The Effect of Prior Knowledge in Undergraduate Performance in Chemistry: A Correlation – Prediction Study

Michael K. Seery

Technological University Dublin, michael.seery@tudublin.ie

The Effect of Prior Knowledge in Undergraduate Performance in Chemistry: A Correlation – Prediction Study

Michael K Seery BA PhD

Thesis submitted to fulfil requirements of

MA (3^{rd} Level Learning and Teaching)

at

Dublin Institute of Technology

Declaration

I certify that this thesis which I now submit for examination for the award of MA (3rd Level Learning and Teaching), is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work.

This thesis was prepared according to the regulations of the Dublin Institute of Technology and has not been submitted in whole or in part for another award in any Institute.

The work reported on in this thesis conforms to the principles and requirements of the Institute's guidelines for ethics in research.

The Institute has permission to keep, lend or copy this thesis in whole or in part, on condition that any such use of the material of the thesis be duly acknowledged.

____

Michael K Seery

July 2009

While the individual man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty. You can, for example, never foretell what any one man will be up to, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician.

Sir Arthur Conan Doyle

Acknowledgements

This finishing of this thesis is due to a lot of support from colleagues and friends. My thanks go to the following;

My supervisor Brian Bowe for his support in the development of the work and discussion of results and the MA coordinator Marian Fitzmaurice for ongoing support and encouragement, timely advice and excellent proof-reading;

Mark Russell, whose thesis offered some guidance and structure on quantitative work in education and Joe Condon from Statistical Advice Unit, DIT, for telling me things I should have known and lots of things I never would have;

Diana Mitchell and all the DIT library staff for their help in sourcing materials;

David Treagust and his colleagues for providing the Diagnostic Assessment tests;

My academic colleagues, particularly Claire and Christine, for ongoing encouragement and support.

Sincere thanks to you all for your support. Special words of thanks go to Niall for putting up with it all.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Leaving Certificate examination is the final exam taken by students at second level, and intends to award students credit for their studies at second level. With increasingly levels of participation in post-secondary education, it is now primarily viewed as a stepping stone to further (FE) and higher (HE) education. For the HE sector, the Leaving Certificate is used by colleges to allocate places in a process administered by the Central Admissions Office (CAO). Students are allocated a maximum of 600 CAO points on the basis of their performance in six subjects in the Leaving Certificate examination. Points for courses such as medicine and health-related courses, law and certain arts courses have traditionally being high, with students requiring five to six A grades in their Leaving Certificate subjects, amounting to 550 – 600 points. Points for science-based courses have steadily decreased over the last decade, from a range of 400 – 500 to a range of 300 – 400, with an average of 320 in 2008 (Childs, 2008). This is due to several factors. The CAO points requirement for a programme is ultimately decided by interest in places and number of places available. There has been a steady to declining interest in science in Ireland in the last 10 – 15 years as students have opted for information technology courses in the late 1990's, and then courses related with building and economic boom – architecture, construction related courses, business courses – in the decade from the millennium. The Irish government, through programmes such as Discover Science and Engineering, has placed significant resources in promoting science at primary and secondary level and there are signs that the decline in interest is beginning to level off. The second reason for decline in science is the explosion in the number of science-related courses in the third level sector. Tertiary level courses in science were traditionally (pre 1997) delivered by the Universities with applied courses in science being delivered by the Dublin Institute of Technology (DIT). Latterly, Dublin City University and University of Limerick were accorded University status, and the country's Regional Technical Colleges were reformed into Institutes of Technology. These, along with existing Universities and DIT offer science courses, with the Institutes of Technology and DIT[1] increasingly offering courses at level 8 (honours degree level). In addition, larger universities offer high number of places (> 300) on their first year courses. The result is a low demand for a large number of courses. In relation to chemistry courses in particular, the Leaving Certificate subject Chemistry is taken by approximately 10 - 15% of the Leaving Certificate cohort each year. As such, requirement of precursor knowledge for chemistry based degrees is not feasible, due to the limited pool of applicants and the high number of places discussed above.

In this study, students are completing a course of study which is an honours degree in analytical chemistry. Approximately half of the intake have studied Chemistry at Leaving Certificate level. The first year programme therefore starts the subject *ab initio* and delivers a 15 European credit transfer system (ECTS) module in *Introductory Chemistry* to bring the level of chemistry knowledge to slightly above Leaving Certificate level (approximately equal to A-level chemistry), so that all students entering year 2 are at an equal knowledge base in chemistry.

A question that arises is whether students who *have not* completed Chemistry at Leaving Certificate level are at a distinct disadvantage to those who have, both in terms of their Year 1 performance and their performance subsequently in years 2 to 4. In this study, it is planned to examine whether there is a correlation between prior knowledge and level of chemistry with grades achieved in Years 1 – 4. In doing so, factors such as CAO entry points, distance from college, gender and perceived motivation and interest in the course will be examined and incorporated into the correlation.

A motivation for this work is to examine whether the year 1 programme allows equalization of the chemistry knowledge across the group, or if more assistance is required for students who have no prior knowledge of chemistry. To complete this, a correlation between CAO points and annual grades in Years 1 – 4 will be examined.

There is extensive literature on the role of prior learning, but a limited amount in the area of chemistry, and none on the particular Irish context. However, there are interesting questions in the Irish context. The primary research question is:

[1] DIT is usually referred to separately from the group of Institutes of Technology for historical and operational reasons.

"*How does students' prior chemistry knowledge influence their performance in year 1 chemistry?*"

For the purposes of this study, prior knowledge is measured by performance at Leaving Certificate level, in that the performance here closely matches the expectations in the first year at degree level.

Additional sub-questions of interest are as follows:

(i) Is there a correlation between CAO entry points and Year 1 performance?

(ii) How does students' performance in end of year exams in subsequent years differ based on their level of prior knowledge on entry?

(iii) Do other variables exist that impact year 1 performance and if so to what extent do they predict year 1 grades. Variables such as laboratory performance, gender, semester tests and commuting distance will be examined.

(iv) Does the level of understanding of basic chemistry concepts, as measured by conceptual testing, correlate to prior knowledge.

The literature as discussed in the next chapter brings together interesting issues for this work which may confound the study because of the context in which it lies. Given that the *prior knowledge* will be based on students' Leaving Certificate performance, the question arises as to what is the *nature* of this prior knowledge. Leaving Certificate Chemistry has considerable acknowledged simplifications which may enforce misconceptions in students' understanding, which they can carry through to college level. Additionally, there is an emphasis on particular areas in the Leaving Certificate syllabus that would not carry through to the college syllabus. This may initially challenge students' self-beliefs, especially if they arrive at college in a situation where they are perceived to know more chemistry than some of their peers.

The study comparing aptitude and prior knowledge can also be similarly examined by considering CAO entry points as a level of student study aptitude. In this case, comparison of students who have a high CAO score but no prior learning with those who have prior learning will be interesting.

Therefore, in light of all of these factors, it is acknowledged that this study is limited in scope, but the question examining a correlation is a valuable one, and will be the precedent to the other issues and questions discussed.

Finally, all of the present work takes place in the context of the student undergoing a huge transition in their own lives. They are leaving school and entering a new world where independence is expected and maturity is assumed. In a traditional discipline like chemistry, taking stock of the student experience, as measured in this case by their performance in college, is I believe a worthwhile exercise.

Chapter 2

Literature Review

2.1 Introduction

Students attending third level institutions face many new and exciting experiences and challenges and possibly move away from home for the first time. As such, their performance at third level – especially in the earlier stages – may be attributed, to and influenced by, any of several factors. These include how well a student adjusts and settles into college routine, what their motivation is on the course they are studying, what their general level of study aptitude is, how well their learning styles adjust to college education system, what teaching methods are employed, *etc.* This research aims to study an additional factor – the effect of students' prior knowledge on their performance in first year as well as subsequent years – and survey to what extent prior knowledge can predict year 1 performance over a range of other measurable factors.

Two important factors relevant to the study are considered below. Firstly, the nature and role of prior knowledge is examined by surveying how previous researchers have defined and assessed prior knowledge. A theoretical perspective for prior knowledge based on the literature is outlined.

Secondly, informative case studies on the assessment and use of prior knowledge in predicting academic performance, both in a general context and in a context specific to chemistry are presented. Some of the latter studies provide a useful template for the methodological approach used in the current study. This leads into an examination of prior knowledge at the specific boundary of the school-university transition in chemistry.

This boundary is a unique experience for students, as they leave the school system and enter the tertiary education system, and the analysis must be based in this context. The first year transition has been called a "betwixt space" (Palmer, O'Kane, & Owens, 2009), where students adapt from the school/home life to the university one. This adaption involves turning point experiences – both positive (new learning experiences, independence) and negative (leaving family life, isolation) – and the students ability to cope with these.

2.2 Prior Knowledge as a Concept: Theoretical Framework

2.2.1 Introduction

The concept of prior knowledge and the underlying theoretical framework in cognitivism is summarised below. This summary is primarily based on Dochy's reviews of the topic (Dochy, De Ridjtt, & Dyck, 2002; Dochy, Segers, & Buehl, 1999) as well as Bloom's original work (Bloom, 1976).

2.2.2 Origins

The origins of prior knowledge as a theoretical framework can be sourced in the work of Bloom in the 1970's (Bloom, 1976) who was interested in the extent that human characteristics such as intelligence and motivation could be influenced by experience (Bloom, 1964; Education-Encyclopedia, 2009). Bloom discussed the concept of 'cognitive entry behaviour', (a term he borrowed from the work of Glaser (Dochy et al., 2002)) which he determined to account for more than half (r = 0.7, r^2 = 0.49)[2] of the variance in cognitive achievement in subsequent learning tasks (Bloom, 1976). According to Dochy, it wasn't until the late 1980's and early 1990's that researchers began to define and study 'cognitive entry behaviour' with work by Alexander (Alexander & Judy, 1988) and Dochy himself (Dochy, 1992, 1994). The terms used varied widely and as discussed later, definitions were not consistent, but the term 'prior knowledge' is now favoured by Dochy and most other modern studies surveyed (*vide infra*). Prior knowledge is distinguished from aptitude, which takes into account motivation, learning styles and individual characteristics. Finally, prior knowledge itself can be

[2] r is a correlation factor, a measure of how strongly two variables are related. The square of this value, r^2, is a measure of how much variance in the value is predicted by the term under consideration. Detailed explanations of these terms are provided in Chapter 3.

sub-divided into many forms, declarative – procedural – conditional knowledge and domain specific – domain transcending knowledge to name two favoured sub-divisions (Dochy et al., 2002).

2.2.3 Development of Bloom's Concept

According to Dochy, Bloom presented convincing arguments to support his central thesis that what he termed cognitive entry behaviours were crucial to learning, with "…an overview of research that only lunatics would doubt. At least at that time…" (Dochy et al., 2002). A multitude of subsequent papers on cognition (Alexander & Judy, 1988; Dochy, 1992, 1994; Dochy et al., 1999) served to demonstrate that prior knowledge was the most significant element in learning. An important conclusion to this work, was given by Glaser and De Corte in Dochy (Dochy, 1992; Dochy et al., 2002):

> *'Indeed, new learning is exceedingly difficult when prior informal as well as formal knowledge is not used as a springboard for future learning. It has also become more and more obvious, that in contrast to the traditional measures of aptitude, the assessment of prior knowledge and skill is not only a much more precise predictor of learning, but provides in addition a more useful basis for instruction and guidance'*

The latter point, that a tutor's knowledge of the lack of or misconceptions in prior knowledge can be used effectively in teaching strategy is a recurrent theme in the literature on prior knowledge and a useful practical outcome of an understanding of the role of prior knowledge.

2.2.4 Modelling the Role of Prior Knowledge

Dochy reviews a range of causal modelling techniques to describe the concept of prior knowledge which are reproduced here (Dochy et al., 2002). The purpose here is to examine how prior knowledge is incorporated into each of the models presented and the predominance of prior knowledge in determining achievement, as well as the lessons that can be drawn from an understanding of prior knowledge. More detailed analysis may be found in Dochy's review or the original papers.

Figure 1 shows a model which describes the various factors resulting in achievement. The numbers indicate the degree of correlation (*r*) between the factors, essentially a degree of association. Note especially the large correlation between prior knowledge and achievement.

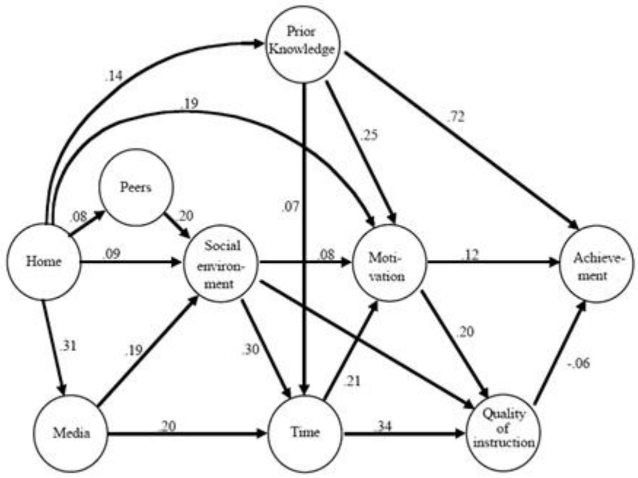


Figure 1: A 'complex causal model' on the factors leading to achievement, presented in Dochy (Dochy et al., 2002) based on work completed by Parkinson

Another more schematic representation is shown in Figure 2. This illustrates two important points, originally developed by Dochy and Alexander (Dochy & Alexander, 1995). The first is that it demonstrates how prior knowledge affects learning – by means of an overall facilitatory effect; by means of the inherent qualities of prior knowledge enhancing this facilitation and by means of interaction between impact of prior knowledge and this facilitation (Dochy et al., 2002). Secondly, it describes how this understanding can influence better teaching in the classroom. The facilitation effects may be direct (*i.e.* prior knowledge leads to better results) or indirect (by means of optimising clarity of study materials and by way of optimising instruction and study time).

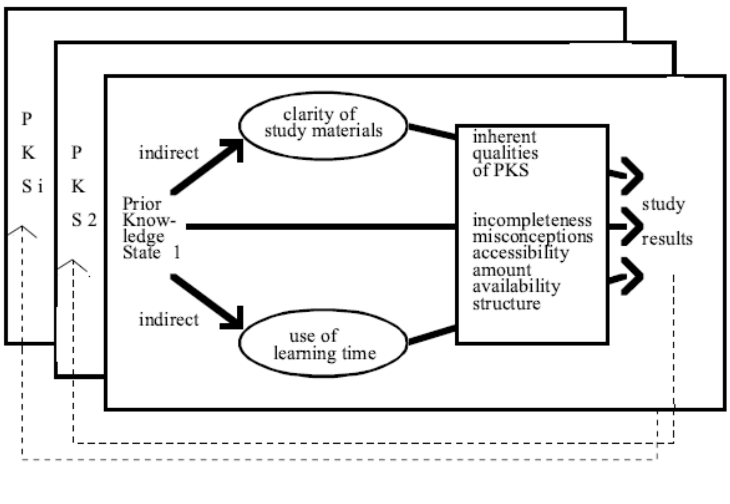


Figure 2: Interaction of 'inherent qualities' of prior knowledge and its 'facilitating effect', based on work by Dochy presented in Dochy (Dochy et al., 2002)

2.3 Role of Prior Knowledge

Prior knowledge and its impact on student performance is a subject that has been widely studied in the education literature. This review concentrates on some of the important aspects to be considered when discussing prior knowledge, namely:

- The definition of prior knowledge and how it is assessed
- The role of prior knowledge in subsequent academic performance
- The use of prior knowledge as a predictor in student achievement

2.3.1 Current Definition of Prior Knowledge

As mentioned above, the meaning of the term prior knowledge varies over the literature surveyed and over the development of the concept of prior knowledge from Bloom's original work. At a simple level it can be considered to be the "knowledge, skills, or ability that students bring to the learning process" (Jonassen & Grabowski, 1993). Dochy argues that this definition is too vague and proposes that prior knowledge should be defined as "the whole of a person's actual knowledge that (a) is available before a certain learning task, (b) is structured in schemata, (c) is declarative and procedural, (d) is partly explicit and partly tacit, (e) is dynamic in nature and stored in the knowledge base" (Dochy, 1994). Dochy again uses this definition in his 2002 review (Dochy et al., 2002). A more specific form of prior knowledge (and hence one more readily quantified) is domain-specific prior knowledge (also called topic-relevant prior knowledge). This is the level of prior knowledge of a particular area being studied, for example in mathematics or chemistry (Dochy, 1992). This (domain-specific knowledge) can differ widely in terms of the quality and relevance to what is currently being studied (Hailikari, Nevgi, & Komulainen, 2008) (see discussion on the quality and depth of prior knowledge, *vide infra*). The quantification of prior knowledge, based on the definition espoused by Dochy, above, is rarely explicitly stated in studies on prior knowledge and depends of course on the instruments used to measure the level and quality of prior knowledge. In studies surveyed for this work, prior knowledge was most commonly defined and quantified by means of a performance in one of a number of test methods used. Therefore a crucial element in understanding what is meant by prior knowledge is how prior knowledge is assessed.

