). *Communities of practice: Learning, meaning, and identity*. Cambridge Univ Pr.

Wenger, E., McDermott, R. A., & Snyder, W. (2002). *Cultivating communities of practice: A guide to managing knowledge*. Harvard Business Press.

White, M. D., Cruz, I. R., Cameron, B. D., & Drabinski, E. (2008). RefViz 2 (review). *Portal-Libraries and the Academy*, *8*(3), 343-344.

Wild, K. P., & Schiefele, U. (1994). Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens [Learning strategies in university studies: Results concerning the dimensional structure and reliability of a new questionnaire]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *15*(4), 185-200.

Williams, K. (2001). What is the digital divide. In *d3 workshop (digital divide doctoral students), Ann Arbor, August*. Citeseer.

Williams, K. (2003). Literacy and Computer Literacy: Analyzing the NRC's Being Fluent with Information Technology. *Journal of learning technology*, *3*(1), 1–20.

Winer, G. A., Cottrell, J. E., Gregg, V., Fournier, J. S., & Bica, L. A. (2002). Fundamentally misunderstanding visual perception: Adults' belief in visual emissions. *American Psychologist*, *57*(6-7), 417-424.

Witkin, H. A., & Goodenough, D. R. (1981). *Cognitive styles: Essence and origins*. International Universities Press New York.

Witkin, H. A., Lewis, H. B., Hertzman, M., Machover, K., Meissner, P. B., & Wapner, S. (1954). Personality Through Perception New York. *Harper & Row*, *359*, 451-460.

Witkin, H. A., Moore, C. A., Goodenough, D. R., & Cox, P. W. (1977). Field-Dependent and Field-Independent Cognitive Styles and Their Educational Implications. *Review of Educational Research*, *47*(1), 1-64.

Witten, I. H., Bray, Z., Mahoui, M., & Teahan, B. (1999). Text mining: A new frontier for lossless compression. In *dcc* (p. 198). Published by the IEEE Computer Society.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann Pub.

Wolfe, R. N., & Johnson, S. D. (1995). Personality as a Predictor of College Performance. *Educational and Psychological Measurement*, *55*(2), 177-185. doi:10.1177/0013164495055002002

Wooldridge, A. (2005). The brains business. *The Economist*, *8*.

Zhang, L. F. (2004). Revisiting the predictive power of thinking styles for academic performance. *The Journal of Psychology: Interdisciplinary and Applied*, *138*(4), 351-370.

Zhang, L. (1999). Further cross-cultural validation of the theory of mental self-government. *The Journal of psychology*, *133*(2), 165–181.

Zhang, L., & Sternberg, R. J. (2000). Are learning approaches and thinking styles related? A study in two Chinese populations. *The Journal of psychology*, *134*(5), 469–489.

Zhang, L., & Sternberg, R. J. (2005). A Threefold Model of Intellectual Styles. *Educational Psychology Review*, *17*(1), 1–53. doi:10.1007/s10648-005-1635-4

Zmud, R. W. (1979). Individual differences and MIS success: A review of the empirical literature. *Management Science*, *25*(10), 966-979.

Zuckerman, M., Kuhlman, D. M., Joireman, J., Teta, P., & Kraft, M. (1993). A comparison of three structural models for personality: The Big Three, the Big Five, and the Alternative Five. *Journal of Personality and Social Psychology*, *65*(4), 757-768.

Appendix 1. Web server logs and Web usage mining

This document is loosely based on the Web Characterization Terminology & Definitions Sheet (WCA 1999).

Definitions of terms

The term *Internet* (or the Net) is quite recent: most definitions agree that it is a network of computer networks which operates world-wide using a common set of communications protocols (W3C). However, Bartolini suggested “a temporal information space supported by a computer network infrastructure through which information is transferred according to agreed common standards” (2001). This is preferred to shift the focus from the infrastructure to a definition which encompasses its fluid and dynamic nature.

