Chapter 3

Analysis of the results

3.1 Core competencies

Before it is possible to compare performance in assessments over the period of the study on an inter-cohort basis, the possible variation in core competencies in the academic subjects studied at A level needs to be considered. These can be measured by an analysis of the second part of the pre-test questionnaire, dealing with technical questions. It may then be possible to determine if there is any relationship between the falling score for the overall test mark, and if one aspect of the test is responsible. The second part of the test itself - see Appendix 1 - covers four basic areas. The main one is electronics, which is itself made up of several areas, such as basic electrical theory, devices and applications. For this analysis all these have been grouped into a single variable. The other three areas are physics, computing and mathematics. The physics questions were mainly concerned with basic physical phenomena such as electromagnetism, electrostatics and dimensions/ units. The computing section was basic binary concepts, whilst the maths was concerned with trigonometric concepts used in electronics. Figures 3.1 to 3.2 show the changes in percentage of the class giving correct responses to each grouping.

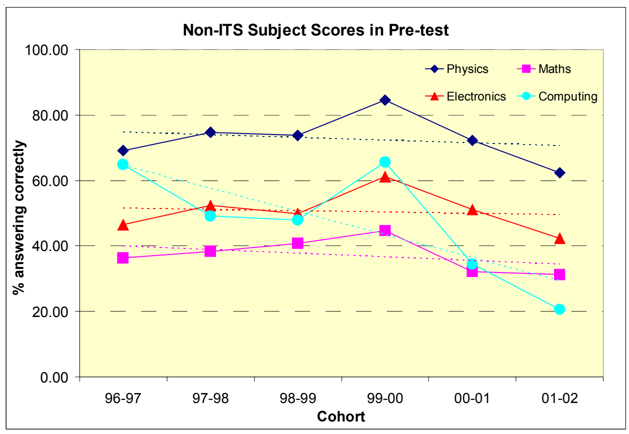

Figure 3.1 Percentage correct responses to the pre-test by subject area of the question for Non-ITS group

As can be seen from Figures 3.1 and 3.2, other than for electronics which remained fairly constant, all the other subject areas showed a decline in correct responses over the period of the study. The surprising decline is in the understanding of basic computing theory, which shows a significant decrease in correct responses. The implications of this are discussed below, as well as in Bradbeer (2002a and 2002b).

In Figure 3.3, both courses are considered together, and the results for the 1999-2000 cohort have been taken out of the trendline, owing to the different nature of the testing conditions, and the resulting anomaly (In 1999-2000 the pre-test questionnaire was administered in a different way to the other years. The instructor allowed discussion amongst the students, so that the result reflects a group response more than an individual response, as in other years).

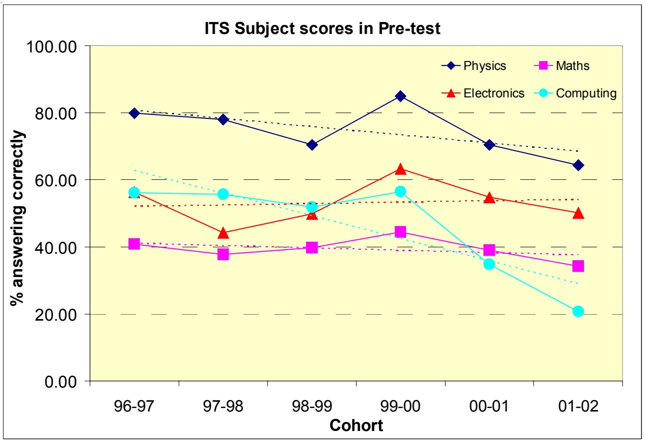

Figure 3.2 Percentage correct responses to the pre-test by subject area of the question for ITS group