2.3.2 Assessment of Prior Knowledge

There are several approaches to assessing the level of prior knowledge. In an extensive review of the topic, Dochy and co-workers identified six approaches to the assessment of prior knowledge; multiple choice and recognition tests, association methods, questionnaires, checklists and free recall. The Dochy study argues that depending on the type of assessment used, different amounts/types of information will be elicited and hence a range of assessment methods will give a good overview of prior knowledge (Dochy et al., 1999) (this relates well to the discussion on misconceptions, *vide infra*). A subsequent study by Dochy added two more methods (Dochy et al., 2002) incorporating many of the ideas considered below – use of external testing or other previous discipline or aptitude testing. Prior knowledge assessed by any method is “a snapshot in time” (Dochy et al., 1999) and several authors have argued that prior knowledge should be assessed by a variety of methods to give a more extensive picture of the nature and breadth of prior knowledge (see for example (Hailikari & Lindblom-Ylanne, 2007)).

2.3.3 Role of Prior Knowledge on Performance

2.3.3.1 Overview

Several studies have examined the role of prior knowledge on student performance and the majority of these conclude that prior knowledge of a subject has a positive impact on student learning. Dochy (Dochy et al., 1999) consider 183 articles, including those early seminal studies listed above, which examine the role of prior knowledge. All but 11 studies found that prior knowledge had a positive effect on student learning, although some of these studies which found no effect were probably not valid due to poor methodology (for example students with little difference in prior knowledge were studied or familiarity was used to determine prior knowledge). The authors of this review draw several important conclusions. Prior knowledge does have a positive impact on student learning, a fact that appears universally acknowledged in the literature, where prior knowledge can explain from 30 – 60% of the variance in student performance. Other learning variables (motivation, time on task, quality of instruction) also impact on performance although these are related to prior knowledge. As discussed above, the assessment method used to determine prior knowledge should be varied to examine different elements of prior knowledge.

2.3.3.2 Nature of Prior Knowledge

The authors in the above study identify that misconceptions may hinder the performance of a student and the "accessibility, availability and structure" of the prior knowledge should be measured (Dochy et al., 1999). In examining students' performance as a result of their prior knowledge, the question arises whether their prior knowledge, if wrong, might actually hinder their learning and hence performance, as students with pre-existing prior knowledge may be reluctant to change their mental model of a concept they believe to be true. There is considerable variance in the literature on whether this is indeed the case. In their review, Dochy et al consider this point and surmise that even though students with inaccurate prior knowledge may be at a disadvantage, they still have the advantage over students with no prior knowledge as the latter group do not have relevant knowledge frameworks to validate and structure new information (Dochy et al., 1999). Furthermore, the 2002 review examines eight different approaches to assessing prior knowledge for different stages of learning and conclude that all data agree with the view that prior knowledge positively influences new knowledge acquisition. Some simple, effective examples of this are wide-spread in the literature. Byrnes and Guthrie found that students with a level of prior knowledge were more capable of searching a textbook in search of answers to a question than those who had no prior knowledge, as the framework was already present with which to reference what they were looking for (Byrnes & Guthrie, 1992). Students with prior knowledge were found to be more adept and discerning when note-taking in lectures (Etta-AkinAina, 1988).

However there is a significant body of literature arguing that misconceptions or incorrect prior knowledge may in fact hinder future achievement. These include a review by Alexander and Judy (Alexander & Judy, 1988), as well as more recent work by Thompson and Zamboanga (Thompson & Zamboanga, 2003). The argument is that if prior knowledge is incomplete, it can hinder understanding because students' beliefs in the accuracy of their misconceptions becomes a barrier to greater understanding. They compared their work to similar studies in the area of psychology performance (Taylor & Kowalski, 2004).

As mentioned above, both models of the cognitive role of prior knowledge and practice based research have indicated that a tutor's understanding of prior knowledge, and misconceptions in this knowledge, can have a beneficial impact on practice if that knowledge is used in teaching practice. Treagust has pioneered a number of studies in the study of misconceptions in science disciplines, and these are discussed later in this chapter. A common theme in the literature on misconceptions is that if they can be taken into consideration early in teaching, the use of formative and self-assessment to overcome misconceptions can be a powerful teaching tool (Hailikari & Lindblom-Ylanne, 2007). On a more intricate level, the nature of prior knowledge can vary in terms of quality and depth. Therefore students whose level of domain specific prior knowledge is surface level and who operate at a low cognitive level may not benefit from their prior knowledge (Hailikari & Lindblom-Ylanne, 2007).

2.3.4 Prior Knowledge as a Predictor of Student Achievement in Other Disciplines

There are some excellent recent studies on the use of prior knowledge to predict student achievement, which are summarised in Table 1. As well as providing insight into the nature and role of prior knowledge as discussed above, these studies provide a useful template for the current work in terms of methodological approach, data analysis and conclusions drawn in light of the literature on the topic discussed herein.

A study of both prior knowledge of topic and general academic aptitude was conducted on a group of freshman psychology students in order to examine whether general aptitude has an affect on performance and to examine the prediction power of prior knowledge over other factors (Thompson & Zamboanga, 2004). They cite their own earlier work completed that showed that prior knowledge was a positive and significant predictor of exam performance, even when factors quantifying student achievement were controlled (Thompson & Zamboanga, 2003). This second paper aimed to consider general aptitude. The data gathered is listed below:

- Measures of course achievement by means of taking four multiple choice tests during the semester
- Student ACT scores which are used as means to predict course achievement
- Two pre-tests as indicators of prior knowledge
- Measures of attendance, homework and recitation exam performance as indices of course involvement

The study is in two parts. The first is a correlation table examining the inter-correlation between the different measures gathered (11 in total) to see which has effect on others. The study finds good intercorrelation between exam performance (r = 0.67) and good intercorrelation between ACT and exams (r = 0.5) as well as amongst the various other factors.

The second part of the work is more powerful – it involves examining which of the data can be used to *predict* the end of year achievement. In this case, data is grouped into various blocks and used in a hierarchical linear regression model to predict the course achievement measure. The hierarchical model indicates that each block group is tested in turn, and each subsequent analysis of a block takes into account (or controls for) the previous step in the model. The blocks included background variables (ACT score, year in school and major), prior knowledge (pre-tests, prior psychology coursework) and course specific variables (course participation and involvement). The result was to demonstrate that while several of the factors correlated with exam performance, prior knowledge as examined in the pre-test was a unique predictor of variance in exam scores.

The studies on Mathematics (Table 1) were in two stages – the first was a similar analysis to what was discussed; examining the role of previous factors such as study success and prior knowledge as well as student perceptions. What is interesting about this study is that in the second part of the analysis, differences between procedural and declarative prior knowledge began to evolve. Procedural prior knowledge is where a student can reproduce an approach, and is obviously of crucial importance in a scientific discipline. This involves at the basic level algorithmic problem solving and at the more advanced level synthesis of known procedures to develop strategy to solve a problem. This study found that this type of prior knowledge showed a strong predictory trend with performance, whereas declarative (*i.e.* ability to state facts or recall) prior knowledge did not. These studies again used multi-step regression in their approach, and considered a range of variables in each of their models. The final paper on Accounting (Table 1) takes a different approach but is useful in developing the model for the present research in that it used comparison of means in pre- and post-test scores to address the question of the use of students' prior knowledge in assisting access to new materials, which demonstrates the role of prior knowledge in developing a mental model/framework. These studies which articulate a similar approach to using prior knowledge as a predictor are summarised in Table 1.

Table 1: Examples of Non-Chemistry Studies Surveyed which use Prior Knowledge to Predict Student Performance

Subject Discipline	Data Gathered	Correlational Analysis	Regression Analysis	Reference
Psychology	Course achievement (MCTs), ACT scores, pre-tests, measures of attendance and homework	Good correlation observed between prior course achievement, in-class work and performance	Regression demonstrated that prior knowledge was sole predictor of performance	(Thompson & Zamboanga, 2004)
Mathematics	Previous study success, student expectation of success, self-efficacy, self-perception of ability, prior knowledge tests	Strong correlation and intercorrelation between all components (except self-perception)	Prior knowledge predicted performance over all other variables (55%); academic self-beliefs had a strong influence on prior knowledge	(Hailikari et al., 2008)
Mathematics	Similar to above, distinguished between procedural prior knowledge and declarative prior knowledge	Positive correlation between performance and prior grade; (procedural most strongly intercorrelated)	Regression demonstrated that procedural and not declarative prior knowledge has influence on performance	(Hailikari & Lindblom-Ylanne, 2007)
Accounting	A study which compares pre- and post-test scores to examine role of prior learning	Correlation not used, differences between means of pre- and post-test scores used	Lack of prior knowledge made students ability to access new material difficult, inaccurate prior knowledge hindered learning process	(Addison & Hutcheson, 2001)

2.3.5 Prediction of Chemistry Achievement

2.3.5.1 Predictors of performance (other than prior knowledge)

There is a large body of work on predicting chemistry achievement using a range of factors. Some of the more relevant studies to this work are detailed below. Studies using prior knowledge as a predictor are discussed separately in 2.3.5.2. As early as 1929, Smith and Trimble were examining how to predict students' performance in chemistry based on their "past records". These authors use aptitude tests to predict student performance in chemistry and found a reasonable correlation of scores with their test (not described), but their overall conclusion was that it was generally "possible to predict the performance of the best and of the poorest students" (Smith & Trimble, 1929). More recent work has discussed the correlation of mathematical SAT scores with performance and found that gender, prior college experience and ethnic background were not important factors (Spencer, 1996), the correlation of group assessment of logical thinking to identify at risk students and advise on tutoring and educational aids (Bunce & Hutchinson, 1993) and the correlation of diagnostic tests with performance (A. A. Russell, 1994). These studies are essentially (or completely) correlation studies, despite what some of their titles and abstracts state. Some prediction studies include the use of ACT and GPA scores in predicting achievement of African Americans with regression analysis (r = 0.65) (Carmichael, Bauer, Sevenair, Hunter, & Garnbrel, 1986) and non-cognitive predictors (student attitudes where they rated their academic abilities and expectancies) which demonstrated among other factors that a self-rating of mathematical ability was a significant predictor, and the unusual (in the context of these studies) finding that students' attitudes were a better predictor of grade achievement than ACT scores of the number of years of high school maths (House, 1995). More recently a study in University of South Florida used a test of logical thinking with SAT scores to identify at-risk students in chemistry, and concluded that the process of assisting such students once identified should develop their formal thought processes as well as content review (Lewis & Lewis, 2007). This study differs from others in that it used a universal assessment (American Chemical Society Assessment) to provide a more generalised result.

2.3.5.2 The role of prior knowledge in achievement in chemistry

Treagust has completed much work on the role of prior knowledge in chemistry. In an early study which examined four factors; formal reasoning ability, prior knowledge, field dependence/ independence, and memory capacity on performance in chemistry found that both prior knowledge and formal reasoning ability accounted for significant variance in performance (Chandran, Treagust, & Tobin, 1987). The tests Treagust used in this and subsequent studies have been published and are available for use on request (Chandrasegarana, Treagust, & Mocerino, 2007; Treagust, 2008). These provide standardised methods for testing students' knowledge of basic concepts in chemistry as well as examining the quality of that knowledge, by means of a two-tier assessment process.[3]

Other prediction studies include using of Maths SAT scores, a college entrance exam score and High School chemistry grades to predict performance in year 1 was conducted on a group of allied health students (Craney & Armstrong, 1985). This study found that in contrast to others it refers to, Maths SAT scores was not alone a powerful predictor, and the prediction capability was enhanced when combined with other factors, such as high school chemistry.

In a study on the effectiveness of a bridging course in chemistry for those with no prior knowledge, researchers found that attendance at the course contributed to a better performance in subsequent year 1 university examinations than students who had no prior knowledge, although not as well as students with a strong background in chemistry (Youl, Read, George, Masters, & Schmid, 2005). This fact was attributed to the increased level of prior knowledge and/or improved self-efficacy.

This is similar to findings by Boujaoude and Giuliano (Boujaoude & Giuliano, 1991) who found that prior knowledge is the factor of greatest significance when considering student achievement in chemistry. This study is interesting in that is uses a variety of instruments to assess 199 students in terms of approaches to study, prior knowledge, logical thinking ability, attitude as well as performance in college level chemistry exams. It also considers the effect of gender as well as the effect of prior knowledge (and the observation that the students performed better in "reproducing" than "meaning", logical thinking also contributed to the correlation variance). The study also found that males performed better than females in logical thinking. More recently, a gender-based study found that girls perform better on reading-based skills whereas boys perform better in measure of scientific knowledge (O'Reilly & McNamara, 2007), which is a point worth considering in the Irish context as girls tend to outperform boys in the sciences in Leaving Cert (HEA, 2007).

An interesting study on the impact of a tutorial programme in introductory chemistry was outlined by Braathen and Hewson (Braathen & Hewson, 1988). This case-study, conducted within a constructivist theoretical framework, examined a small groups of students engaged on

[3] Prof. Treagust and his colleagues have kindly provided these tests for use in this study.

a tutorial programme and found that their learning was enhanced if they were positively disposed to learning. Another study from a constructivist framework argued that the teaching philosophy of chemistry is rooted in constructivism, in that students are encouraged to consider (and build) models of sub microscopic representations. Given that chemistry (and indeed science) is constructivist means that prior knowledge is of fundamental importance to scientists (students and experts) as science is a successive building of prior knowledge and experience (Harrison, 2003).

These studies regarding chemistry at the school-university transition consistently refer to and are based upon the notion that it is not only students declarative (or indeed domain specific) knowledge that is of importance, rather that their procedural knowledge is of more value. This is of importance in the current context, as the Leaving Certificate examination format is generally accepted to encourage rote learning rather than procedural, in-depth knowledge. (Whether this is the case for this group of students is examined in this work). Treagust completed a study outlining how an introductory chemistry course emphasised the role of rote learning (Chittleborough, Treagust, & Mocerino, 2002). This study, from the students' perspective, displayed a lack of development of suitable mental models by the method of teaching (because of assessment, teaching speed and clarity, *etc*) which resulted in a knowledge framework being developed, which was scant and compartmentalised. Given the similarities between these introductory courses and the Leaving Certificate Chemistry syllabus (DES, 1999), it could be argued that a similar outcome is likely to be achieved by Leaving Certificate Chemistry students.

2.5 Summary and Impact of the Literature Review on this Study

The literature discussed above provides an overview both to the theoretical basis of prior knowledge and its assessement and effect on future learning and achievement in a general and specific context.

This literature raises interesting issues and indicators for the current study. Dochy urges that the researcher accurately define what they mean by prior knowledge and outline how it is assessed. This is of ultimate importance to the validity of a study. Leading on from the assessment of prior knowledge, issues arise too on the nature of this knowledge, where it exists. Practically all studies concerned with the impact of the study of prior knowledge in the classroom/lecture hall discuss the idea that an understanding of prior knowledge and gaps within this knowledge can actually facilitate the tutor promote a richer learning environment by taking account of gaps and/or misconceptions. Treagust has done much work in this regard, and his work is used as a basis for the present study in examining the quality of prior knowledge. Treagust, too, has completed work on the analysis of an introductory chemistry course, and the means by which it encourages rote learning. The subsequent impact on the quality of prior knowledge has immediate relevance to this study, and is one which will be considered by comparing the nature of school and freshman college education.

However there are gaps and limitations apparent in the work on "predictor papers" (Lewis & Lewis, 2007). Almost all of the work has been completed in the American system, where because of the systematic differences in undergraduate education, students may have different motivations to students in this study, who are on a dedicated chemistry course. Most of the work cited above involved a prediction for an introductory/general chemistry course taken by students on different major courses. Little/no work has been completed in the Irish context in this area, and it is hoped that this study will make a useful contribution to that knowledge gap. At a technical level, much of the research completed focuses on a single predictor, a point made well in an excellent paper examining the effects of a remedial course on chemistry over a six-year period (Bentley & Gellene, 2005). Such studies have collected a single data fact about students in the study (e.g. Maths SAT score) and examined its use as a predictor. More detailed studies have benefited from using a range of different scores and measurements, and that approach is taken here.

Chapter 3

Methodology and Methods

3.1 Introduction

This study takes a quantitative approach to studying the role of prior knowledge in undergraduate performance. Qualitative approaches have been widely used in chemical education research (Bodner & Orgill, 2007), but because of the type of research question and answers required, and the inherent motivation of determining whether bridging courses/additional assistance was necessary, a quantitative approach was favoured.