The *World Wide Web* (WWW) is a hypertext system of cross-linked data sources, which permits easy access to or publication of complex data types, including text, graphics, sound and animation, across the Internet. Initially developed at CERN (the European Centre for Nuclear Research) in Geneva, Switzerland, it became the most common and most used manifestation of the internet.

The basic elements of this system are a *server* and a *client*. The classic definition for client is the role adopted by an application when it retrieves and/or renders resources or resource manifestations, whereas the specific one for the Web defines the Web client as an application "capable of accessing Web resources by issuing requests and render responses containing Web resource manifestations".

The server is "the role adopted by an application when it is supplying resources or resource manifestations" to the requesting client. In most of the cases, spiders and crawlers apart (robot applications which automatically parse the internet for a variety of reasons), a request from a client is issued by a user.
In the context of the WWW, to distinguish a resource from another it needs to be identified. The term Uniform Resource Identifier (URI) is “a compact string of characters for identifying an abstract or physical resource.”

In the case of web documents, we use the URL: the term Uniform Resource Locator which is “a subset of URI that identify resources via a representation of their primary access mechanism (e.g., their network 'location'), rather than identifying the resource by name or by some other attribute(s) of that resource”.
A URL is something that every Internet user manages daily and it an essential reference to understand web traffic.

Within this framework, it is usually very hard to identify a user (not to be confused with a client, which has unique parameters), however, in the case of a private network or a Virtual Learning Environment (VLE) like WebCT, users can access a specific resource only if permission is granted by some sort of authentication and the user becomes clearly identifiable.

A resource manifestation takes place after a client has performed a request to a server and this provides a response. In Web architecture, there is another layer that plays an important role in the communication system: the HTTP proxy.

A proxy is "an intermediary which acts as both a server and a client for the purpose of retrieving resources or resource manifestations on behalf of other clients. Clients using a proxy know the proxy is present and that it is an intermediary".

In many cases the server is not able to identify the proxy. A very important feature of proxies is proxy caching, which can be used to reduce the downloading time of resources requested by users and the load of network traffic between the client and the server. Since the early days of the WWW, caching is also commonly performed at client side by the internet browsers which keep a certain amount of data stored locally and allows faster loading time of the documents (especially pictures and other media).

All internet activity is usually recorded in some way; there are three ways in which data is recorded:

- **server level collection**: the server stores data regarding requests performed by the client, thus data regard generally just one source;
- **client level collection**: it is the client itself which sends to a repository information regarding the user's behaviour (this can be done either with an ad-hoc browsing application or through client-side applications running on standard browsers);
- **proxy level collection**: information is stored at the proxy side, thus Web data regards several Websites, but only users whose Web clients pass through the proxy (this level is ignored in our analysis).

Server side data

Server side data or HTTP server information is automatically gathered by Web servers and is usually collected in access log files. CERN and NCSA specified a *Common Log Format* (CLF) for every access stored in a log file, and this is supported by most of the HTTP servers.

```
129.215.50.195 - s9906490 [28/Nov/2005:18:23:08 +0000] "GET
/SCRIPT/PS0002_7/scripts/serve_home HTTP/1.1" 200 46189
"https://www.webct.ed.ac.uk/webct/homearea/homearea?" "Mozilla/5.0 (Windows; U;
Windows NT 5.1; en-GB; rv:1.7.5) Gecko/20041110 Firefox/1.0"
```

Fig. 1. A single log line from an APACHE HTTP server

Every log entry conforming to the CLF contains these fields:

- **client IP address** or hostname (if DNS lookups are performed);
- **user id ('-'** if anonymous);
- **access time;**
- **HTTP request in 3 parts:**
  - **method** (GET, POST, HEAD, ...)
  - **path** of the resource on the Web server (identifying the URL);
  - the **protocol** used for the transmission (HTTP/1.0,HTTP/1.1);
- the **status code** returned by the server as response (200 for OK, 404 for not found);
- the **number of bytes** transmitted.
- **referrer** or referring resource, which contains the URL of the document issuing the link to the current requested page.
- **user agent,** which identify the client application used to retrieve the resource

Client side data

As mentioned earlier, clients store some information in local files (cached information). When a user visits a website, the browser downloads all the necessary resources (text and media) and stores a temporary version locally.