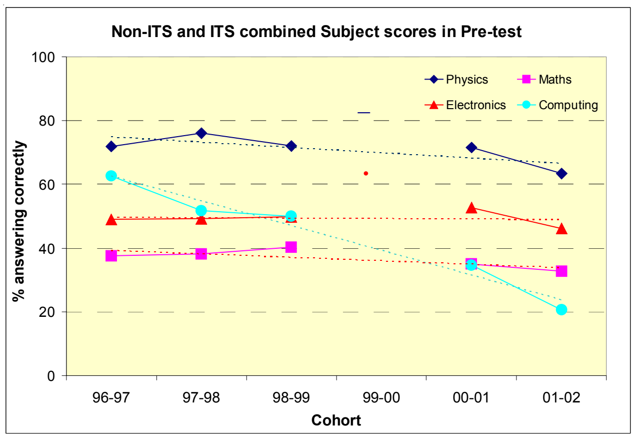

Figure 3.3 Pre-test subject scores for combined results of Non-ITS and ITS groups with anomalous data from 1999-2000 cohort removed.

3.2 Implications for course content

The changing extent of the basic knowledge of new entrants has meant that assumptions made in 1996 cannot be applied in 2002. This has had implications in the syllabus content of the courses themselves. The amount of time needed to cover basic theory that should have been covered in the A level syllabus and was not - or may have been but was not understood by the students - increased substantially over the period of the study. So much so, that the courses themselves have been drastically rewritten to cope with this. From 2002, two completely new courses replaced those taught for the previous 6 years. These now emphasise the design aspects of electronics and not concentrate so fully on analysis.

Also, a more problem-based, student-centred learning approach is being taken, based upon the experiences of using integrated studio teaching for the BEMTE course.

A brief look at the first semester syllabuses for 1996, 2001 and the new 2002 courses will show clearly the changes that have had to be made to accommodate the changing environment.

> 1996: Revision of basic electric and magnetic fields. Inductance, self-inductance, mutual inductance. Transformers; principles of operation and applications. Revision of circuit theorems and laws; Simple dc transient analysis. Revision of ac fundamentals; Phasors and complex numbers. Three phase systems. Revision of basic semiconductor devices; modes of operation. Amplifier circuits. Feedback. Introduction to the operational amplifier. Power amplifiers. The transistor as a switch.
>
> 2001 Basic Magnetic Fields: Revision of basic magnetic laws. Inductance, self-inductance, mutual inductance, Magnetic circuits. Transformers; principles of operation and applications. Basic Electric Fields: Capacitance and capacitors, energy storage in capacitors. DC Circuit Analysis: Revision of circuit theorems and laws. Simple dc transient analysis. AC Circuit Analysis: Revision of ac fundamentals. Phasors and complex numbers; reactance, impedance, power and power factor. Semiconductor Devices: Revision of basic semiconductor devices; pn junction, characteristics of junction diode, diode circuits, bipolar transistors, field effect transistors, modes of operation.
>
> 2002 Circuit analysis techniques, basic discrete semiconductor devices, integrated circuit fundamentals, the transistor as an amplifier, the transistor as a switch.

It can be seen that by 2001 most of the first semester course had been taken up with basic revision of fundamentals, which were assumed to be generally known by the students in 1996, an assumption which could not be made in 2001. A subjective analysis of tests, quizzes and coursework over the period of the study support these assumptions.

The 'knock-on' effects of having to cover fundamentals in the first semester meant that less coverage could be given to the more design aspects of electronics in the second semester and this itself meant that courses taken in the second and third year were also affected, in most cases adversely. This was especially true of the virtual elimination of power electronics from the syllabus. There was also some criticisms that basic electronics courses, which were designed to support the manufacturing and mechatronics engineering programmes should not become applied physics courses!

Another reason for changing the course structure was the change in the BEME programme to a BEng in Manufacturing and Information Systems Engineering (BEMISE). The basic electronics courses for this new programme would become second year electives, not first year core courses. The BEMTE programme was also revised to become more design mechatronics based, although still keeping the basic electronics courses in the first year core.

This led to a complete rethink of what such a basic course in electronics should provide, both as a student learning experience and as a basis for further study in later years. It was decided that most theoretical 'applied-physics' fundamentals should be ignored completely, and that a more systems approach should be taken in teaching the basic electronic circuits. In other words, any design processes should be based around the use of 'black-box' modules, which would correspond roughly to the most popular integrated circuit packages, such as logic gates, operational amplifiers etc.