There are two strands to the work: a correlational study and regression analysis using multivariate linear modelling. The former examines the nature and strength of relationship between two variables and the latter aims to predict outcomes based on these relationships. Together they provide a powerful analytical protocol for the data of interest. The methodological approach and methods used are based on similar studies in mathematics (Hailikari & Lindblom-Ylanne, 2007; Hailikari et al., 2008) and psychology (Thompson & Zamboanga, 2003, 2004) as outlined in the literature review. A correlational study on the role of first year mathematics on performance in engineering was conducted by Russell (M. Russell, 2004), and his methods for reporting correlation and summary data are used as a template for similar analysis protocols here.

3.2 The Research Question and Null Hypothesis

As discussed in the introduction, the research question for this study is:

"*How does students' prior chemistry knowledge influence their performance in year 1 chemistry?*"

For the purposes of this study, prior knowledge is measured by performance at Leaving Certificate level, in that the performance here closely matches the expectations in the first year at degree level.

Additional sub-questions of interest are as follows:

(i) Is there a correlation between CAO entry points and Year 1 performance?

(ii) How does students' performance in end of year exams in subsequent years differ based on their level of prior knowledge on entry?

(iii) Do other variables exist that impact year 1 performance and if so to what extent do they predict year 1 grades. Variables such as laboratory performance, gender, semester tests and commuting distance will be examined.

(iv) Does the level of understanding of basic chemistry concepts, as measured by conceptual testing correlate to prior knowledge.

The null hypothesis was that prior knowledge did not have a positive influence on the end of year performance.

3.3 Overview of Sample and Data Gathered

The students surveyed are the first year students in a chemistry based course at a large third level institution in Dublin. There is on average 30 students per year, and data from academic years 04/05 to 08/09 comprising five years in total was available. As this course was first registered on the Central Applications Office (CAO) in 03/04, this sample set represents approximately 75% of the entire cohort of this course since its inception.

The data gathered was grouped into four categories: background, prior knowledge, course involvement and performance:

- background information on the student: CAO points, distance to college and age;
- level and quality of prior knowledge: level of prior knowledge of chemistry, diagnostic assessment

- course involvement and participation: Laboratory mark, perceived level of study, perceived level of interest, attendance;
- course performance: Semester 1 test, Semester 2 test, end of year exam mark, subsequent years exam marks.

Students entering this course of study are required to have at least one honour (HC3 or better) in a science subject in their Leaving Certificate, although chemistry is not a pre-requisite. Therefore there is a good range of students who both have and have not studied chemistry, an important point when considering the validity of the data (Dochy et al., 1999).

3.4 Data Sources

The data was obtained from a variety of sources including induction day surveys, college information system, college admissions officers and student surveys. CAO entry level points, distance to college, and Leaving Certificate chemistry results were obtained from 1st year induction day surveys for the students in question. In addition, CAO and Leaving Certificate chemistry levels for 07/08 and 08/09 were provided by the Admissions Office of the institution and cross-checked with the data gathered from the induction day surveys.

Examination performance (semester tests, lab mark, year 1 exam, subsequent years' exam marks) were obtained from the college information management system and cross-checked with year coordinator records. In the first year, there was only one chemistry module (15 ECTS) and the mark in this exam was the first year mark. An overall first year mark, taking into account semester 1, semester 2 and lab marks, as well as exam mark was not used except in the section on subsequent years. In subsequent years, there were several chemistry modules, so the average mark for each year was computed from the average of all chemistry modules. In all cases, first sitting marks were used except where a student repeated an element without prejudice where the supplemental sitting mark was used.

The student rated interest, level of study and attendance was obtained by surveying each of the four years of students in May 2008, and the cohort who were on an industrial placement in October 2008. Students were asked to respond to Likert questions rating their level of study, interest and attendance. It is acknowledged that there are inherent flaws in this data, as students in Year 2, 3 and 4 are being asked to recall their opinions when they were in Year 1, and some over-positive bias is expected. For the students in 08/09 an attendance record was maintained for the year. These surveys also asked students to provide the distance to college in year 1, and again asked students their level of chemistry at Leaving Cert and their CAO points obtained. This data was cross-checked with the Year 1 induction surveys, and where a discrepancy arose (usually students in the second survey overstated their CAO), the induction day survey data was used. There was however a very good degree of correlation between the two surveys. Students' date of birth was obtained from the college information system.

The diagnostic assessment data was gathered for students in Year 1 in 08/09. The assessment was used midway through semester 1 and early in semester 2. The same assessment was used in each case. The assessment questions were chosen so that they tested the key basic elements on both the Leaving Cert chemistry syllabus and in the Year 1 chemistry programme.

3.5 Data Treatment

3.5.1 Prior Knowledge

Students with prior knowledge of chemistry were scored according to the CAO points awarded for the grade achieved, according to the scheme in Table 2. Therefore a student who obtained a HB1 was given a prior knowledge score of 85 whereas a student who obtained an OD2 was given a prior knowledge score of 10. There were two exceptions to this. Students who scored higher level fail grades were considered to have some prior knowledge and were awarded 30 points for a Higher Level grade E (HE) or 20 points for a Higher Level grade F (HF). One student who had studied Leaving Certificate Physics and Chemistry was awarded half the score that would have been awarded for chemistry. Choosing this score or removing this student from the dataset had the same impact on the data (*i.e.* very little).

The advantage of this scoring is that it is assessing prior knowledge using externally validated data. However, students who had no Leaving Cert chemistry were awarded a score of zero. This causes some difficulties in the data analysis. The CAO scale is not continuous, although it is a good approximation to say that it is.[4] A more serious issue to consider is that the scale between zero and the rest of the scale is not uniform. This fact is taken into account in the correlational analysis (by treating zero and non-zero (*i.e.* those without and with prior knowledge) separately) and in regression analysis, *vide infra*.

[4] A continuous scale is one without break – so scores of 6, 28, *etc* would be possible. The fact that the CAO scale ranges from 5 – 100, albeit in units of 5, means that assigning it as continuous is a good approximation (Condron, 2008).

Table 2: CAO points awarded for grades at Higher and Ordinary Levels

	Higher Level	Ordinary Level
A1	100	60
A2	90	50
B1	85	45
B2	80	40
B3	75	35
C1	70	30
C2	65	25
C3	60	20
D1	55	15
D2	50	10
D3	45	5

3.5.2 CAO Entry Points

The course of study has a minimum entry point requirement set by the CAO each year based on the number of places available and the demand for the course. In line with most science based courses, and because the numbers completing the Leaving Certificate has dropped over recent years, the CAO entry points for most science based courses has fallen. In the case of this course, the points have dropped by over 100 points over the years studied (CAO, 2008). Although this is disappointing for the course itself, it makes for a broad range of statistical data! The CAO points and median values for the students on the course are presented in the Results section.

Because of institutional policies, (*e.g.* non-traditional Access routes, Mature Student entry), some students did not have the minimum points requirements. In most of these cases (e.g. Access), students did meet all other minimum requirements (*i.e.* subject requirements) and were included in the dataset. However some students were excluded after consideration; *e.g.* where the CAO entry points were very low and the student scored well in Year 1. In these unusual cases, these points were considered statistical outliers as they were non-traditional students returning to education, whose CAO points did accurately not reflect their level of ability and they were removed from the dataset.

3.5.3 Distance to College

Students provided distance to college in terms of commute time and the students' address was also available, giving a physical distance score. Given the non-linear relationship between physical distance and commute time, (*e.g.* Finglas (8.2 km) is a similar commute time to Greystones (28.2 km) because of public transport options), commute time as provided by students was used to score distance to college.

3.5.4 Student Perceptions

Students answered Likert scale questions (Very Good – Very Poor (5 points)) giving their perceived level of interest, study and attendance in Year 1. The data was converted to a scale using a score or 1 for "Very Poor" and 5 for "Very Good" in line with common practice for these questionnaires (Reid, 2006). The student survey used is available in Appendix 1.

3.5.5 Course Achievement

Course achievement was scored according to the mark achieved by the student in any of the elements surveyed (*e.g.* semester test, lab mark, diagnostic assessment, end of Year mark). In Year 1, there is one 15 ECTS module in chemistry. The end of year mark was obtained from the end of year *exam* mark rather than the module mark, as the latter takes into account the lab and assessment marks. Russell (M. Russell, 2004) considers progression rates into Year 2, but in this cohort of students, there were practically no supplemental or failed students, so year performance was used as a more appropriate indicator.

As indicated above, performance in Years 2 – 3 were computed by compiling the average of the chemistry based modules (eleven 5 ECTS modules in Year 2, six 5 ECTS modules in Year 3 because of placement) which incorporated exam and laboratory mark (weighted 70:30). The Year 4 mark consisted of ten 5 ECTS modules and a project module. The exam – lab weighting in Year 4 is 80:20. The final degree mark differs slightly in that it includes a contribution from Year 3 (20%) and adjustments made by external examiners, and so was not used.

3.6 Ethical Considerations

This study required obtaining primarily data that was readily available on the institute information management system and as such was in the domain of the academic staff. However, in order to comply with good ethical practice, students were informed about the project being undertaken. This was done using a cover letter on the survey given to all students. This letter and survey was mailed out to all students on their student VLE one week prior to the survey being given out in class to give advance notification and time to digest the details of the survey and the reasons for the study. The letter given to students is given in Appendix 1. By adhering to the BERA guidelines (BERA, 2004), this research:

- complies with the principle of voluntary informed consent by providing information on the nature of the project prior to the survey being presented, an explanation of the role of the data pertaining to them in the project and how it will be used and reported;
- avoids deception by providing explicit details the role of the research as well as contact points for more information;
- informs of the right to withdraw from the research at any time;
- does not work with students under the age of 18;
- does not provide incentives to completing the survey or participating in the research;
- ensures there is no detriment arising from the research because of participation – this is explicitly achieved by removing all names from the data set once it was compiled;
- secures anonymity (as outlined above) and by not naming or identifying by inference any student in the research;
- complies with data protection legislation as outlined above;
- ensures that all of the above is continued into the disclosure stage of the research.

The Head of School, who has ultimate responsibility for the welfare of the students in the School was informed about the research project and informed of the above ethical considerations and provided consent for the project, subject to the ethical guidelines being followed.

3.7 Data Analysis

3.7.1 Introduction

This section explicitly considers the methodology and methods used in this research. The approaches to data analysis are outlined along with the theoretical rationale for these methods. The methods are discussed below. Advice and suggestions from Dr. Joe Condron, Statistical Advice Unit, DIT are acknowledged. Any errors in interpretation of this advice are the author's alone.

3.7.2 Methodology

The methodology which will be used in this project is quantitative, correlational research with regression analysis. Quantitative research is a predominant research methodology in educational research and has its roots in late 19th century studies of children's behaviour by Hall (Creswell, 2008). It arose out of physical sciences and as such has an objective epistemology with a positivist theoretical perspective. The student related data is treated as numbers and statistical functions applied to those numbers, from which conclusions can be drawn. It is accepted here that this method will not be sensitive to several of the nuances and differences between different groups as indicated above (*e.g.* students with high self-belief who have not done chemistry and those who have done chemistry and have lower self-belief) but this study is intended to be an initial step in examining correlation and use of the data as prediction for future students. Additionally, as the literature review indicates, prior knowledge is either the sole or most significant factor in predicting future achievement, even taking factors like efficacy into account.

There are three main branches of quantitative research: experimental, correlational and survey (Creswell, 2008, 2009). Experimental research takes a sample for study, measures some data using an appropriate instrument both before and after the sample is exposed to some planned change/experience. The difference in the pre- and post-test data (or sample and placebo data) is used to determine whether the experiment had an effect. Survey research monitors the views, attitudes or any other measurable attributes of a sample and summarises them using basic statistics.

Correlational research has its origins along with the earliest of educational research, and was first used in the 1880's to quantify the level of association between two variables. Little has changed in the intervening time, and correlation is still a powerful technique in current quantitative research. The extent of correlation is quantified by a term known as the Pearson

Product Moment Correlation Coefficient, (introduced, rather confusingly, by Francis Galton in the 1880's) which is usually called 'Pearson's *r*' (see below). Correlational research has since been divided into two sub-categories: explanatory and predictory (Creswell, 2008).

Explanatory correlational research aims to examine the strength of association between two variables and explain this association based on knowledge of the two variables. Predictory correlational research is much more powerful. It aims to examine correlations and use knowledge of these correlations to *predict* future events based on the correlations observed. Some care is required, and a general rule of thumb is that 'correlation does not imply causation'. In this regard, predictory correlational research is usually used along with regression analysis, *vide infra*.

3.7.3 Data Analysis Methods

There are several data-analysis methods used in this study ranging from basic statistical calculations and tests, through correlation, to more advanced multilevel regression modelling. All data analysis was performed on SPSS 15.0 for Windows, using standard techniques and approaches for this software (Kinnear & Gray, 2006; Muijs, 2004).

3.7.3.1 Basic Analysis

Basic statistical calculations such as mean, median and standard deviation are used to summarise large numbers of data effectively. The mean is the computed average score. Two mean values (*e.g.* a mean value for students with prior knowledge and one for students without) can be compared using a statistical test called the t-test, which determines whether there is a statistical difference between the means or whether that difference is too small to distinguish if it arises from chance or not. A threshold of 95% significance is the norm and is used here. This means that a test is deemed to be significant if the possibility that the true result arises from chance is 1 in 20 or less. 95% significance is also termed "alpha = 0.05", and a value that is statistically significant at 95% confidence is said to be significant at $p > 0.05$. A value which is significant at 99% (*i.e.* a 1 in 100 chance that the result is due to random factors) is obviously more significant, and is said to be significant at $p > 0.01$, *etc*.

Mean values only give part of the picture, and other useful values are the median (the middle value in a range) and standard deviation. These give some indication of the range of data. A useful way of summarising the range is to use box-plots. These divide the data into four quadrants (inter-quartile ranges) and use a box to represent the range of values in the middle 50% of values, with 'whiskers' to indicate the 25% each side. In this study, box-plots, mean values and t-tests are used extensively to compare data between students with and without prior knowledge. Standard deviations are used to quantify the range more explicitly – a large standard deviation value indicates a broad range of numbers.

3.7.3.2 Correlation

Creswell (Creswell, 2008) recommends the use of scatter plots to examine regression data visually, before performing any correlational analysis. This allows for a general impression of the data to be gained and means that the statistical values subsequently calculated can be checked to see if they tally with what would be expected. There are three different scenarios (Figure 3) that can arise from correlational analysis:

- As one variable is increased, the other is observed to increase as well. This is a positive correlation.
- As one variable is increased, the other is observed to decrease. This is a negative correlation.
- As one variable changes, there is no observed trend in determining how the other variable will change – there is no correlation.

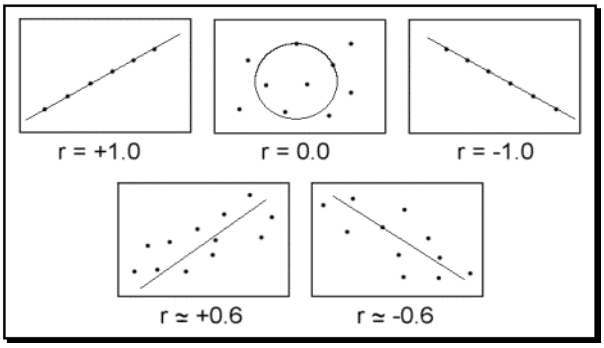


Figure 3: Model scatter-plots showing positive, negative and no correlation with intermediate scenarios (allpsych.com, 2008)

The *strength* of correlation, in terms of how closely one variable follows the other on changing, obviously varies and as indicated above value called Pearson's *r* is used to quantify this strength, according to the following arbitrary bands:

$r = \pm 1$: a perfect correlation; $r > \pm 0.8$: a very strong correlation; $r > \pm 0.5$: a strong correlation; $r > \pm 0.3$: a moderate correlation; $r < \pm 0.3$: a modest correlation; $r < \pm 0.1$: a weak correlation.

A positive sign indicates that it is a positive correlation. Correlation is also checked for statistical significance, and only those values which are considered *significant* correlations can be considered. Correlations are effectively summarised in a correlation matrix, which shows how each of the several factors under study correlate with each other.

The Pearson's *r* is used when both variables are continuous. In cases were variables are ordinal, (*e.g.* Likert responses), Spearman's rho (ρ) is used. The value and ranges obtained are analysed analogously to Pearson's *r*.