An internal caching mechanism was built in early days to speed up the communication between the client and the server, therefore, for example, if the user returns to the same page later, unless there is a discrepancy in the file details (i.e. creation date), the browser displays the local files rather than downloading the files again from the server. With broadband largely available this is not much of a issue anymore, with rich multimedia content able to be streamed directly from the browsers, however, the function is still very useful for mobile applications as transactions are usually charged according to the amount of data transferred. Another piece of information stored by the clients is the history of the recent websites and some site-specific information. These are stored in *cookies*, which are parcels of text sent by a server to a web browser and then sent back unchanged by the browser each time it accesses that server. HTTP cookies are used for authenticating, tracking, and maintaining specific information about users, such as site preferences, which are useful to customize or personalize user experience.

To create cookies some scripting component is embedded within the hypertext document. This was also the case of the tracking system adopted for the STEER project (more details in Hardy & al 2006).

HTTP specifications

In this section some useful technical definitions are explained

Hypertext Transfer Protocol (HTTP) is an Application layer protocol. Basically it is a method used to transfer or convey information on the World Wide Web. Its original purpose was to provide a way to publish and retrieve HTML pages.
Development of HTTP was coordinated by the World Wide Web Consortium and the Internet Engineering Task Force, culminating in the publication of a series of RFCs, most notably RFC 2616 (1999), which defines HTTP/1.1, the version of HTTP in common use today.

HTTP is a request/response protocol between clients and servers. The originating client, such as a web browser, spider, or other end-user tool, is referred to as the user agent. The destination server, which stores or creates resources such as HTML files and images, is called the origin server. In between the user agent and origin server may be several intermediaries, such as proxies, gateways, and tunnels.

A HTTP client initiates a request by establishing a Transmission Control Protocol (TCP) connection to a particular port on a remote host (port 80 by default). A HTTP server listening on that port waits for the client to send a request message.

Upon receiving the request, the server sends back a status line, such as "HTTP/1.1 200 OK", and a message of its own, the body of which is perhaps the requested file, an error message, or some other information.
Resources to be accessed by HTTP are identified using Uniform Resource Identifiers (URIs) (or, more specifically, URLs) using the http: or https URI schemes.

Request message
The request message consists of the following:

- Request line, such as GET /images/logo.gif HTTP/1.1, which requests the file logo.gif from the /images directory
- Headers, such as Accept-Language: en
- An empty line
- An optional message body
- The request line and headers must all end with CRLF (i.e. a carriage return followed by a line feed). The empty line must consist of only CRLF and no other white space.

In the HTTP/1.1 protocol, all headers except Host are optional.

Request methods
HTTP defines eight methods (sometimes referred to as "verbs") indicating the desired action to be performed on the identified resource: head, get, post, put, delete, trace, options and connect.
HTTP servers are supposed to implement at least GET and HEAD methods and, whenever possible, also the OPTIONS method.

HEAD
Asks for the response identical to the one that would correspond to a GET request, but without the response body. This is useful for retrieving meta-information written in response headers, without having to transport the entire content.

GET
Requests a representation of the specified resource. By far the most common method used on the Web today. Should not be used for operations that cause side-effects (using it for actions in web applications is a common misuse).

POST

Submits data to be processed (e.g. from an HTML form) to the identified resource. The data is included in the body of the request. This may result in the creation of a new resource or the updates of existing resources or both.

Status codes

In HTTP/1.0 and since, the first line of the HTTP response is called the status line and includes a numeric status code (such as "404") and a textual reason phrase (such as "Not Found"). The way the user agent handles the response primarily depends on the code and secondarily on the response headers. Custom status codes can be used since, if the user agent encounters a code it does not recognize, it can use the first digit of the code to determine the general class of the response.