At the same time, there would be a more 'hands-on' experience with simulation and experiment replacing basic theory. This would seem to be shifting the course more towards the technician engineer pedagogy compared to the more traditional university approach. Also, group based projects would replace more individual learning experiences.

Bradbeer (2003) discusses the implications for the changing course content, and it is clear from the data presented in that paper and also presented above, that there has been a gradual decline in the basic knowledge of physical fundamentals, maths and computing over a six year period from 1996, even though the grades achieved at A level have been rising. This may be a consequence of the Hong Kong government's policy of rapidly increasing the number of university places available for 18 year olds, from around 6% of school leavers to today's 18%, starting in the early 1990s.

It would also seem to indicate that the Hong Kong Examination Authority has condoned a gradual and sustained inflation in grades over that period of time. Also, even though universities and departments are quite happy to publicise the fact that quality of their student entrants is getting relatively better, in fact this hides the fact that, in absolute, terms they are not.

The implications for syllabus and course design are even more profound, and means that a constant and continuing shift in course content and level is needed to give the students a meaningful learning experience that is suited to their level of knowledge. Unfortunately, programme leaders and course lecturers who just take the raw entrant examination grades as an indication of how to 'pitch' their courses are in danger of getting it wrong, with disastrous consequences, which can be seen in the lack of commitment and energy that students have for their studies. This problem will be addressed in greater detail in Chapter 5.

The implications for the intra-cohort comparison is that it would not be possible to compare cohort with cohort over the period of the study as the course content changed during this period. However, there would be few problems with an inter-cohort comparison as the changes were identical for each of the two courses under study.

3.3 The assessments used in the comparison.

Although the students studied two consecutive courses, only a detailed analysis was made of the first one. The main objective of the assessments used was to determine whether there was an educationally significant difference between the performance of the two groups. Many students in Hong Kong have been taught to study in a strategic manner, mainly for the purpose of passing examinations. This is one of the main failings of the government school education system in Hong Kong. At the same time, students are not prepared for a university environment, where a less exam oriented assessment is used, with far more emphasis on project and lab based work than usually found in local secondary schools. The assessment tools designed to determine the learning outcomes also had to be acceptable to the university assessment scheme, which limited the amount of innovation that could be used. Consequently, a series of tests, projects and reports were designed which closely mirrored accepted practice for the programmes, whilst trying to probe more deeply into the learning taking place.

One of the objectives was to establish whether deep or surface learning was taking place. Deep and surface are two approaches to study, derived from original empirical research by Marton and Säljö (1976) and since elaborated by Ramsden (1992), Biggs (1987, 1993) and Entwistle (1981), among others. According to Atherton (2005), it is important to clarify what they are not:

> Although learners may be classified as "deep" or "surface", they are not attributes of individuals: one person may use both approaches at different times, although she or he may have a preference for one or the other.
>
> They correlate fairly closely with motivation: "deep" with intrinsic motivation and "surface" with extrinsic, but they are not necessarily the same thing. Either approach can be adopted by a person with either motivation.

There is a third form, known as the "Achieving" or strategic approach, which can be summarised, as Atherton does (ibid), as "a very well-organised form of surface approach, and in which the motivation is to get good marks. The exercise of learning is construed as a game, so that acquisition of technique improves performance. It works as well as the analogy: insofar as learning is not a game, it breaks down".

Ramsden (1998) summarised the two approaches as follows: Deep learning focuses on "what is signified", relates previous knowledge to new knowledge, relates knowledge from different courses, relates theoretical ideas to everyday experience, relates and distinguishes evidence and argument, organises and structures content into coherent whole and emphasises is internal, from within the student. Surface learning, on the other hand, focuses on the "signs" (or on the learning as a signifier of something else), focuses on unrelated parts of the task, information for assessment is simply memorised, facts and concepts are associated unreflectively, principles are not distinguished from examples, task is treated as an external imposition, and the emphasis is external, from demands of assessment.

Although both Entwistle (1981) and Biggs (1987) had developed methodologies for analysing whether deep or surface learning was taking place, it was not possible to introduce these into the courses as they were structured at the time of the study. From 2005, all first year students at CityU have to take part in a Learning and Study Strategies Inventory (LASSI) survey, based upon the work of Weinstein, Schulte, and Palmer, Ph.D. at University of Texas (Weinstein et al, 2000).