3.7.3.3 Regression

Correlation has a use in examining the strength of relationship between two variables but it does not imply causation – *i.e.* that one variable caused another. Regression takes this next step, by examining how much of the variance can be ascribed to one particular factor. This can be further extended to examining two, three or several variables at once, and examining which of these influences the variable of interest the most (Bartholomew, Steele, Moustaki, & Galbraith, 2008). It is then possible to *predict* the variable of interest based on the analysis. This approach is multiple regression. The approach is to examine several factors in turn and examine how the 'fit' (R^2 value; a perfect R^2 is 1) improves as each variable is added. Therefore, in this study, a hypothetical regression may be performed by using the model equation:

$$[\text{Year 1}] = a + \beta_1[\text{CAO}]$$

This would return values for $\boldsymbol{a}$ and $\boldsymbol{\beta_1}$ from the modelling and would allow calculation of a Year 1 score based on an inputted CAO value. For example, suppose $\boldsymbol{a}$ is returned as 55, and $\boldsymbol{\beta_1}$ is 0.0025, with $R^2 = 0.3$; then a CAO score of 400 would predict a Year 1 score of:

$$[\text{Year 1}] = 55 + 0.025 \times 400 = 65\%$$

This is clearly a poor regression fit, as the $\boldsymbol{a}$ value is large and unexplained, and the R^2 value is low. If a second factor, for example, attendance, is introduced, thus:

$$[\text{Year 1}] = a + \beta_1[\text{CAO}] + \beta_2[\text{Attendance}]$$

and values for $\boldsymbol{a}$, $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$ and R^2 are 35, 0.08, 3 and 0.5 are obtained, the prediction for a student who obtains 400 CAO points and has an attendance score of 70% (0.7) is:

$$[\text{Year 1}] = 35 + 0.08 \times 400 + 3 \times 0.7 = 69.1\%$$

This second equation gives a better R^2 value indicating it is a better fit and it now models for a second term. The actual data can then be viewed to see if a student who meets these criteria of inputted CAO and attendance obtains a score of 69.1%. Obviously there will be some degree of variability with this value, and the confidence limits are also calculated, to give an indication of this variability. In other words, the first model with only one term may return a value of Year 1 = 65% $\pm$ 15% (where 15% is the confidence limit), whereas the second improved model may return a value of 69.1% $\pm$ 3%. This allows for greater confidence in the value and the range of the value being quoted. Of course, all of these values and the confidence limits are subject to the scrutiny of significance, again chosen to be 95%.[5]

This process can be continued for as many terms as required. Whether that term significantly improves the model can be examined by looking at how the R^2 value changes.[6] Note that the absolute β values depend on the number they represent; for a large score like CAO they will be smaller. The SPSS output also gives the relative $\boldsymbol{\beta}$ values (*i.e.* relative weightings) which can be used for comparative purposes.

[5] Modern statistical analysis is moving away from such a rigorous cut-off point ("the altar of $p > 0.05$" (Osborne, 2008)) for null-hypothesis significance testing and toward confidence ranges and likelihoods. Although the concept of significance of $p > 0.05$ is still universally accepted in both the social science and physical science communities, this work makes attempts to incorporate best practice from modern statistical approaches as espoused by more recent writers on the topic (Fidler & Cumming, 2008).

[6] The contribution of Dr Joe Condron, Statistical Advice Unit, DIT to the development of the regression model used in this study (as outlined in the Results Chapter) is gratefully acknowledged (Condron, 2008).

Chapter 4

Results

4.1 Profile of Students Studied

This study involves the examination of the academic profiles of students who completed Year 1 of a programme in analytical chemistry at a large third-level institution in Dublin in the academic years 04/05 to 08/09. In order to best facilitate the examination of data over the time period of the study, two separate strands were pursued. The first is the study of students who have already completed Year 1, for whom the data indicated in Table 3 is known. These data were collected and scored as described in Chapter 3. The second are students who were in Year 1 (08/09) as the study was being conducted and whose end of year marks were not available until the latter end of this study. Some additional data was gathered for this group, as indicated in Table 3. The data discussed below refers to the main cohort (04/05 – 07/08). Students from this year's group are considered separately in Section 4.5.

4.1.1 CAO Profile

CAO points were available for all students in the data set. The required points for entry into the course have declined year-on-year in line with a general drop in points for Science related courses across the HE sector. The average CAO entry points for the group was 414 with a range from 300 – 555. A proportion (< 5) were accepted onto the course with lower than the required amount of CAO points by alternative entry (*e.g.* Access) routes. The CAO entry points of students who have and have not prior knowledge of chemistry is shown in Figure 4. This plot shows the distribution of CAO points indicating the inter-quartile range as well as identifying the minimum and maximum points in each group. It is apparent that students who have not selected Leaving Certificate Chemistry have in general a lower CAO score than those who have (note in particular the median value). The average CAO score of the former group is 394 points whereas that for the latter is 425 points. A t-test demonstrated that there was a significant difference between these means (t = 2.684, dF = 87, p = 0.009).

Table 3: Data gathered for student cohorts studied

Data Type	Code	04/05 – 07/08	08/09
CAO Points	CAO	☑	☑
Leaving Cert Chemistry Mark	PK	☑	☑
Year 1 Exam Mark	Y1	☑	☑
Semester 1 Test Mark	S1	Except 04/05*	☑
Semester 2 Test Mark	S2	Except 04/05*	☑
Year 1 Lab Mark	LAB	☑	☑
Distance to college (time)	DIS	☑	☑
Student Rated Attendance (Y1)	SRA	☑	☑
Student Rated Interest (Y1)	SRI	☑	☑
Student Rated Study (Y1)	SRS	☑	☑
Gender	GEN	☑	☑
Age	AGE	☑	☑
Diagnostic Test	DIAG	-	☑
Attendance (Recorded)	AR	-	☑
Sample Size N		89	27

*A combined semester test mark is available

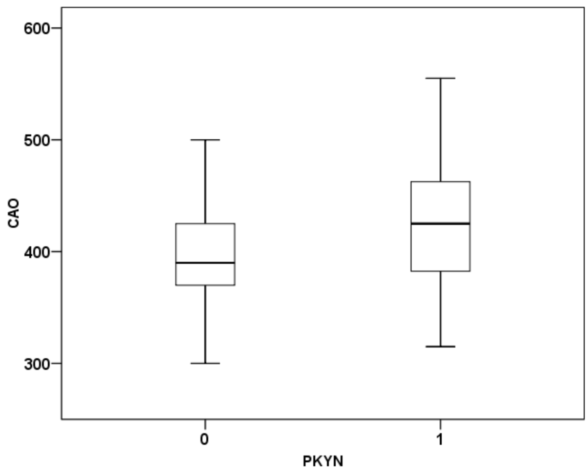


Figure 4: Range of CAO entry points for students with (PKYN = 1) and without (PKYN = 0) prior knowledge

In examining students on a year by year basis, it is clear that this emphasis is mainly due to the 04/05 cohort (Figure 5) which has a significantly higher median than the other three cohorts. Analysis of the data by gender shows no consistent trend of achievement of one group over another.

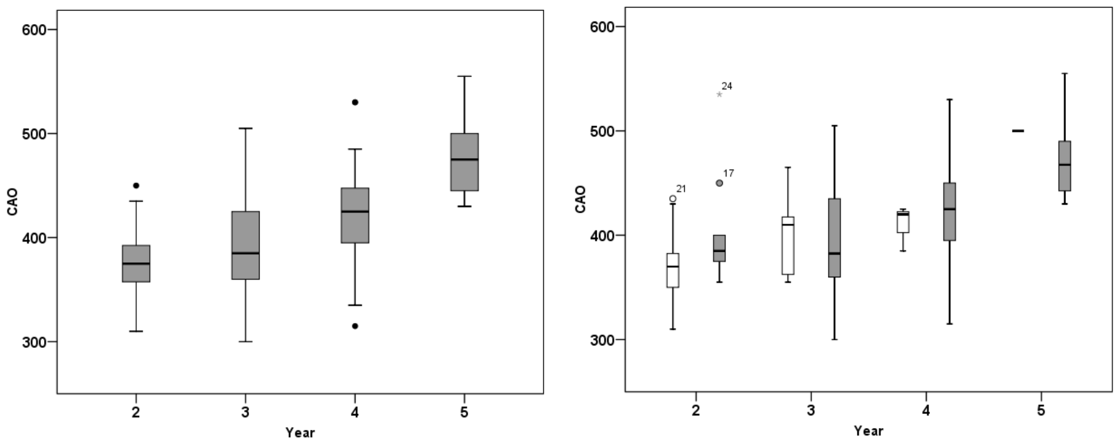


Figure 5: Range of CAO entry points for students by year (left) and by year as distinguished by gender (right, male = unfilled, female = filled) with Year 2 = 07/08 and Year 5 = 04/05 cohort

4.1.2 Prior knowledge

Prior knowledge, defined as having Leaving Certificate chemistry was quantified as described in Chapter 3. Of the 89 students studied, 56 had prior knowledge with grades as outline in Figure 6.

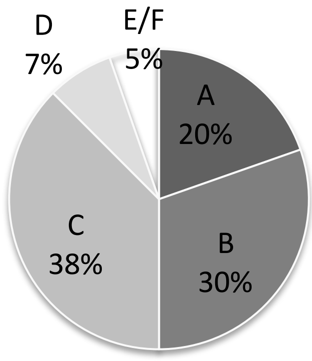


Figure 6: Grades obtained by students who had prior knowledge of chemistry

The year 1 scores of students with and without prior knowledge was examined, by comparing the mean scores for students with prior knowledge (mean = 64%, N = 56) and those who have not PK (mean = 50%, N = 33) using a t-test. This concluded that there is a significant difference between the mean Y1 score for the two groups of students; $t = 4.288$, $dF = 91$, $p < .0005$.

4.1.3 Semester Tests

Two semester tests are held each year for this year of the programme. The breakdown of these marks is shown in Figure 7. A comparison of the average score between those with prior knowledge and those without for both of the semester marks are shown in Table 4. The results indicate that there is a highly significant difference between the grouped averages for both tests ($p < 0.0005$).

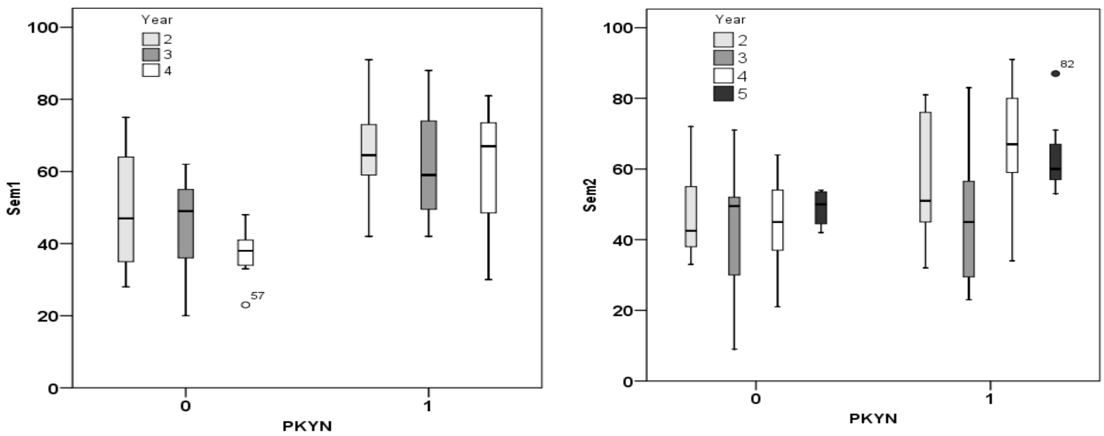


Figure 7: Semester 1 (left) and Semester 2 (right) scores distinguished by year. (2 = 07/08, 5 = 04/05) Only the overall semester test marks are known for 04/05 cohort and these are shown on the Semester 2 plot

Table 4: Data analysis of Semester 1 and Semester 2 Tests

	Mean (%) (Overall)	Mean (%) (PK = 0)	Mean (%) (PK ≠ 0)	Sig**	dF	t
Semester 1	56	44	63	.000	71	5.166
Semester 2*	55	46	60	.000	75	4.133

*Includes 04/05 cohort overall Semester 1 and Semester 2 mark (Not including these marks gave similar results)

** A significance value of .000 indicates significant to $p < 0.0005$ – *i.e.* highly significant.

4.1.4 Lab Marks

Laboratory classes are held over the two semesters and a combined mark for each of the cohorts were compiled. These are shown in Figure 8. The average for those who had no prior knowledge was 69%, whereas that for those who did was 75%. A t-test found a small but significant difference between the means ($t = 3.494$, $dF = 87$, $p < 0.05$).

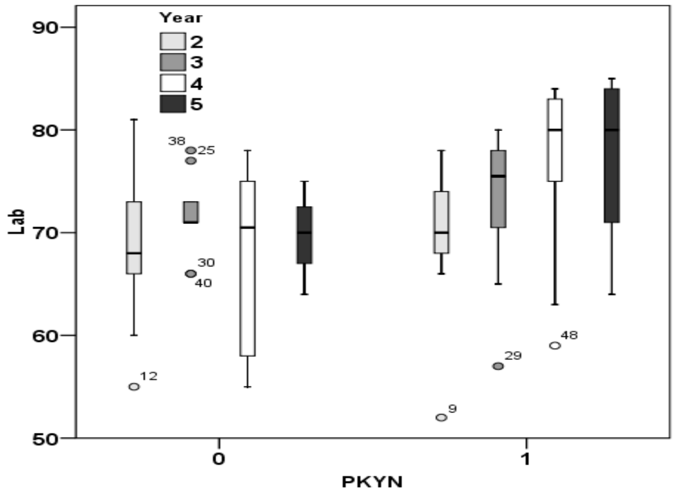


Figure 8: Lab marks as distinguished by year (2 = 07/08, 5 = 04/05)

4.1.5 Distance to college

Data was collected to determine how far students lived from college. Students were asked to provide an estimated journey time. The data collected is shown in Figure 9. The trend observed is that slightly more students that do not have prior knowledge commute further distances. The means travel times for each of the years are provided in Table 5. t-Tests showed that there were no significant difference between mean travel times across the years surveyed nor between students with and without prior knowledge.

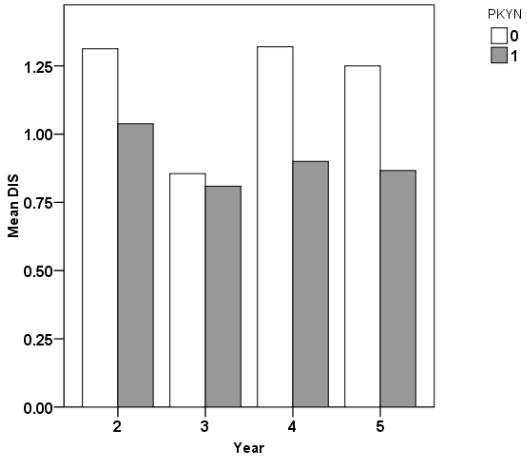


Figure 9: Average travel time (hours) distinguished by students who have (filled) and have not (unfilled) prior knowledge

Table 5: Average commute times for student cohorts studied (with 95% confidence limits)

Year	Average Commute Time (hours)
07/08	1.2 ± 0.3
06/07	0.8 ± 0.2
05/06	1.0 ± 0.25
04/05	1.0 ± 0.4

4.1.6 Student Perceptions of Year 1

Students from all cohorts 04/05 to 07/08 were surveyed on their perceptions of year 1 in terms of their interest, their level of attendance and their level of study. While these surveys are inherently subjective and depend on the subjective views of the students, they are of use in examining whether these parameters have an effect on year 1 performance. The survey is outlined in the Chapter 3 and asks students Likert style questions. As can be seen from the range of answers in Figure 10, the responses reflected that students considered their attendance (4/5 SA or A), level of study (2.2/5 SA or A) and level of interest in the course (3.3/5 SA or A) above average. There were 59 responses from the group of 89 students in the dataset.

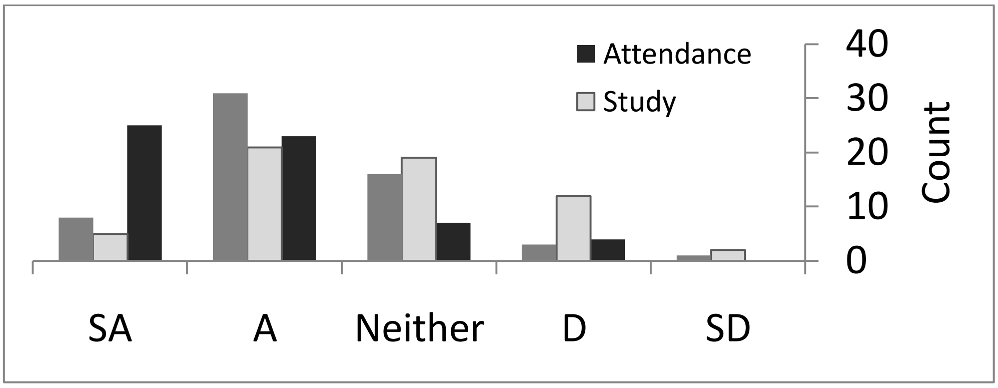


Figure 10: Student responses to the questions gauging their level of attendance, their level of interest and their level of study in year 1

4.2 Bivariate Correlational Studies

In this section the various data collected is compared to examine whether and to what extent a relationship exists between two variables studied. Correlation is quantified using Pearson's *r*, as described in Chapter 3. The strength is assigned arbitrary bands; if $r < \pm 0.1$, then the association is weak; $r < \pm 0.3$ is modest; $r < \pm 0.5$ is moderate; $r < \pm 0.8$ is strong and $r > \pm 0.8$ is very strong. For ordinal values, correlation was quantified using Spearman's rho, with similar bands assigned.