The standard also allows the user agent to attempt to interpret the reason phrase, though this might be unwise since the standard explicitly specifies that status codes are machine-readable and reason phrases are human-readable.

Common HTTP error codes table can be found elsewhere:

- http://support.microsoft.com/?id=318380
- http://en.wikipedia.org/wiki/List_of_HTTP_status_codes

Error analysis could have been possible only for the logs before 2006, but we did not pursue this further because the focus was on the user rather than the evaluation of the system.

Some examples message codes are:
200 OK
400 Bad Request
401 Unauthorized
404 Not Found
500 Internal Server Error

Definition of activity from data

W3C (World Wide Web consortium) has defined several data abstractions for Web usage. These can be used as a suitable data model for performing usage analysis, however data needs to be modified in a suitable format to be processed for data mining.

A **user** is defined as a single individual that is accessing files from one or more Web servers through a browser. While this definition seems trivial, in practice it can be very difficult to uniquely and repeatedly identify users unless a specific authentication mechanism is used. Even with the use of cookies, a user may access the Web through different machines, or use more than one browser on a single machine. Also, several users may use the same machine and browser. In our case identification was not an issue since the introduction of WebCT vista 6.
However the earlier logs (before 2005/06) presented some problems common to standard logs and are addressed in the cleaning data section

A **page view** consists of the set of files that contribute to the display on a user's browser at one time. Page views are usually associated with a single user action (such as mouse-clicks) and can consist of several files such as frames, graphics, and scripts. In the consideration of user behavior, it is really the aggregate page view which is important.
The user does not explicitly ask for "n" frames and "k" graphics to be loaded into her browser, the user requests a complete "Web page", and the composition is rendered on the browser from a number of resources, being different files or data streams from a database. In this context in which WebCT is serving data from a database, there are multiple requests to the web server to compose the page, which makes it particularly hard to identify a web page for the logs before 2005/06.

A **click-stream** is a sequential series of page view requests. Again, the data available from the server side does not always provide enough information to reconstruct the full click-stream for a site. Any page view accessed through a client or proxy-level cache will not be "visible" from the server side. Fortunately this was not the case for our data.

A **user session** is the click-stream of page views for a single user across the entire Web. Typically, unless client-side data collection is available, only the portion of each user session that is accessing a specific site can be used for analysis, since access information is not publicly available from the vast majority of Web servers.

The set of page views in a user session for a particular Web site is referred to as a **server session** (also commonly referred to as a **visit**). A set of server sessions is the necessary input for any Web Usage analysis or data mining tool.

The end of a server session is defined as the point when the user terminates activity or the session times out. The *real* termination, however, is impossible to track reliably with server-side data since the "next-click" at a different Web site is not logged. Also, there is no way of knowing when a user exits a browser when tracking from the proxy or server levels, therefore the last action is always excluded in the computation of durations.

Any semantically meaningful subset of a user or server session is referred to as an **episode** by the W3C.

Term	**Definition**
Browser	Client-side software that is responsible for displaying Page Views and making HTTP requests to a Web Server.
Web Server	Server-side software that is responsible for handling incoming HTTP requests.
Content Server	Server-side software that is responsible for serving Page File in response to requests.
User	Single individual that is accessing files from one or more Web servers through a Browser.
Page File	Files or data from DB that is served through HTTP protocol to a User.
Page View	Set of Page Files that contribute to a single display in a Web Browser.
Click stream	Sequential series of Page View requests
Server Session (visit)	Set of page views served due to a series of HTTP requests from a single User to a single Web Server.
Episode	Subset of page views from a single User or Server Session.

Table 1. Frequently Used Definitions and Abstractions.

Basic operations and tools used

Two different procedures were implemented for the logs from different systems.
Logs before 2006 are the same as logs in the CLF standards (shown in figure 1 and in the body of the thesis.

Logs after the 2006 were a reduced version of the CLF with specific reference to containers rather than all the resources making much easier to *interpret* content.