This is beginning to give some of the detailed information needed to make such a comparison, but as they were not able to be used at the time of the study, another approach was needed. Consequently, a simplified approach had to be taken, and this was based around a mixture of multiple choice tests, group based projects (for the ITS group), lab work for the non-ITS group, and descriptive examination questions that probed more deeply. The analysis of the results could then determine whether one group scored higher on the multiple choice questions - possibly indicating more surface learning, as compared to the descriptive questions in the final examination, which hopefully indicated the level of deep learning.

A multiple choice test was given midway through the first semester. This multiple choice test was based upon material taught in the first half of the semester, and consisted of 30 questions, mainly to do with the basics of electricity, magnetism and simple electric circuit theory. The test was the same for each cohort. The answers were not given to the students, just their marks.

At the end of the first semester the students sat an examination which consisted of two parts. The first was a multiple choice section of 25 questions, based on work for the whole semester. This accounted for 25-30% of the final mark. The rest of the exam was a more traditional one, with students have to answer three questions from four in a more descriptive manner. Again, the questions covered the whole of the syllabus.

Three sets of multiple choice questions were used, and rotated on a 3 year pattern, so that no test was used twice during a three year period. To ensure that the standard of the multiple choice test was approximately consistent each year, a number of colleagues compared the papers, and adjustments were made accordingly. As mentioned at the beginning of the chapter, intra-cohort comparison was not an easy task to undertake, owing to the changes in the syllabus during the period of the research. However, as the same test was given to both courses, inter-cohort comparisons could be reliably made. No analysis of the answers was made to determine whether any of the four subject areas tested in the pre-test, and reported above, varied form cohort to cohort.

The final grading for the semester was based upon a combination of coursework, which included two assignments (not part of this study), mid semester test, and laboratory/group work, and examination performance. For the first two years of the study this split was 60:40 examination:coursework changing to 70:30 in the third year. To determine whether there was any relationships between the results, the two examination components, and the final marks including coursework, were kept separately.

Although the second semester course was not studied in such detail, the students sat a mid term test, all questions being descriptive with some calculation, followed by a final exam that was of a more traditional style. They also carried out a number of assignments and laboratory based project work. Again, final grading was based upon a combination of coursework and examination performance, in the same ratio as Semester A. The results of the mid-semester test in Semester B for 1996-97 and 1997-98 were not available for analysis, and have been left out of the data. The results from the assessments in the second semester course are not analysed in great detail, but will be given for completeness only.

Owing to the nature of the course structure, there are always a number of repeat students in each class. These have been eliminated from the analysis. Similarly some students are given exemption from taking the Semester A course. These students have been eliminated from the analysis of the Semester B results.

The results are given as percentages of maximum marks. Grades have not been shown, as the change in course structure affected the grading system but not the marking system. In Semester A, the final mark is shown in three sections - first, the total mark for both coursework and examination, then as the mark for the multiple choice examination, and then for the descriptive examination. In Semester B, the final mark is shown, first as a total for coursework and examination, and then as examination only.

3.4 Implementation of the Integrated Teaching Studio.

One other factor that has to be taken into account when analysing the results is the fact that the construction and inauguration of the ITS was fraught with problems, which resulted in some variation of teaching pattern compared with that originally planned. In the first year, 1996-97, the interface cards to provide the laboratory component of the course were not working properly. It was therefore decided to use a normal laboratory for all experimental based work. In 1997-98, the interface cards provided proved impossible to use for the experimental work in Semester B, so this was also held in a normal laboratory. There have been few problems with this aspect of the implementation since then, and all experimental work for the ITS-based classes have been run as planned.

From 1998-99, the multimedia courseware used in the ITS was also used as part of the presentation graphics in the lecture theatre for the non-ITS class. In 1999-2000 the classes were taught by colleagues as I was on sabbatical leave at Rennsalaer Polytechnic Institute in New York for Semester A. These colleagues did not use studio teaching and taught both classes in traditional mode. Also, each colleague only taught one of the courses, so there were different styles for each course. The results from these classes in semester A have been included in this study as they provide an interesting comparison for the effectiveness of the ITS based teaching, although only the pre-test and final examination have any data that can be useful. In Semester B of that year I only taught the ITS class, and not the non-ITS class. The results for both groups in Semester B 1998-99 have, therefore, been ignored in this analysis.