4.2.1 CAO

4.2.1.1 CAO and Year 1 Performance

The correlation between CAO on entry and year 1 performance is shown in Figure 11. Correlational analysis to determine Pearson's *r* shows that there is a moderate correlation (r = 0.366**, $p < 0.01$) between the two variables. It can be observed on the data set that there are four outliers in the high CAO – high Y1 area of the scatter-plot. The correlation reduces when these outliers are removed (r = 0.214*, $p < 0.05$). The correlations between CAO and Y1 are summarised in

Table 6. The correlation is moderate to strong for students who have completed Leaving Certificate chemistry but there is no significant correlation for students who have no prior

knowledge. Separate scatter-plots for CAO with and without prior knowledge are provided in Appendix 2.

Table 6: Correlation between CAO and year 1 performance

Variables Compared	Correlation coefficient *r*
CAO – Y1	0.366** - 0.241* (see text)
CAO (with PK) – Y1	0.515**
CAO (no PK) – Y1	No significant correlation

* = p < 0.05; ** = p < 0.01

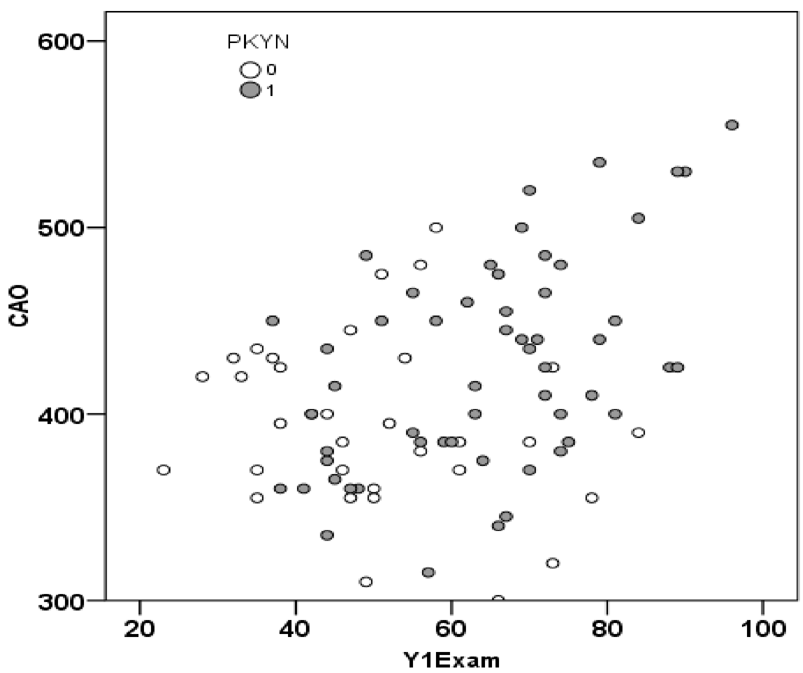


Figure 11: Scatter plot showing CAO points and year 1 exam performance for students with (filled) and without (unfilled) prior knowledge

4.2.1.2 Other indicators

CAO shows a modest correlation with year 1 performance, above, but correlation between prior knowledge and other indicators ranges from insignificant to modest for the range of values. Apart from Year 1, the strongest correlation is that with semester 2 performance (*r* =

0.337**). No significant correlation or a very weak correlation was observed between CAO and S1, Lab (*r* = .213*), DIS, SRA, SRI and SRS.

4.2.2 Prior knowledge

4.2.2.1 Prior knowledge and year 1 performance indicators

From the above data, it can be deduced that there is a correlation between students who have completed Leaving Cert Chemistry and their performance in Year 1 exams. In order to examine this important facet of this work directly, a correlational analysis was conducted between the variables prior knowledge and year 1 performance.

The scatter-plot of the data is shown in Figure 12. It can be seen that there is a visual correlation between students with prior knowledge and year 1 performance. Examining the data-set as a whole, a correlation coefficient or $r = 0.569^{**}$ ($p < 0.01$) was calculated. This indicates a moderate to strong correlation between year 1 scores and prior knowledge. As described earlier, a t-test indicates that there is a significant difference between the mean Y1 score for students who have prior knowledge (mean = 64%, N = 56) and those who have not prior knowledge (mean = 50%, N = 33); $t = 4.288$, dF = 91, $p < .0005$. These data suggest that prior knowledge has an impact on Y1 performance.

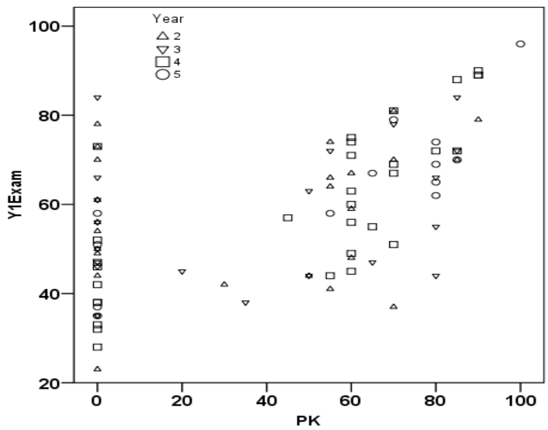


Figure 12: Scatter-plot showing the year 1 scores based distinguished by year (2 = 07/08, 5 = 04/05)

Incorporating the data where students had no prior knowledge is may be statistically flawed, as this assumes the zero value is on the continuous scale (see Methodology chapter). Therefore the correlation coefficient, excluding these values, *r* increases to 0.654***.

The effect of gender was also examined and the results are shown in Table 7. The average Y1 mark for males (N = 26) was 60%, whereas that for females (N = 63) was 58%. There was no significant difference between the means of the male and female scores in Y1 ($t = 0.641$, $dF = 93$, $p < 0.05$). A correlation of 0.561** for male students (N = 26) and 0.583** for female students were calculated. This should be taken in the context of the fact that the male population of the sample is predominantly in the 07/08 cohort.

Table 7: Pearson's r correlation study between PK and Y1 by gender

	N	Mean (%) (Y1)	r
Male	26	60	0.561**
Female	63	58	0.583**

**p < 0.01

Similar findings are observed for prior knowledge and other Year 1 performance indicators S1 and S2, both of which show strong correlations with prior knowledge, and to a lesser extent Lab performance (Table 8). The correlation between prior knowledge and semester 1 exam performance is the strongest correlation observed among the four year 1 indicators (S1, S2, Y1 and Lab).

4.2.2.2 Prior knowledge and other indicators

Prior knowledge does not significantly correlate with any of the student rated replies SRS, SRI and SRA (Table 8) suggesting that students' engagement in the programme in year 1 was not influenced by not having prior knowledge of chemistry. (A correlation is calculated between prior knowledge and term-time distance in year 1, which is clearly a meaningless result.)

Table 8: Correlation matrix showing inter-correlation for all variables in dataset. All values are Pearson's r except those for ordinal variables - SRA, SRI and SRS - which are Spearman's rho.

	PK	Y1	S1	S2	LAB	DIS	SRA	SRI	SRS
CAO	.746**$^{\Phi}$	.366**$^{\Delta}$	.205	.337**	.213*	-.102	-.201	-.093	-.015
PK		.569**	.592**	.529**	.382*	-.241*	.006	.123	.124
Y1			.654**	.684**	.264*	-.238*	.240	.326*	.293*
S1				.611**	.188	-.046	.139	.452**	.050
S2					.353**	-.020	.074	.092	.227
LAB						0.031	.073	.155	.084
DIS							-.075	-.218	.024
SRA								.295*	.285*
SRI									.435**

* = $p < 0.05$, ** = $p < 0.01$

Φ: This value excludes those without Leaving Certificate chemistry for reasons discussed in Chapter 3.

Δ: Having removed four outliers, the value is 0.263*. See text for details.

4.2.3 Year 1 performance indicators

Strong correlation was observed between the three main year 1 performance indicators: S1, S2 and Y1, the strongest on the table apart from (the obvious) correlation between CAO and PK. S1 and S2 exhibited a correlation r = .611**, an interesting follow-on from the correlation between Y1 and S1. The second semester test showed a somewhat higher correlation with the year 1 score which is unsurprising as the test is usually held in the middle of semester 2 and the exam is held in May. The year 1 performance negatively correlated with commuting time (DIS), although as observed above there was no significant difference between the average travel times of each group.

4.2.4 Student-Rated Data

There was no observed correlation with student rated attendance and other non-student rated variables. Student rated interest correlated moderately with both the year 1 and semester 1 mark. Student rated study showed a modest correlation with year 1 performance.

Each of the three student rated variables demonstrated inter-correlations, with interest and level of study showing a moderately strong correlation.

4.3 Regression Analysis

As described in the Methodology chapter, regression provides for a powerful analytical progression from correlation analysis by means of examining the effect strength of each of the variables. There are three steps to the regression analysis approach used in this research:

- Consideration of prior variables: CAO and distance of commute
- Consideration of prior knowledge
- Consideration of engagement in year as measured by laboratory work
- Consideration of performance in year as measured by semester tests

The rationale for this approach is outlined in the Discussion. The summary of important data outputted from the model is shown in Table 9. All stages in the model were statitistically significant according to ANOVA testing ($p = .006$ for step 1 and $p < .0005$ for steps 2 – 4).

4.3.1 Regression Step 1: Background Variables

The first step in the regression was to model the year 1 performance solely on the CAO score and commute distance, two background variables of the student involved. Age was not included in the analysis as there were not a diverse range of ages and therefore was not a discriminating factor. The regression output is given in Table 9. The CAO value is significant ($p = .008$), whereas the distance value is not. The R^2 value is 0.174, which indicates that these variable only account for 17% of the variance in the year 1 score. This regression implies that a student who had the average score of 414 points and commute took the average time of 1 hour to college would get an end of year score of 59%, according to the generated equation (using actual rather than standardised β values):

$$\text{Year 1} = a + \beta_{CAO}[CAO] + \beta_{DIS}[DIS]$$

The average year 1 score is 59%. However, the R^2 value is low, which means it may be possible to improve the model, and the commute time is found to be insignificant to the model.

4.3.2 Regression Step 2: Prior Knowledge

The second step incorporates prior knowledge, again in line with the approach used by Thompson (Thompson & Zamboanga, 2004). Several observations are noted. The first is that both the R^2 and adjusted R^2 increase by .172 and .166 respectively, indicating that the prior knowledge, coupled with the variables in the previous step account for 35% of the variance in the year 1 score. The former will increase somewhat on addition of a subsequent variable to regression, so the latter measure is useful to take into account what actual increase is due to the contribution of a new variable to the regression model. The coefficient is now significant (p = 0.02) and now the only significant variable is that of prior knowledge, which is highly significant (p = .001). In testing the model, using average scores, the model again returns a score matching the average of 59%.

4.3.3 Regression Step 3: Level of Engagement

The third step in the regression was to examine course engagement as measured by laboratory work. The laboratory scores were incorporated into the model. This has no effect on the model. The laboratory scores are insignificant, and the prior knowledge scores are again highly significant. Additionally, there is little change in the standardised β values for CAO, distance and prior knowledge again indicating that the lab score has no effect. The R^2 value increases marginally (which would happen in any case) but the adjusted R^2 decreases, indicating that the laboratory score has a negligible explanation for the variance in the year 1 exam score.

4.3.4 Regression Step 4: Course Performance

4.3.4.1 Incorporating Prior Knowledge

The final step is to consider course performance at the end of the year by incorporating course performance indicators from during the year – the semester 1 and semester 2 tests. When these are included, several interesting observations are noted. The first is that the R^2 and adjusted R^2 values increase to .786 and .619 respectively. This complete model now accounts for 75% of the variance in the year 1 exam score. Additionally, distance (p = .04) is significant, and each of the semester tests which are highly significant for both semester 1 (p = .008) and semester 2 (p = .002). Prior knowledge is highly insignificant in this model. Modelling the equation with average values gave a predicted scores of 52%.

4.3.4.2 *Excluding Prior Knowledge*

The insignificance of prior knowledge in the latter stage of the above analysis prompts further analysis. The multi-step model was again completed, except in this case, prior knowledge (step 2) was omitted. The final step in the model resulted in similar values being obtained as shown in Table 9. The R^2 and adjusted R^2 values were .786 and .618 respectively. The standardised β coefficients for CAO (.128), distance (-.201), lab (.033), semester 1 (.367) and semester 2 (.402) were similar to those observed for the analysis incorporating prior knowledge (*c.f.* Table 9). The testing of the model with average values gave a predicted score of 55%. The relevance of this analysis is considered in the Discussion.

Table 9: Hierarchical Regression Analysis of Predictors of Year 1 Exam Performance

Step	Variables entered	R^2	df (1,2)	ΔF	Coefficient	β (Step 1)	β (Step 2)	β (Step 3)	β (Step 4)	β (Standardised)
1. Background		.174	2, 53	5.59**	21.72					
	CAO					0.104**				.345**
	Distance					– 5.61				–.202
2. Prior Knowledge		.346	1, 52	13.71***	37.13*					
	CAO						0.037			.123
	Distance						– 3.00			–.108
	Prior Knowledge						0.227***			.484***
3. Course Engagement		.350	1, 51	.301	27.56			27.56		
(Laboratory)	CAO							0.036		.121
	Distance							–3.25		–.117
	Prior Knowledge							0.214***		.458**
	Lab Mark							0.146		.068
4. Course Performance		.618	2, 49	17.255***	6.95			6.95		
(Semester Tests)	CAO							0.035		.115
	Distance							–5.34*		–.192*
	Prior Knowledge							0.019		.042
	Lab Mark							0.051		.024
	Semester 1 Mark							0.326**		.349**
	Semester 2 Mark							0.371**		.398**

Note: Unstandardised beta values shown in each of the steps 1 – 4, standardised beta values are shown in right hand column, * $p < .05$; ** $p < .01$; *** $p < 0.001$

4.4 The Role of Prior Knowledge in Subsequent Years of Study

4.4.1 Introduction

The impact of prior knowledge on Year 1 performance is the main emphasis of this work. However, an interesting sub-question is to consider the impact of prior knowledge on subsequent years, and ultimately the degree classification. Preliminary studies are presented here, as the data available reduces from year 2 through to year 4, both because of student withdrawal and the fact that the study is centred on students currently completing the course, so the number of results for each subsequent year reduces accordingly. Nevertheless, it is intended that these preliminary results will be followed up as more results come available each year and the aim of this section of the study is to seed that work.

4.4.1 Descriptive Statistics

The year totals for each of the subsequent years, along with the year total for year 1, prior knowledge score and CAO point score were subjected to correlational analysis. The descriptive statistics for the data available is shown in Table 10. The number of data values for the later years are lower, as students have not yet completed these years.

The mean score for each of the years shows that, considering overall marks, students perform best in first and fourth year with an average fourth year mark of 63.7%. The final degree mark consists of 80% of this Y4 mark and 20% of the third year (Y3) mark. The year three mark is probably a little lower than expected as all of year three exams take place in the first semester, whereas students traditionally do better in their second semester marks. As mentioned above, the mean degree grade is not considered, as it incorporates adjustments made at exam boards, external examiner modifications, *etc.*

In order to examine whether there was a difference between subsequent years' results for students with and without prior knowledge in first year, the average marks in each case where compared using a t-test. Table 11 shows the results. It can be seen from these results that both the Year 1 Exam and the Year 1 Overall (end of module mark) show highly significant differences between the mean results for students with and without prior knowledge, as discussed earlier in this chapter. However, there is a surprising change in later years. For Year 2, the means for those without and with prior knowledge are 58% and 60% respectively, with the t-test result determining that the difference between the means is highly insignificant. In year 3, the mean values for those without and with prior knowledge are 54% and 60%, with an insignificant difference (just about the 0.05 threshold) between the means and in year 4 the difference between the means for those without and with prior knowledge (62% and 64% respectively) is highly insignificant.

Table 10 : Descriptive statitistics for subsequent years

Variable	Mean	Standard Deviation	N
CAO	**414**	**54.6**	**89**
PK	**42.2**	**35.0**	**89**
Y1-Overall	**63.2**	**10.5**	**89**
Y2	**59.6**	**7.2**	**70**
Y3	**58.0**	**9.8**	**50**
Y4	**63.7**	**7.3**	**38**

Note: Y1-overall is the combined year 1 chemistry module mark (S1, S2, Lab and Y1 Exam); Y2 – Y4 are the aggregate average for each of these years.

Hence it can be concluded that despite the highly significant difference between mean scores in year 1 for semester tests, lab marks and year 1 exams, there is no significant difference in subsequent years' scores between students who entered this course without and with prior knowledge in the form of Leaving Certificate Chemistry.