The top frame in the figure below shows an example from the old format and the bottom frame one from the new logs.

Although the core structure information is similar, the difference in the type and amount of information conveyed is substantial, particularly in the definition of the content.
By looking at the bottom frame it is quite simple to identify the activities: login and organizer viewed. In the top frame similar information needs to be abstracted/deduced from the URI and might lead to an inaccurate definition of the resources.

In this particular example the users displays a simple sequence of two actions: login and view announcement.

```
87.82.10.84 s0123456    s0123456    27/04/2006  08:46:37    +0100
      GET   /SCRIPT/PS0001_8/scripts/serve_home HTTP/1.1    200
      45623
      https://www.webct.ed.ac.uk/webct/public/autosignon?IMS%20id=
s0233198&Time%20Stamp=1146123991&URL=https://www.webct.ed.ac.uk:44
3/SCRIPT/PS0001_8/scripts/serve_home&MAC=15e97d680d9236475ae22fbcc
6024471     Mozilla/5.0 (Windows; U; Windows NT 5.1; en-GB;
rv:1.7.12) Gecko/20050919 Firefox/1.0.7   1

87.82.10.84 s0123456    s0123456    27/04/2006  11:07:18    +0100
      GET   /PS0001_8/announce.htm  HTTP/1.1    304   -
      https://www.webct.ed.ac.uk/PS0001_8/announce.htm
      Mozilla/5.0 (Windows; U; Windows NT 5.1; en-GB; rv:1.7.12)
Gecko/20050919 Firefox/1.0.7  1
```

```
zs_PS0001_9_epack,970766018021,lorenzov,970766017021,2006-11-09
15:02:56.000,1163080976000,logout,login,The University of
Edinburgh,0
zs_PS0001_9_epack,1228278932011,lorenzov,1228278931011,2007-01-26
12:34:26.000,1169811266000,organizer-viewed,organizer,Course
Content Home,6
```

Once a master list of links was available, unique Ids were assigned to each page and copied into a MySQL database. Log files were also transferred in the database for further processing. Page names (long Uris in the path and referrer fields) were substituted with unique Ids of each Uri reducing drastically the size of the files.

Sessionising the logs was the following steps. To do this, a list of unique users was compiled and ‘students users’ were separated from ‘staff users’.
Sessions were then constructed based on the timestamp, the user ID and the IP (client ID).

Click streams were then generated building the sequences of pages, tools, actions and abstractions into strings. These sequences were also copied and filtered showing only the unique instances (i.e. one sequence could have the following tools [login, content, content, discussion, discussion] which is 5 pages, but in reality only 3 unique tools were accessed in this session).

The final steps were to derive a characterisation of the sessions and the users, both based on the metrics explained in chapters 5-8.

References

Web Characterization Terminology & Definitions Sheet. http://www.w3.org/1999/05/WCA-terms/ . W3C Working Draft 24-May-1999. 1999.

Berendt, B., Mobasher, B., Spiliopoulou, M., & Wiltshire, J. (2001). Measuring the accuracy of sessionizers for web usage analysis. Proceedings of the Web Mining Workshop at the First SIAM International Conference on Data Mining, 7–14.
Robert Cooley, Bamshad Mobasher, Jaideep Srivastava (1997). *Grouping Web page references into transactions for mining World Wide Web browsing patterns*.
Robert Cooley, Bamshad Mobasher, Jaideep Srivastava (1999). *Data preparation for mining World Wide Web browsing patterns*.

Appendix 2. Clustering algorithms

- K-means clustering (used with: OmniViz, WEKA and PASW17)
- Two-steps clustering (PASW17)

NOTE: this document is heavily drawn from the PASW guide to algorithms, particularly for the Two-Steps algorithm which is proprietary.

K-means clustering

This procedure is a well know technique which attempts to identify relatively homogeneous groups of cases based on selected characteristics. The algorithm requires the user to specify the number of clusters.
In the various software used, both WEKA and PASW require this to be specified, OmniViz has a proprietary algorithm to infer the number of clusters based on the semantic ‘closeness’ of the text field used.