From the start of the 2000-2001 academic year, the studio has been functioning well, and there have been few problems since then that might affect the overall results.

3.5 Cohort analysis

The results for each assessment in each cohort are shown in Appendix 6. Each year shows the results in three ways, always comparing the inter-cohort analysis. Where available the results compare each of the two groups in each cohort for the following:

Pretest
Mid-semester test Semester A
Final total mark for Semester A end-of-semester assessment (coursework and exam)
Total examination mark for Semester A end-of-semester exam
Multiple choice section mark for Semester A end-of-semester exam
Descriptive section for Semester A end-of-semester exam
Mid-semester test for Semester B
Final total mark for Semester B end-of-semester assessment (coursework and exam)
Examination mark for Semester B end-of-semester exam

The first is a box and whisker plot, with confidence intervals. The second is a table showing the data used to draw the box-plots and parametric statistics for each year. The third is a table showing the calculation of effect size for each type of assessment, or outcome measure. Each cohort will be examined in turn, with a meta-analysis carried out on that cohort's data to determine the overall effect size.

Effect size has been used in preference to the more common methods of statistical significance for three main reasons. First, it is the size of the impact that is of substantive importance, and yet statistical significant testing is dependent on sample size – the same effect acquires statistical significance with larger samples. Secondly, effect sizes can be compared from one study to another and under suitable circumstances can be combined using meta-analysis. Thirdly statistical significance testing employs arbitrary cut-offs. Such issues became apparent during the study. Initially, all the data analysis was based on statistical significance assuming a null hypothesis.This gave some results that were hard to interpret. For example, some of the intra-group comparisons showed a statistical significance with $p<.05$ but only just ($p=.048$) whilst others were not significant ($p>.05$) but only just ($p=.052$).. Is there really any difference, educationally between the two? The use of Effect Size overcomes this problem.

Effect Size is usually defined as the (mean of the experimental group - mean of the control group)/ standard deviation of the control group. However, a number of assumptions are made in this formula. First, the control group is assumed to be large. If it is not, and this is true in this study, a pooled estimate value of standard deviation is used. Secondly, if the population of each group is small, then there will be a bias in the calculation. Thus Hedges correction is used (Hedges and Olkin, 1985). Similarly, it is useful to know the CI of the Effect Size calculated. For the analysis used in this study, an Effect Size calculator devised by Robert Coe (2000) was used. A more detailed explanation is given in Appendix 6.

In the following sections only the standardised Effect Size (with Hedges transformation) will be shown. The treatment group is the ITS group, and the control group is the non-ITS group. Reference should be made to the full data in Appendix 6. In the following analyses, the overall effect size for Semester A and Semester B assessments does not include the pre-test results. As there were no significant differences between the pre-test scores for all but the final year of the study, this appears a rational decision. Appendix 6 also contains data analysis of the confidence interval and p value for each cohort.

3.5.1 1996-97

pretest	mid a	fin a mark	fin a exam	fin a mc	fin a desc	fin b mark	fin b exam
0.18	0.94	1.10	0.95	0.67	0.88	-0.11	0.35

Table 3.1 Standardised Effect Sizes for 1996-97 cohort for each assessment.

It is clear from the data for the pre-test that there is no significant difference between the groups. However there is a significant difference between the two groups for the Semester A mid-test, as well as for all measures of the final Semester A assessment. However, there is little significant differences between the Semester B results.

For Semester B, although the non-ITS group gained higher marks in the overall assessment, the ITS group did better in the examination. The laboratory marks for the ITS group were lower than that for the non-ITS group, which may have been a reflection of the rather chaotic nature of the practical work that year.

3.5.2 1997-98

pretest	mid a	fin a mark	fin a exam	fin a mc	fin a desc	fin b mark	fin b exam
-0.02	0.30	0.36	0.26	-0.09	0.38	0.04	0.06

Table 3.2 Standardised Effect Sizes for 1997-98 cohort for each assessment.