Table 11 : Comparison of Means of Years 1 – 4 Scores (including Year 1 exam and Year 1 overall) for students with (PKYN = 1) and without (PKYN = 0) prior knowledge (* = average across all modules)

	Mean	N	t	dF	p
Year 1 Exam					
PKYN = 0	50	33	4.288	91	< .0005
PKYN = 1	64	56			
Year 1 Overall					
PKYN = 0	56	33	5.524	91	<.0005
PKYN = 1	67	56			
Year 2*					
PKYN = 0	58	22	0.808	68	0.422
PKYN = 1	60	48			
Year 3*					
PKYN = 0	54	16	1.971	48	0.055
PKYN = 1	60	34			
Year 4*					
PKYN = 0	62	9	0.792	36	0.433
PKYN = 1	64	29			

4.4.2 Bivariate correlational analysis

Bivariate correlational studies were carried out on the data available to examine whether there were correlations between students CAO and/or prior knowledge and each of the years of study. The results are shown in Table 12. There is a strong to very strong inter-correlation between each of the year marks and the year 4 mark, with the marks in year 3 correlating very strongly (r = .811, $p < 0.01$) with the year 4 mark. The CAO score correlates to a greater extent to year 1 (overall) and year 4 than prior knowledge. Year 3 is probably anomalous for reasons of exam timing described above. These data are therefore in general consistency with the descriptive statistics above, showing that prior knowledge is not a significant predictor of subsequent years study.

Table 12: Correlation matrix showing CAO and PK to Y1-Total and subsequent year's performance Y2 – Y4 (values are Pearson's *r*)

	PK	Y1-Total	Y2	Y3	Y4
CAO	.746**$^{\Phi}$	.304**	.403**	.302*	.544**
PK		.633**	.254*	.398**	.336*
Y1-Total			.627*	.498**	.575**
Y2				.691**	.748**
Y3					.811**

* = $p < 0.05$, ** = $p < 0.01$

Φ: This value excludes those without Leaving Certificate chemistry for reasons discussed in Chapter 3.

4.5 Quality of Prior Knowledge: Analysis of Misconceptions

The current group of dirst year students, who completed examinations in June 2009 were analysed in an analogous manner, including their diagnostic assessment and recorded attendance. Table 13 shows the summary of descriptive statistics.

Table 13: Descriptive Statistics for Current First Year Cohort (N = 24)

Variable	Mean	Std Dev
CAO (points)	387	58.5
PK (%)	58.6	15.7
Y1 (%)	48.3	16.3
S1 (%)	40.9	17.7
S2 (%)	42.7	19.3
Lab (%)	69.9	6.2
Recorded Attendance (RecA) (%)	76.7	14.3
Diag (%)	49.6	11.1

This sample group is much smaller (N = 24) than the main group, but interestingly here the average CAO score for those without Leaving Certificate Chemistry the almost the same as those with Chemistry (and a t-test confirmed that the difference between the means was statistically insignificant). The overall average score is 387 points. The mean prior knowledge score among students who have prior knowledge is 58.6%.

The correlation between the variables shown in Table 13 were determined with a view to examine how the diagnostic assessment correlates with prior knowledge. Table 14 shows the results.

Table 14: Correlations between variable recorded for 08-09 cohort (N = 24)

	PK	**Y1**	**S1**	**S2**	**Lab**	**AR**	**Diag**
CAO	.181	.170	.257	.043	.000	-.043	.204
PK		.264	.565**	.162	-.286	-.083	.340
Y1			.596**	.807**	.286	.388	.040
S1				.674**	.164	.387	.138
S2					.448*	.699**	-.023
Lab						.634**	-.203
AR							-.090

*p < 0.05, ** p < 0.01

These tests throw up some interesting findings relative to the main body of work, but it is difficult to say whether this is due to the small sample set. However, it can again be seen that each of the year performance indicators S1, S2 and Y1 all strongly inter-correlate. Both CAO score and prior knowledge do not significantly correlate with any variable, except PK – S1. The recorded attendance correlated strongly with S2, but not with Year 1, probably because the average attendance was very high in any case. Finally, the diagnostic assessment does not correlate with any variable studied.

4.6 Summary of Results of Study

This study has examined the entry CAO points, prior knowledge and a range of in-programme performance factors for several cohorts of students on an analytical chemistry programme in a large third level institution in Dublin.

The average CAO score for students of entry was found to be 414 points. CAO points differed significantly between students who had prior knowledge of chemistry (*i.e.* had completed Chemistry at Leaving Certificate Level) and those who had not. However, examining the relationship between students with/without prior knowledge, it was found that the former group showed a strong correlation with year 1 exam scores, whereas the latter group showed a weak correlation. In addition, the mean scores in semester tests for students with prior knowledge were significantly higher than those without. This allowed for the distinction between the role of the CAO and prior knowledge in the performance in year 1 exams, with prior knowledge demonstrably the determining factor. Prior knowledge correlated significantly with semester tests and to a lesser extent laboratory scores. The beneficial impact of prior knowledge was further investigated by regression analysis, which again showed that prior knowledge was an important, statistically significant factor, even when taking CAO points score into account.

Some intergroup trends were observed in the analysis. In line with the fall in the CAO cut-off points requirement, the CAO points range for each successive cohort has dropped, resulting in an overall drop in median value of approximately 100 points. There were fewer males in the earlier, higher points, cohorts but as the points have dropped, the number of males enrolled on the programme has increased. Unlike at Leaving Certificate on the whole however, there was no significant difference between the average CAO points of males and females enrolling on the course, nor was there a significant difference between gender for year 1 performance.

From the descriptive statistics and correlation analysis, regression analysis was used to develop models for predicting future scores. A four stage model was used to probe the use of an array of variables in predicting the year 1 score. Prior knowledge was highly predictory, until semester scores were incorporated, at which point it became insignificant. This is due to multi-collinearity – the prior knowledge value is already considered in the semester scores at this stage of the model. Laboratory scores were insignificant and poor predictors of scores and distance from college had a slight detrimental effect on year 1 scores. Using these results, the discussion section will consider two models which were developed out of the study – a model for early in the semester when background data (CAO, prior knowledge, distance) are known and a model for later in the year when semester results are known. The models are compared and discussed in terms of usefulness.

Finally, a surprising, but pleasing result is that despite strong dependence of year 1 scores on prior knowledge, subsequent year marks are not distinguishable as year 1 scores were. There are no significant differences between mean scores for each group and prior knowledge correlates weakly (and to a lesser extent than CAO) to each of the year scores. These observations will be discussed in the next chapter.

He diagnostic assessment showed little predictory power in terms of year 1 performance and other year 1 indicators. This may be due to the small sample set or other factors (for example the test was run in open conditions where students could look at each other answering).

Chapter 5

Discussion

5.1 Introduction

This work aims to study the importance of prior knowledge in students' performance at undergraduate level. The study focuses primarily on Year 1 performance, but also gives consideration to the performance in years 2 - 4. There are several motivations for the study. Primarily, the low numbers of students choosing science at senior cycle at school level means that third level institutions have to compensate in some way for the absence of knowledge in a subject discipline for a degree. Irish institutions cope with this in two main ways; they either provide bridging courses to get the fundamentals across and then progress with Year 1 'as normal', or they use Year 1 to ensure all students are at an equal level entering Year 2. Therefore, by examining whether prior knowledge has a role in student performance at Year 1 level, this study aims to address whether students truly do achieve a 'level-playing field' at the end of year 1. Consequent analysis of later years provides insight as to whether any implicit differences are carried through to later years.

As outlined in the literature review, it is generally accepted that prior knowledge has a beneficial influence on future performance. Several studies and reviews surveyed outlined the reasons for this observation; prior knowledge essentially provides a language and a framework for students to build upon, whereas students who do not have prior knowledge have to first establish that framework. Additionally, students with a prior knowledge may have a greater confidence in a subject and approach it with a more positive attitude. The debate with regards to prior knowledge is based on the constructivist perspective, and discusses whether incorrect prior knowledge has a negative impact on students' performance. While there appears to be some cases and arguments where this can be the case, prior knowledge is thought to have an overall positive effect, given the already mentioned concepts of framework, confidence to topic, as well as isolated issues such as ability to approach text books, ability to take notes which have an indirect effect on performance. Additionally, a tutor's knowledge of misconceptions and their effective isolation and correction has been shown to be a powerful teaching strategy.

The context of the current work is to examine the role of prior knowledge in a system which has a year 1 system that aims to bring all students up to an equivalent level to approach senior years of the degree with equality in content knowledge and understanding of chemistry, at a syllabus level at least. The nature of the study is both powerful and limited. Its power lies in the fact that the analysis is based on above three quarters of the population studying for this programme, and therefore statistical analysis and conclusions can provide powerful statements, both to tutors and to academic management, in the case of arguing, for example, for additional support for students considered to be at risk of underperforming. The limitations are that while the general nature of the outcomes makes for useful summary statements and generalisable approaches (favoured by managers and decision makers) and are the strength of this analysis, it is also, echoing the quote at the start of the text, a great weakness. Students approach college with a range of different inherent motivations, reasons and individual stories that can never be captured in the overall picture that is presented in this work. Time and resources unlimited, individual case studies of students' stories would provide a interesting progression to this study. However, despite these acknowledged limitations, this work purports to be the first step in an important analysis of the experience of year 1 students at college level chemistry in Ireland, based on their ability to perform at the standards required. Subsequent studies analysing further aspects of this experience are welcomed.

5.2 Variables Studied

An important question in the study of prior knowledge is whether students perform better because of their prior knowledge is more extensive or whether they are better at approaching academic tasks, summarised in literature studies as academic aptitude versus prior knowledge. Therefore in this study, the crucial data was the CAO points of students and their prior knowledge of chemistry. The former reflects students' ability to process examinations across a range (six) of different subjects,[7] in an intense examination environment where memory capacity is arguably the most important skill. Choice of this variable reflects many other studies which use ACT, SAT scores or equivalent to examine students general aptitude, for example those of Thompson and Zamboanga discussed in the literature review (Thompson & Zamboanga, 2003, 2004).

The prior level of chemistry was measured by performance in Leaving Certificate chemistry, scored according to the CAO points that would be awarded for a grade achieved. Dochy and others have stressed the need for appropriate assessment of prior knowledge, and there are several reasons why this choice is appropriate (Dochy et al., 1999). Firstly, the content level of both the year 1 programme (DIT, 2009) and the Leaving Certificate Chemistry Syllabus (DES, 1999) are similar, the former being somewhat more advanced in content. The speed of delivery is of course more advanced at third level. Secondly at early stages in third level, the emphasis (whether academics agree on it or not) in a subject like chemistry is on basic facts and procedures, in a sense developing the basic ‘mental model’ to build future knowledge upon. Again, expectations on how students deal with this knowledge is more advanced at third level. In both these cases, one expects that familiarity with content would assist students in processing slightly more advanced content based upon this basic knowledge. Finally, and importantly for this study, an objective is to build a model upon which a students achievement can be predicted, primarily to identify needs for additional tutorials/support to weaker groups identified by the prediction model. As will be discussed later, two prediction models evolved from this study, but a tutor at the beginning of the year has very little data available to identify cohorts of students who may need additional tutorial/other support. Therefore data such as CAO and level of prior knowledge provide readily accessible, externally validated data on which such predictions can be made for students completing the course *in the future*.

In order to facilitate a more in-depth analysis of the sample of students being studied *retrospectively*, in this work, a range of other data was collected as indicated in the results chapter. These consisted of background factors including CAO, the distance students commute to college, their age and gender. Age was not a useful factor in this study as most students were of a similar age, and so it was a poor discriminatory variable. Commute distance was surprisingly long, on average, with average commute time being one hour.

[7] Students may (and usually do) choose more than six subjects, but the CAO score is computed from the best six marks achieved.

The average CAO score for students in the study was 414. In 2008, 55.5% of the Leaving Certificate cohort (52,144 students) scored points of 300 or greater, with 29.7% scoring 400 points or greater (Childs, 2008). As mentioned in the results chapter, falling points for science courses over the last five years gave a broad range of CAO data scores, with an upper limit of 555 points and a lower limit of 300, (Access students are required to have minimum subject requirements but do not need to obtain minimum points). This range makes for a good statistical dataset. Some criticisms of previous studies on prior knowledge were based on the narrow range of the profile of students sampled (Dochy et al., 1999). A striking result from the analysis of CAO points was that students who had studied chemistry at Leaving Certificate level obtained a significantly higher CAO score than those that did not, probably due to the perceived notion that chemistry is a difficult subject and therefore students may be reluctant to choose to study it at this level. The percentage of students choosing higher level Chemistry and achieving an honours grade is 78% averaged for the years 2005 – 2008, with 22% obtaining an A grade, 32% obtaining a B grade and 24% obtaining a C grade (Childs, 2008). This compares with an average of 72% for biology and 71% for Physics over the same time period.

This immediately raised a challenge to the research in that if any correlation was observed between prior knowledge of chemistry and year 1 performance, could this bias be attributed to the fact that these students were more capable at performing well in this type of assessment anyway. This again reverts to the question persistent in prior knowledge debate: academic aptitude versus prior knowledge. This question is emphatically answered in the regression analysis in this study which finds that prior knowledge is a significant and important predictor of performance, even when CAO performance is considered.

The second block of data gathered was a survey of students querying their impressions on their level of study, attendance and interest when they were in year 1. The validity questions surrounding these data have been acknowledged in the Results section, but the data, even considering these issues makes for interesting analysis. Each of the three variables, in each of the years studied show a strong positive bias towards interest and attendance, with a reasonable approximation of a normal distribution for the attitude to study. The latter point encourages confidence in this data. Apart from the averages of the scores, the positive bias indicates something much more important; that students on the course are in general motivated and interested in the course and engage well. Therefore for this group of students, in this context, issues such as motivation may not be as pertinent. A formal attitudes to study survey or motivational study may be relevant in future analyses, but it is the conclusion in this study that correlation and prediction can be confidently conducted knowing that there is a general positive bias to engagement and attendance.

The laboratory score was used as an indirect measure of course engagement. The laboratory programme in Year 1 is highly structured and prescribed and good marks are awarded for attention to detail and hard work in compiling the lab notebook. Therefore it is argued here that diligent, involved students will score well in the laboratory programme, and therefore it is a useful score of course involvement. Reflecting the student survey data, above, there is a positive bias to this data (the range of lab data is 52 to 85 with a mean of 73 compared to the range of year 1 exam scores which is 23 to 96 with a mean value of 59). In comparing the correlation of semester 1 and semester 2 with year 1 performance, which are both strong, with the correlation of the lab mark with year 1 exam performance, which is modest, it is clear that the lab mark reflects something different than just the ability to know about the chemistry content. Therefore it is considered an appropriate score for course engagement.

The final block of data was the performance during the year and the end of year mark. The former was measured by semester tests. The semester 1 test is usually held in the second half of semester 1, and it is usually observed that students who do not have prior knowledge do not perform as well in this test, as they are just getting to know the subject. This was borne out by the correlational analysis, which found that prior knowledge and semester 1 performance were strongly correlated. Similarly, the semester 2 test was strongly correlated with prior knowledge although the effect was slightly less pronounced as for semester 1. This may be due to the fact that students are becoming more familiar with the subject (an internal prior knowledge effect!) and therefore there is less of a difference between students with and without prior knowledge. There are of course other factors such as coming to terms with college life and subsequent challenges, although one can assume that these factors equally affect both sets of students.

The year 1 exam score was the variable that was used for prediction purposes. This was chosen above the module mark, in that the module mark incorporates the lab and assessment mark, so there would be internal conflicts in the analysis. Until Sept '08, there was no minimum mark required for the exam component of the module, (there was a minimum of 35% on the combined exam-assessment component, appropriately weighted). The situation from Sept '08 is that students are required to achieve 35% in the exam component. However, for this data set this means that students may have failed their exam component but still progressed into year 2 because of the positive influence of their assessment score. As such, progression rates approached 100% and were not a suitable score to measure success.

5.3 Correlational Studies

As stated above, the inherent motivation for this research is to identify 'flags' which will indicate whether a student needs support based on the score of any of the variables studied. Correlational analysis provides a useful indication whether any two variables are interlinked to a significant extent. The correlation matrix presented in the Results chapter summarises the work on this element of the research. Hidden amongst the numbers are some useful findings.

The above discussion on the nature of students studying chemistry doing better in year 1 because they have on average better exam performance is answered somewhat by the correlation analysis. A simple correlation between CAO score and year 1 performance returns a modest to moderate correlation between the two variables. However, when the correlation is examined for those with and without prior knowledge, it is found that there is a strong correlation between those with prior knowledge but none for those without. This important result indicates that it is prior knowledge of chemistry rather than general study aptitude as measured by CAO that is influencing year 1 performance. If the reason was due to general study aptitude, similar correlation would be expected for both groups; *i.e.* better students do well. The point is further analysed in the regression analysis. This is perhaps the most significant finding from the correlation work.