Two methods for classifying cases are implemented in PASW: either updating cluster centers iteratively or classifying only. Analysis of variance F statistics can be computed, but these statistics are opportunistic (the procedure tries to form groups that do differ), the relative size of the statistics provides information about each variable's contribution to the separation of the groups.
The first iteration of the algorithm involves 3 steps:

- Step 1: Select Initial Cluster Centers
- Step 2: Update Initial Cluster Centers
- Step 3: Assign Cases to the Nearest Cluster

Data considerations.

- Variables should be quantitative at the interval or ratio level.
- No missing values are allowed. Cases with missing values are deleted on a listwise basis.
- Order of variables and scale (consider standardizing) used are important and affect the solution
- The Euclidean distance measure is used for computations

Procedure.
The default algorithm for choosing initial cluster centers is not invariant to case ordering, therefore it is recommended that several different solutions with cases sorted in different random orders are obtained to verify the stability of a given solution.
Scaling of variables is an important consideration. If variables are measured on different scales (for example, one variable is expressed in dollars and another variable is expressed in years), the results may be misleading. In such cases, it should be considered standardizing the variables before the k-means cluster analysis is performed.

1. Select Initial Cluster Centers.
To select the initial cluster centers, a single pass of the data is made. The values of the first NC cases with no missing values are assigned as cluster centers, then the remaining cases are processed as follows:

If minid(xk,Mi) >dmn and d(xk,Mm)>d(xk,Mn), then xk replaces Mn.
If minid(xk,Mi)>dmn and d(xk,Mm)<d(xk,Mn), then xk replaces Mm.

That is, if the distance between xk and its closest cluster mean is greater than the distance between the two closest means (Mm and Mn), then xk replaces either Mm or Mn, whichever is closer to xk.

If xk does not replace a cluster mean in (a), a second test is made:

Let Mq be the closest cluster mean to xk.
Let Mp be the second closest cluster mean to xk.
If d(xk,Mp)>minid(Mq,Mi), then Mq=xk;

That is, if xk is further from the second closest cluster's center than the closest cluster's center is from any other cluster's center, replace the closest cluster's center with xk.

2. Update Initial Cluster Centers
Starting with the first case, each case in turn is assigned to the nearest cluster, and that cluster mean is updated. Note that the initial cluster center is included in this mean. The updated cluster means are the classification cluster centers.
Note that if NOUPDATE is specified in the options, this step is skipped.

3. Assign Cases to the Nearest Cluster
The third pass through the data assigns each case to the nearest cluster, where distance from a cluster is the Euclidean distance between that case and the (updated) classification centers. Final cluster means are then calculated as the average values of clustering variables for cases assigned to each cluster. Final cluster means do not contain classification centers.

When the number of iterations is greater than one, the final cluster means in step 3 are set to the classification cluster means in the end of step 2, and the algorithm repeats step 3 again. The algorithm stops when either the maximum number of iterations is reached or the maximum change of cluster centers in two successive iterations is smaller than ε times the minimum distance among the initial cluster centers.

References

Hartigan, J. A. (1975). *Clustering algorithms*. New York: John Wiley and Sons.

Two-steps clustering

This algorithm is an extension of the well known hierarchical clustering. This refers to a process by which groups are recursively merged, until at the end of the process only one cluster remains containing all records.

The TwoStep cluster method is a scalable cluster analysis algorithm designed to handle very large data sets. It requires only one data pass and the procedure has two steps:

1) pre-cluster the cases (or records) into many small sub-clusters;
2) cluster the sub-clusters resulting from pre-cluster step into the desired number of clusters.

It can also automatically select the number of clusters (used in our application).
Data considerations.