Furthermore, the influence of prior knowledge is provided by the correlation analysis between prior knowledge and semester tests and year 1 performance. The semester tests and year performance are all strongly inter-correlated, the strongest observed on the table apart from the (obvious) correlation between prior knowledge and CAO. These results are unsurprising and indicate that students who perform well in semester tests also perform well in the end of year exam. However, the correlation between prior knowledge and these factors strongly suggests that students who have prior knowledge are more likely to do well in these tests. The simple comparison of mean year 1 scores between both groups which shows a significant difference is further evidence on the role of prior knowledge in influencing performance. The fact that data is not distinguished by gender is contrary to what is observed at Leaving Certificate level (HEA, 2007) but encouraging.

As mentioned above, the correlation with lab mark is much weaker, an interesting observation which is probably attributable to a number of factors. Among these are that labs are continuously assessed in a supportive environment where assistance can be sought as required. In addition, the structured lab programme is delivered to a wide range of students from diverse academic backgrounds on different programmes, and hence it is unsurprising that prior knowledge is not a strongly contributing factor in lab performance. Certainly the support and capabilities of the lab supervisors have a role to play in facilitating students without prior knowledge performing equally well in their assessment!

Of interest to any Dublin based institution, there is a modest significant negative correlation between commute distance and year 1 performance. However, this observation should be taken in light of the fact that neither semester test is influenced by distance of commute, which somewhat diminishes the significance of this finding. Distance of commute was found to be a factor in the regression analysis, below, but as is described, can probably be confidently omitted from the model.

Finally in the correlation analysis, the student survey data inter-correlates with modest to moderate stength. The interest component of the survey correlates with both year 1 performance and semester 1 test, while the level of study correlates with exam performance. Level of attendance does not correlate with any factor in the table. The attendance levels in the current year of study were 76.7%, which reflects well on the students' responses from previous years that 80% perceived their attendance to be very good or good.

5.4 Regression Analysis

For the amateur statistician, like the author, regression analysis is a pot of gold at the end of a rainbow that is partially obscured by clouds. It offers untold insight to data and hidden meaning and understanding resulting in new interpretations beyond the relative simplicity of correlation giving the ultimate desire: the full picture. In reality, regression is a reality check. With its power lies some warning. Entering all of the variables and expecting an all-knowing equation to emerge is, as the author has experienced, not what happens. In short, multiple regression should not be used "as a fishing expedition" (Pallant, 2007).

The approach taken here is hierarchical linear regression. This analyses variables in a stepwise fashion. The first variable (or block of variables) is considered and a regression performed. The second step incorporates the second variable(s) and regresses *taking into account the first* *step.* The process continues for as many steps as are entered. Herein, four steps are used, considering the approach by Thompson, who performed similar analysis (Thompson & Zamboanga, 2004).

The key element here was a consideration of the factors that should go into the regression equation, to avoid the so-called fishing expedition. In approaching this, the rationale for the research was recalled. The aim is to examine whether prior knowledge is a factor in student performance, which it is according to correlational studies and to examine how students 'at-risk' of under-performing can be identified early on for tutorial support. Therefore the factors for inclusion in the regression model were ones that would be available to the tutor. As such, the model considered the variables CAO and distance of commute, available from the induction day, and indicators of background, prior knowledge as measured by Leaving Certificate Chemistry, course engagement as measured by lab score based on the rationale above, and course performance as measured by semester tests.

The regression analysis is presented in the Results chapter. The first question asked is an extension of the correlation work – is the CAO score the sole predictor of academic achievement or is prior knowledge a factor? The answer is emphatically answered by the analysis. In the first step of the model, CAO is observed to be an adequate predictor of year 1 performance. However, when prior knowledge is incorporated in the second step, taking into account the influence of CAO, it is observed that prior knowledge that is the significant factor. This is reflected in the high degree of significance for prior knowledge in the model, the increase in the goodness of fit value (R^2) and the relative weightings of the standardised beta coefficients. Incorporating the lab score does not improve this model; the goodness of fit remains relatively unchanged as do the beta coefficients.

It is the final stage in the model which offers some surprise. In this step, the effect of prior knowledge is not insignificant, and instead the emphasis in the model lies with semester test scores, both of which significantly and largely effect the year 1 score, with an accompanied goodness of fit increase.

These observations are in line with the correlational analysis and are explained thus. The correlation study indicated that it was prior knowledge rather than CAO performance that influenced year 1 performance, with the important result that year 1 was correlated with CAO for those who had prior knowledge only. Furthermore, the weaker correlation between year 1 and lab score is borne out by the regression, which shows it to be a poor predictor of end of year scores. The stark effect of the semester tests on the regression model is probably due to the fact that prior knowledge correlates strongly with both semester tests, and these correlate in turn with year 1 performance. Therefore the influence of prior knowledge on year 1 is effectively already incorporated in the semester 1 and 2 test scores. This is called multi-collinearity and occurs when highly correlating variables are included together in regression analysis (Bartholomew et al., 2008).

Two models for implementation evolve out of this data. The first is for the year tutor early in the year and is based on steps 1 and 2 of the regression model, using data that is known on student entry. Predictions on students' performance at the end of the year (and implicitly, therefore, difficulties in the material covered) can be identified by the regression model:

$$[\text{Year 1}] = 37.1 + 0.037[\text{CAO}] - 3[\text{Distance}] + 0.227[\text{Prior Knowledge}]$$

As discussed earlier, the standardised beta variables (*i.e.* those that take into account the magnitude of the variable they are multiplying) show that prior knowledge is almost four times the size of the CAO variable, indicating its importance. There is some merit in considering the role of distance in the model. Commute time data is very easily obtained from induction day surveys and so is of use to a year tutor. Additionally, distance is not specific to either group, so inclusion is relevant to both groups.

The second model that arises out of the data is one that can be used when the most of the year 1 data is known, (for example prior to semester 2 test) and is based on the third and fourth steps of the regression analysis. Two considerations are worthwhile here. The first is that the lab mark is not necessary for the model. As discussed in the results chapter, the laboratory score was found not to significantly affect the regression model. The second is that the semester tests can be considered to account for the prior knowledge bonus and so prior knowledge can be omitted from the model. This was tested as described in the Results chapter, to afford the following regression model:

$$[\text{Year 1}] = 0.046[\text{CAO}] + 0.352[\text{Sem 1}] + 0.375[\text{Sem 2}]$$

(The intercept value (a) is so small in this model it is omitted). This provides the year tutor with a model which can inform the student after the semester 1 test of the support or work that may be required to prepare for semester 2 test. The point to be emphasised with this model is that although prior knowledge is not explicitly in the model, each of the semester tests are assumed to depend on prior knowledge.

In order to check the validity of these models, the available data from previous years was retrospectively incorporated into each equation, to see, had that data been available at the time, would it have correctly predicted the students' scores. For the first model, 53% of student scores were predicted to within 10% of the year 1 score. From analysis of the data that is not within the 10% range, it would appear that students with prior knowledge and grades at the lower end of the range (less than 50 points) are being over-predicted, whereas students with higher grades are being under-predicted. This would indicate that the prior knowledge score is not linear, and that weighting would have to be applied, based on analysis of a larger dataset to the one available. There are some additional individual circumstances which can be surmised from looking at the raw data – namely students lacked interest and dropped out in subsequent years, students worked extremely hard and performed much better than predicted, which contribute to the number of students outside the 10% threshold. Nevertheless, the model has merit in predicting, often with great precision, the scores of over half the cohort of students. In terms of its practical use, it is recommended that until a large data set is available to facilitate non-linear modelling, this model is used to determine prediction scores. Students who are predicted to be in the low grades, or students who are predicted to be in the 50 – 60% range for their year 1 performance, and who have a prior knowledge score of 50 or less, would warrant attention regarding support or advice (see below).

The second model fares a little better, which is unsurprising as it uses actual year 1 performance scores. This predicts 65% of students scores within 10% of the final year score. There are a more diverse range of reasons for the incorrect predictions here, but the more regular ones appear to be that students improved significantly between their semester 2 and year 1 exam score, and students CAO mark over-estimated their performance in year 1. Again, the model is a useful one to flag to students who may be in potential difficulty after their semester 2 test that some additional study or resources are required!

As a caveat to these models, and to show the individual nature of students that cannot be summarised by statistics, the best students of the year in two out of six years did not have prior knowledge. These individuals demonstrate that hard work can override any statistical conclusion!

5.5 Subsequent Years Analysis: Years 2 – 4

The finding that students were not distinguishable in later years was a surprising one. This raises several issues. It would appear from analysis of individual trends in the data set that students subsequent year grades do not necessarily evolve from what would be expected from their CAO and/or prior knowledge (Leaving Certificate Chemistry score). In other words, while the year 1 score is predicted reasonably well, there are a lot of changes happening for the student in this year. These may include desire to work harder and hence achieve well, de-motivation because of incorrect course choice/unhappy with direction of course/unhappy with move from family home and hence disengage from the course; taking a "year off" after a hard year's study for the Leaving Certificate[8], so that year 1 marks may not match expectations, and so on. The point is that year 1 now becomes prior knowledge for year 2, and hence what was a mixed deck coming in is now reshuffled, so that the initial sorting criteria which worked reasonably well in year 1 do not have the same discriminating power in year 2 and subsequent years. Evidence for this is the fact that each of the year's scores correlate strongly with each other, much more so than any year after year 1 correlates with prior knowledge.

Another proposal is the question of what actual value year 1 has in subsequent years. The process of modularisation has until this year left year 1 unscathed – the 15 ECTS module in Introductory Chemistry was a mechanism for keeping the status quo of the old system in the modularised environment.[9] However, years 2 to 4 are fully modularised, and each module is individually examined and assessed. Year 2 lecturers generally start from scratch, and assess their own material. An interesting consideration for further study would be to examine the actual value of Year 1 to subsequent years, apart from bringing students to a relatively equal knowledge base.

5.7 Outcome and Reflections on Study

This study has found that there is a clear relationship between students who have prior knowledge and their performance in year 1 exams. The role of prior knowledge is of premium importance above all other factors, including CAO points. However, this accrued benefit to students who have prior knowledge in year 1 is diminished in subsequent years. This may be due to the fact that students have a more equal level of knowledge framework in year 2, or

[8] Anecdotally, the author has heard this several times from students each year.
[9] From Sept 2009, this module will be delivered as a 5 ECTS and a 10 ECTS module.

may be a consequence of the modularisation system assisting packaging of content knowledge, diminishing the need for core fundamentals in subsequent years.

Two prediction models were developed and have reasonable (>52% and >64% respectively) success at predicting student grades within 10% of their year 1 mark. Reasons for a large range of results being outside the 10% threshold are both technical and individual. At a technical level, the CAO score system for prior knowledge is treated as a linear score, but evidence herein suggests that this may not be appropriate, especially between higher and ordinary level. In future work, the lower and higher marks may need to be weighted accordingly.

In terms of the factors considered and not, CAO and prior knowledge score are useful data to accumulate annually to monitor students' progress. The distance term was the subject of much internal debate, and debate with the referees for the journal article accepted on some of this work. The argument against considering distance was that it did not distinguish between groups, but I wished to incorporate it to acknowledge that students have substantial commute times, and that this is slightly to the detriment of their college performance. The laboratory score was not a useful score, perhaps because individual laboratory classes can vary widely.

Several of the previous studies on prior knowledge incorporate a Maths SAT score solely or in combination with some other factors. In future, it would be worthwhile to include a similar term (*e.g.* Leaving Certificate Mathematics) to examine its roll in the prediction of subsequent performance.

There are several options for the next stages for this research. The first would be to use this work as a basis for qualitative analysis – to examine by case-study students from various categories (high CAO, no prior knowledge; low CAO, prior knowledge, *etc*) to study in depth how their prior knowledge assists them throughout the year. More immediately, there are possibilities for incorporating remedial introductory classes or additional tutorials for students who do not have prior knowledge, and examine the effect on the mean value of these students' year 1 scores compared to those that do have prior knowledge. This is the main motivation for this work, and has been for several other studies discussed in the literature review which describe the positive impact of such programmes, except in a number of cases – *e.g.* (Bentley & Gellene, 2005). Finally, analysis of students by means of case study in subsequent years would allow for a more in-depth discussion on the role of prior knowledge for those years (*e.g.* year 1 performance for year 2), which in itself would lend to a discussion on the pedagogic merits of modularisation.

References

Addison, P. A., & Hutcheson, V. K. (2001). *The importance of prior knowledge to new learning.* Paper presented at the 10th Annual Teaching Learning Forum, Perth: Curtin University of Technology.

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain specific and strategic knowledge in academic performance. *Review of Educational Research,* **58**(4), 377-404.

allpsych.com. (2008). Psychology Classroomat AllPsych Online www.allpsych.com. Retrieved December, 2008

Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. I. (2008). *Analysis of Multivariate Social Science Data*. Boca Raton: CRC Press.

Bentley, A. B., & Gellene, G. I. (2005). A Six Year Study of the Effects of a Remedial Course in the Chemistry Curriculum. *Journal of Chemical Education,* **82**(1), 125 - 130.

BERA. (2004). *Revised Ethical Guidelines for Education Research*. Southwell.

Bloom, B. S. (1964). *Stability and Change in Human Characteristics*. New York: Wiley.

Bloom, B. S. (1976). *Human Characteristics and School Learning*. New York: McGraw-Hill.

Bodner, G. M., & Orgill, M. (2007). *Theoretical Frameworks for Research in Chemistry/Science Education*. New Jersey: Pearson Education Inc.

Boujaoude, S. B., & Giuliano, F. J. (1991). *The relationship between students' approaches to studying, formal reasoning ability, prior knowledge and bender and their achievement in chemistry*. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching.

Braathen, P. C., & Hewson, P. W. (1988). *A case study of prior knowledge, learning approach and conceptual change in an introductory college chemistry tutorial programme*. Paper presented at the 61st Annual Meeting of the National Association for Research in Science Teaching.

Bunce, D. M., & Hutchinson, K. D. (1993). The use of GALT (Group Assessment of Logical Thinking) as a Predictor of Academic Success in College Chemistry. *Journal of Chemical Education,* **70**(3), 183 - 187.

Byrnes, J. P., & Guthrie, J. T. (1992). Prior Knowledge and Textbook Search. *Contemporary Educational Psychology,* **17**, 8-29.

CAO. (2008). Central Applications Office www.cao.ie. Retrieved October, 2008

Carmichael, J. W., Bauer, J., Sevenair, J. P., Hunter, J. T., & Garnbrel, R. L. (1986). Predictors of First-Year Chemistry Grades for Black Americans. *Journal of Chemical Education,* **63**(4), 333 - 336.

Chandran, S., Treagust, D. F., & Tobin, K. (1987). The role of cognitive factors in chemistry achievement. *Journal of Research in Science Teaching,* **24**(2), 145-160.

Chandrasegarana, A. L., Treagust, D. F., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice,* **8**(3), 293 - 307.

Childs, P. E. (2008). 2008 LC Results and CAO Points. *Chemistry in Action,* **86**, 35 - 43.

Chittleborough, G. D., Treagust, D. F., & Mocerino, M. (2002). *Constraints to the development of first year university chemistry students' mental models of chemical phenomena.* Paper presented at the 11th Annual Teaching Learning Forum, Perth: Edith Cowan University.

Condron, J. (2008). DIT Statistical Advice Unit Personal Communication.

Craney, C. L., & Armstrong, R. W. (1985). Predictors of Grades in General Chemistry for Allied Health Students. *Journal of Chemical Education,* **62**(2), 127 - 129.

Creswell, J. W. (2008). *Educational Research: Planning, conducting, and Evaluating Quantitative and Qualitative Research*. New Jersey: Pearson Education.

Creswell, J. W. (2009). *Research design : qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks: Sage Publications.

DES. (1999). Department of Education and Science Leaving Certificate Chemistry Syllabus: http://www.education.ie/servlet/blobservlet/lc_chemistry_sy.pdf. Retrieved December, 2008

DIT. (2009). *DT203 Programme Document*.

Dochy, F. J. R. C. (1992). *Assessment of Prior Knowledge as a determinant for future learning: The use of prior knowledge state tests and knowledge profiles*. London: Lemma B.V.

Dochy, F. J. R. C. (1994). *Prior knowledge and learning* (2nd ed.). Oxford/New York: Pergamon Press.

Dochy, F. J. R. C., & Alexander, P. A. (1995). Mapping Prior Knowledge: A Framework for Discussion Among Researchers. *European Journal for Psychology of Educational Psychology,* **10**(1), 123 - 145.

Dochy, F. J. R. C., De Ridjtt, C., & Dyck, W. (2002). Cognitive prerequisites and learning. How far have we progressed since Bloom? Implications for educational practice and teaching. *Active Learning in Higher Education,* **3**(3), 265-284.

Dochy, F. J. R. C., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research,* **69**, 145-186.

Education-Encyclopedia. (2009). B. S. Bloom (1913 - 1999) http://education.stateuniversity.com/pages/1791/Bloom-B-S-1913-1999.html. Retrieved January, 2009

Etta-AkinAina, F. E. (1988). *Note-taking in lectures: the relationship between prior knowledge, information uptake and comprehension.* University of Edinburgh.

Fidler, F., & Cumming, G. (2008). The New Stats: Attitudes for the 21st Century. In J. W. Osborne (Ed.), *Best Practices in Quantitative Research*. Thousand Oaks: Sage.

Hailikari, T., & Lindblom-Ylanne, S. (2007). Exploring alternative ways of assessing prior knowledge, its components and their relation to student achievement: a mathematics-based case study. *Studies in Educational Evaluation,* **33**(3-4), 320-337.

Hailikari, T., Nevgi, A., & Komulainen, E. (2008). Academic self-beliefs and prior knowledge as predictors of student achievement in Mathematics: a structural model. *Educational Psychology,* **28**(1), 59 - 71.

Harrison, A. G. (2003). *Inquiry Learning, Modelling and a Philosophy of Chemistry Teaching*. Paper presented at the Australian Association for Research in Education—New Zealand Education Research Association.

HEA. (2007). Sé Sí: An overview of the issue of gender in education in a life long context http://www.hea.ie/index.cfm/page/news/category/134/section/details/id/324. Retrieved September, 2008

House, J. D. (1995). Noncognitive Predictors of Achievement in Introductory College Chemistry. *Research in Higher Education,* **36**(4), 473 - 490.

Jonassen, D. H., & Grabowski, B. L. (1993). *Part VII, Prior Knowledge*. Hillsdale: Lawrence Erlbaum Associates.

Kinnear, P. R., & Gray, C. D. (2006). *SPSS 14 Made Simple*. New York: Psychology Press.

Lewis, S. E., & Lewis, J. E. (2007). Predicting at-risk students in general chemistry: comparing formal thought to a general achievement measure. *Chemistry Education Research and Practice,* **8**(1), 32 - 51.

Muijs, D. (2004). *Doing quantitative research in education with SPSS*. London: Sage.

O'Reilly, T., & McNamara, D. S. (2007). The Impact of Science Knowledge, Reading Skill, and Reading Strategy Knowledge on More Traditional "High-Stakes" Measures of High School Students' Science Achievement. *American Educational Research Journal,* **44**(1), 161-196.

Osborne, J. W. (2008). Using Best Practices is a Moral and Ethical Obligation. In J. W. Osborne (Ed.), *Best Practices in Quantitative Methods*. Thousand Oaks: Sage.

Pallant, J. (2007). *SPSS Survival Manual*. New York: Open University Press.

Palmer, M., O'Kane, P., & Owens, M. (2009). Betwixt spaces: student accounts of turning point experiences in the first-year transition. *Studies in Higher Education,* **34**(1), 37 - 54.

Reid, N. (2006). *Getting Started in Pedagogical Research in the Physical Sciences: A Physical Sciences Practical Guide*. Hull: HEA Physical Science Centre.

Russell, A. A. (1994). A Rationally Designed General Chemistry Diagnostic Test. *Journal of Chemical Education,* **71**(4), 314 - 317.

Russell, M. (2004). *Academic Success, Failure and Withdrawal amongst 1st Year Engineering Students: Was Prior Mathematical Knowledge a Critical Factor?* , University of Dublin: Trinity College.

Smith, O. M., & Trimble, H. M. (1929). The prediction of future performance of sutdents from the past records. *Journal of Chemical Education,* **6**(1), 93 - 97.

Spencer, H. E. (1996). MAthematical SAT scores and College Chemistry Grades. *Journal of Chemical Education,* **73**(12), 1150 - 1153.

Taylor, A. K., & Kowalski, P. (2004). Naive psychological science: The prevalence, strength, and sources of misconceptions. *Psychological Record,* **54**, 15-25.

Thompson, R. A., & Zamboanga, B. L. (2003). Prior knowledge and its relevance to student achievement in Introduction to Psychology. *Teaching of Psychology,* **30**, 96 - 101.

Thompson, R. A., & Zamboanga, B. L. (2004). Academic Aptitude and Prior Knowledge as Predictors of Student Achievement in Introduction to Psychology. *Journal of Educational Psychology,* **96**(4), 778 - 784.

Treagust, D. F. (2008). *Embedded, formative and diagnostic assessment in science to engender deeper conceptual understanding*. Paper presented at the SMEC: Science and Mathematics Education Conference.

Youl, D., Read, J., George, A., Masters, A., & Schmid, S. (2005). *Bridging the gap - student understanding and the chemistry bridging course*. Paper presented at the Uniserve Science Blended Learning Symposium.

Appendix

Appendix 1: Student Perception Survey

Dear Student,

I am completing an MA in Education and as part of my research, I am looking into the effect of prior learning of chemistry (i.e. Leaving Certificate Chemistry) has on the performance of students in Year 1, as compared to other factors (CAO entry points, home address, term address, etc).

I would be grateful if you could complete the attached questionnaire which should take less than 5 minutes.

The results of this survey will be compiled on a statistical basis and no individual student will be identified in any way. If you have any questions about the dissemination or analysis of the results, please feel free to discuss with me before completing the form. I ask you to put your name on the form so that I can cross-reference with other data, such as end of year mark in Year 1 chemistry.

If you have any questions about the project prior to completing the questionnaire, you can ask me or contact my supervisor, Dr Brian Bowe, whose contact details are available on the form. If you wish to find the outcomes of the study, come and see me in June '09!

Michael Seery

michael.seery@dit.ie

Supervisor:

Dr Brian Bowe

Learning and Teaching Centre

DIT Mount St

Dublin 2

Brian.bowe@dit.ie

1. Name: _______________ _______________

2. Course: DT_______ Current Year of Study: 1☐ 2☐ 3☐ 4☐

3. Did you study chemistry at Leaving Cert level (if PLC see 5 below): Yes ☐ No ☐

   If Yes please state:

   a. Level: Higher ☐ Ordinary ☐

   b. Grade: Grade:____

If you can't remember exact numbers or answer, put a ~ symbol before your answer and I will try to cross-check!

4. How many points did you score in your Leaving Cert (best 6 subjects): _______

5. Did you complete a PLC or equivalent prior to your course in DIT: Yes ☐ No ☐

   If Yes please state:

   a. Name of PLC/equivalent: ____________________________________

   b. Did this course include Chemistry: Yes ☐ No ☐

   c. What grade was obtained: Grade:____

6. What is your home address (town, county only) for the duration of your first year:
   (*e.g. Blanchardstown, Dublin 15; Cahirciveen, Co. Cork)*

7. What is your term-time address (town, county only) for the duration of your first year:
   (Tick if same as home address ☐)

8. How long did it take to commute each way to college from your ***term-time*** address?
   ____

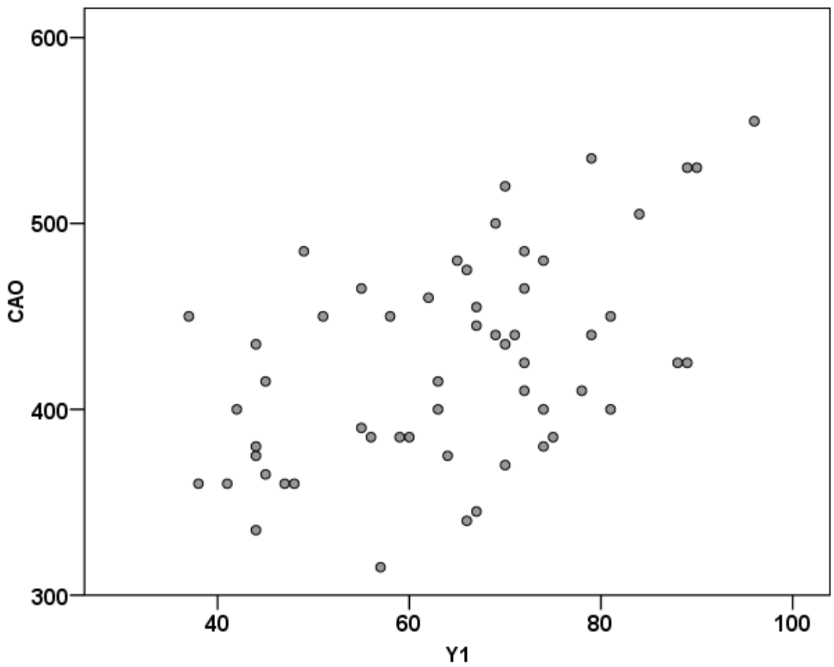
9. Would you consider **your attendance** in Year 1 as:
   Very Good ☐ Good☐ Average☐ Poor☐ Very Poor☐

10. Would you consider **your level of study** in Year 1 as:
    Very Good ☐ Good☐ Average☐ Poor☐ Very Poor☐

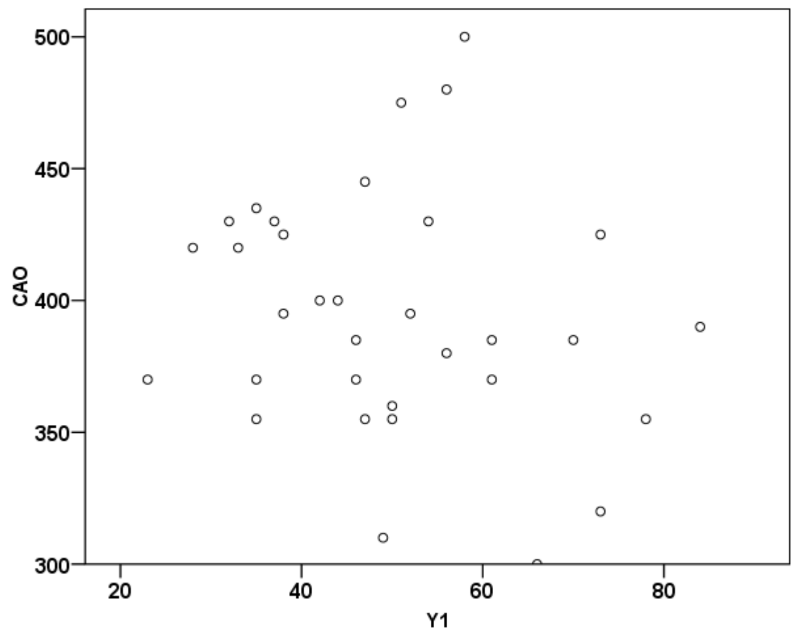
11. Would you consider **your interest** in the course in Year 1 as:
    Very Interested ☐ Interested☐ Average Interest☐ Little Interest☐ No Interest☐

This survey is confidential and no names or identifying details will be used in the results of the survey.
Thank you for taking the time to complete the survey! michael.seery@dit.ie

Appendix 2: Scatter Plots of CAO – Y1 separated for PK = 0 and PK = 1



Scatter-plot of CAO scores and year 1 performance for students with prior knowledge of chemistry



Scatter-plot of CAO scores and year 1 performance for students without prior knowledge of chemistry

Appendix 3: Diagnostic Assessment

1. Element C with electronic configuration $1s^2$, $2s^2$ $2p^6$ $3s^2$ and Element E with electronic configuration $1s^2$, $2s^2$ $2p^5$ reacts to form an ionic compound CE_2.

    1. True
    2. False

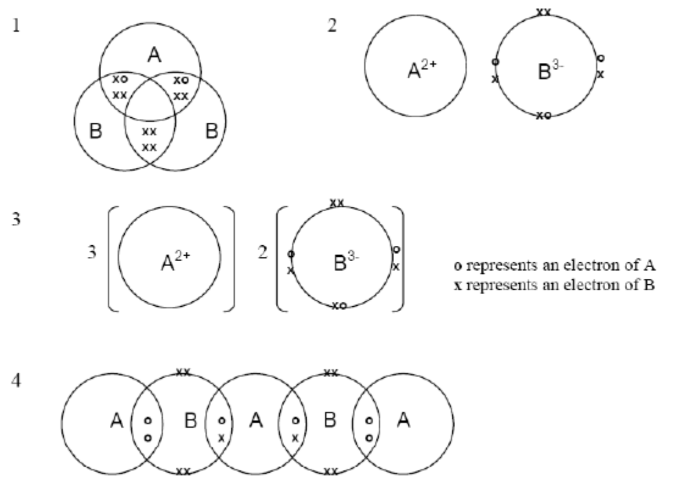
    REASON:

    a. An atom of C will share one pair of electrons with each atom of E to form a covalent molecule, CE_2.
    b. A macromolecule is formed consisting of covalently bonded atoms of C and E.
    c. Atoms of C will each lose two electrons and twice as many atoms of E will each gain one electron to form an ionic compound CE_2.
    d. An atom of C will lose one electron to an atom of E to form an ionic compound CE.

2. An atom of element A has two electrons in its outermost shell while an atom of element B has five electrons in its outermost shell. When A reacts with B, the compound will be:

    1. Covalent
    2. Ionic

    REASON:



3. Water (H_2O) and hydrogen sulfide (H_2S) have similar chemical formulae and structures. At room temperature, water is a liquid and hydrogen sulphide is a gas. This difference in state is due to:
   a. Forces between molecules
   b. Forces within molecules

   REASON:

   1. The difference in the forces attracting water molecules and those attracting hydrogen sulfide molecules is due to the difference in strength of the O-H and the S-H covalent bonds.
   2. The bonds in hydrogen sulfide are easily broken, whereas those in water are not.
   3. The hydrogen sulfide molecules are closer to each other, leading to greater attraction between molecules.
   4. The forces between water molecules are stronger than those between hydrogen sulfide molecules.

4. Sodium chloride, NaCl, exists as a molecule.
   1. True
   2. False

   REASON:

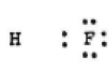
   a. The sodium atom shares a pair of electrons with the chlorine atom to form a simple molecule.
   b. After donating its valence electron to the chlorine atom, the sodium ion forms a molecule with the chloride ion.
   c. Sodium chloride exists as a lattice consisting of sodium ions and chloride ions.
   d. Sodium chloride exists as a lattice consisting of covalently bonded sodium and chlorine atoms.

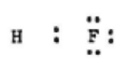
5. Silicon carbide has a high melting point and high boiling point. This information suggests that the bonds in silicon carbide are:
   1. Weak
   2. Strong

   REASON:

   a. Silicon carbide is a simple molecular solid.
   b. Silicon carbide is a macromolecule composed of covalently bonded atoms.
   c. Silicon carbide is a macromolecule composed of covalently bonded molecules.
   d. A large amount of energy is required to break the intermolecular forces in silicon carbide.

6. Which of the following best represents the position of the shared electron pair in the HF molecule:

(1) 

(2) 

   REASON:
   a. Non-bonding electrons influence the position of the bonding or shared electron pair.
   b. As hydrogen and fluorine form a covalent bond, the electron pair must be centrally located.
   c. Fluorine has a stronger attraction for the shared electron pair.
   d. Fluorine is the larger of the two atoms and hence exerts a greater control over the shared electron pair.

7. Nitrogen (a group 5 element) combines with bromine (a group 7 element) to form a molecule. This molecule is likely to have a shape which is best described as:
    1. Trigonal planar
    2. Trigonal pyramidal
    3. Tetrahedral

REASON:

   a. Nitrogen forms three bonds which equally repel each other to form a trigonal planar shape.
   b. The tetrahedral arrangement of the bonding and non-bonding electrons pairs around nitrogen result in the shape of the molecule.
   c. The polarity of the nitrogen-bromine bond determines the shape of the molecule.
   d. The difference in the electronegativity values between bromine and nitrogen determine the shape of the molecule.

8. The polarity of the oxygen-fluorine bond would be best represented as:

(1) $\overset{\delta^-}{O} - \overset{\delta^+}{F}$ (2) $\overset{\delta^+}{O} - \overset{\delta^-}{F}$

REASON:

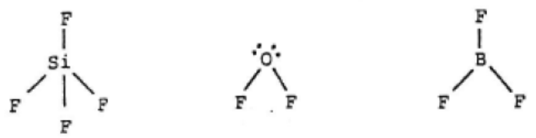
   a. The non-bonding electron pairs on each atom determine the bond polarity.
   b. A polar covalent bond forms as oxygen has six outer shell electrons and fluorine has seven outer shell electrons.
   c. The shared electron pair is closer to fluorine.
   d. The polarity of the O-F bond is due to the oxygen atom forming the O^{2-} ion whereas the flurine atom forms an F^- ion.

9. The molecule SCl_2 is likely to be:
    1. V-shaped
    2. Linear

REASON:

   a. Repulsion between the bonding and non-bonding electron pairs result in the shape.
   b. Repulsion between the non-bonding electron pairs result in the shape.
   c. The two sulfur-chlorine bonds are equally repelled to linear positions.
   d. The high electronegativity of chlorine compared to sulfur is the major factor influencing the shape of the molecule.

10. Which of the following molecules is polar?



(1) (2) (3)

REASON:
   a. The polarity of the molecule is due to the high electrnegativity of fluorine.
   b. Non-symmetrical molecules containing different atoms are polar.
   c. Non-bonding atoms on a molecule produce a dipole and hence are polar.
   d. A large difference in the electronegativities of the atoms involved in bonding result in a polar molecule.