- The procedure accepts both continuous and categorical variables.
- No missing values are allowed. Cases with missing values are deleted on a list wise basis.
- Outliers (data records that do not fit well into any cluster): data records in a leaf entry as outliers if the number of records in the entry is less than a certain fraction (25% by default) of the size of the largest leaf entry in the CF tree.
- The log-likelihood distance measure is used when both continuous and categorical variables are present. It is a probability based distance. The distance between two clusters is related to the decrease in log-likelihood as they are combined into one cluster.

Procedure.
The **pre-cluster step** is implemented by constructing a modified cluster feature (CF) tree. The CF tree consists of levels of nodes, and each node contains a number of entries. A leaf entry (an entry in the leaf node) represents a final sub-cluster. The non-leaf nodes and their entries are used to guide a new record quickly into a correct leaf node. Each entry is characterized by its CF that consists of the entry's number of records, mean and variance of each range field, and counts for each category of each symbolic field.

For each successive record, starting from the root node, it is recursively guided by the closest entry in the node to find the closest child node, and descends along the CF tree. Upon reaching a leaf node, it finds the closest leaf entry in the leaf node. If the record is within a threshold distance of the closest leaf entry, it is absorbed into the leaf entry and the CF of that leaf entry is updated. Otherwise it starts its own leaf entry in the leaf node. If there is no space in the leaf node to create a new leaf entry, the leaf node is split into two. The entries in the original leaf node are divided into two groups using the farthest pair as seeds, and redistributing the remaining entries based on the closeness criterion.

All records falling in the same entry can be collectively represented by the entry's CF. When a new record is added to an entry, the new CF can be computed from this new record and the old CF without knowing the individual records in the entry. These properties of CF make it possible to maintain only the entry CFs, rather than the sets of individual records. Hence the CF-tree is much smaller than the original data and can be stored in memory more efficiently.

The **cluster step** takes sub-clusters resulting from the pre-cluster step as input and then groups them into the desired number of clusters. After merging, the new set of clusters is compared, the closest pair is merged, and the process repeats until all clusters have been merged. (If one is familiar with the way a decision tree is built, this is a similar process, except in reverse.)

Because the clusters are merged recursively in this way, it is easy to compare solutions with different numbers of clusters. To get a five-cluster solution, simply stop merging when there are five clusters left; to get a four-cluster solution, take the five-cluster solution and perform one more merge operation, and so on.

Hierarchical clustering has proved to be a very successful technique in a number of disciplines, and it has been used extensively in market research. However, like the k-means clustering, the limitation is that data needs to be standardised and mixed types attributes are not accepted (unless the ordinal categories are transformed) The two-steps algorithm, therefore offers the opportunity to solve a number of these problems, considering mixed types in a two-steps computation. Bacher and colleagues (2004) conducted a thorough evaluation of this algorithm and whilst they determined that the algorithm performs well on simulated data with continuous variables, with categorical variables the algorithm might lead to misclassification.

This should be carefully considered and might be one of the key reasons for incorrect decisions. In the use of this algorithm we were careful in entering categorical data only to define gender and country of origin (where desirable), minimizing the potential issues indicated by Bacher et al.

References

Bacher, J., Wenzig, K., & Vogler, M. (2004). SPSS TwoStep Cluster–A First Evaluation. Arbeits-und Diskussionspapiere, 2.

Chiu, T., D. Fang, J. Chen, Y. Wang, and C. Jeris. (2001). *A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment.* In: Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining. San Francisco, CA: ACM.

Zhang, T., R. Ramakrishnon, and M. Livny. (1996). *BIRCH: An efficient data clustering method for very large databases.* In: Proceedings of the ACM SIGMOD Conference on ement of Data. Montreal, Canada: ACM.

Appendix 3. Study materials and surveys

Here are some example questions from the different instruments.

ASSIST

- There's not much of the work here that I find interesting or relevant
- When I'm working on a new topic, I try to see in my own mind how all the ideas fit together
- I concentrate on learning just those bits of information I have to know to pass
- I'm pretty good at getting down to work whenever I need to

The Inventory can be downloaded from the ETL project web page:
http://www.etl.tla.ed.ac.uk/

TSI

- When faced with a problem, I make sure my way of doing it is approved by my peers
- I like to check and rate opposing points of view or conflicting ideas
- In dealing with difficulties, I have a good sense of how important each of them is and what order to tackle them in
- I like to follow definite rules or directions when solving a problem or doing a task

The original TSI is included in Sternberg's Thinking Styles book, but more recent versions require approval from the author

CSI

- I am most effective when my work involves a clear sequence of tasks to be performed.
- The kind of work I like best is that which requires a logical, step-by-step approach.

- I am inclined to scan through reports rather than read them in detail.
- I prefer chaotic action to orderly inaction.

The CSI is not publicly available, but can be requested from Dr. Chris Allinson for academic research: c.w.allinson@lubs.leeds.ac.uk

VICS-WA

Screenshots of the items could not be included as protected by copyright

The VICAS-WA is a MS Windows executable which is subject to copyright. Information on the latest version can be obtained directly from Dr. Elizabeth Peterson:
e.peterson@auckland.ac.nz

Appendix 4. WET (Web exploration tool)

The Website Exploration Tool (WET) (Figure 1) is a visual tool to support Web Mining. Its main characteristic is its ability to combine the visualisation of data based on the structure, content and usage of a website.

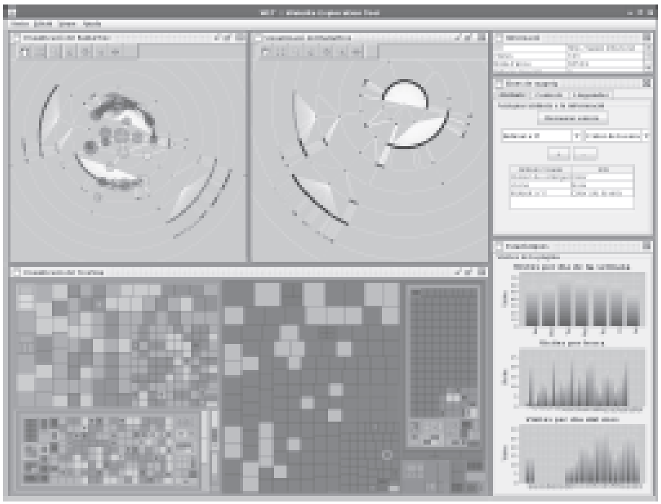


Figure 1: A screenshot of WET.

WET automatically extracts descriptive hierarchies of a website enabling the visualisation of its structure and users' navigation along with its contents and usage. To achieve this, a set of web metrics (e.g. number of hits per page, page rank or landing pages, among others) can be mapped into the visual attributes of glyphs that form a representative hierarchical visual metaphor of the site. In addition, all the available visualisations may be combined through a linking and brushing system that enhances the multi-metaphor user experience.

Procedure.

From the server log 6 steps are necessary in order to prepare the data for WET:

1. Transfer of the logs into a MySQL database;
2. Clean and sessionize the log
3. Extraction of unique URIs and link pairs
4. Enrich the data with topological and semantic information
5. Select sub-samples of users
6. Convert the database into graph xml for WET

I contributed particularly in steps 4 and 5 by providing a new interpretation overlay for the data in our context (see Pascual-Cid et al. 2009 attached)

WET has been developed by Victor Pacual-Cid. He is supervised by Juan Carlos Dürsteler from Universitat Pompeu Fabra, Barcelona and Ricardo Baeza-Yates from Yahoo" Research, Barcelona.
His work is partially supported by the Grant TIN2006- 15536-C02-01 of the Ministry of Science and Innovation of Spain

References.

Pascual-Cid,V., Vigentini, L. And Qixcal, M. (2009) New Techniques for Visualising Web Navigational Data. IV09: Proceedings of the 13th International Conference Information Visualisation (2009) pp. 621 -626

Pascual-Cid,V. (2008) An information visualisation system for the understanding of web data. Visual Analytics Science and Technology, 2008. VAST '08. IEEE Symposium on (2008) pp. 183 -184