For this cohort the message is rather mixed. It is clear that there is no significant difference between the two groups for the pre-test. However, the effect size for the Semester A assessments is low, although indicating some effect in favour of the non-ITS group. This is especially true of the differences between the multiple choice and descriptive components of the final semester exam, where the ITS group scored relatively higher, with a negative ES for the multiple choice part, indicating a higher average mark for the control group. Again, there is no significant difference between the two groups in Semester B.

3.5.3 1998-99

pretest	mid a	fin a mark	fin a exam	fin a mc	fin a desc	mid b	fin b mark	fin b exam
0.16	0.70	0.88	0.89	0.16	0.94	-0.10	0.48	0.57

Table 3.3 Standardised Effect Sizes for 1998-89 cohort for each assessment.

The result for this cohort is very similar to that for that for 1996-97.

3.5.4 1999-2000

pretest	fin a mark	fin a exam
0.04	-0.61	-0.39

Table 3.4 Standardised Effect Sizes for 1999-2000 cohort for each assessment.

This cohort was an anomaly, but, as has been explained above, is kept for comparison. As can be seen from Table 3.4, there was no significant difference between the two groups in the pre-test, but in the final assessment in Semester A, the course which has been previously classified as non-ITS, i.e. the BEME students, scored higher marks than the other, BEMTE, group. As they were both taught using traditional methods in the same class, it is a noted that the consolidated effect size for the semester was - 0.5, indicating a significant difference in favour of what would otherwise have been the control group.

3.5.5 2000-01

pretest	mid a	fin a mark	fin a exam	fin a mc	fin a desc	mid b	fin b mark	fin b exam
0.08	-0.10	0.24	0.37	0.29	0.35	0.29	0.32	0.17

Table 3.5 Standardised Effect Sizes for 2000-01 cohort for each assessment.

Again, it can be seen that there was no significant difference between the two groups for the pretest, but in this cohort, the non-ITS group scored higher for the Semester A mid-test. This was probably an anomaly, as the expected pattern for the final Semester A assessment asserted itself, although with a smaller effect size. The difference between the two groups reduced in Semester B.

3.5.6 2001-02

pretest	mid a	fin a mark	fin a exam	fin a mc	fin a desc	mid b	fin b mark	fin b exam
0.50	0.63	0.48	0.40	0.45	0.35	0.18	0.21	0.01

Table 3.6 Standardised Effect Sizes for 2001-02 cohort for each assessment.

This cohort was slightly different to the others, in that there was a significant difference between the two groups for the pre-test. It is interesting to note this result - as the data provided in Appendix 2 indicates that there should be homogeneity between the groups except in language abilities, where the non-ITS were significantly better than the ITS group! The ITS group continued to score higher than the non-ITS group for all assessments, but again, the narrowing of the difference in Semester B is noticeable.

3.5.7 Overall assessment effect size

If all the cohorts except 1999-2000 are considered together, it is possible, using meta-analysis, to determine what the overall effect size for each assessment will be. This is shown in Table 3.7.

pretest	mid a	fina mark	fin a exam	fina m/c	fin a desc	mid b	fin b mark	fin b exam
0.16	0.49	0.61	0.58	0.25	0.58	0.12	0.19	0.23

1999-2000 cohort results not included

Table 3.7 Standardised Effect Size for each assessment for all cohorts except 1999-2000

Again, it can be seen quite clearly that there was no significant difference between the two groups as far as the pre-test was concerned. However, for Semester A, there was a significant difference between the two groups for all but the multiple choice element of the final examination. The implications will be discussed later.

Similarly, although there was a difference between the two groups in Semester B, it was not so pronounced.

3.5.8 Semester effect size

An alternative way of looking at the results is to perform a meta-analysis on all the assessments, except the pre-test, for each semester for each cohort, except for 1999-2000, which was for the final examination in Semester A only. This is shown in Table 3.8.

	1996-7	1997-8	1998-99	1999-2000	2000-01	2001-02
sem a	0.91	0.24	0.72	-0.50	0.23	0.46
sem b	0.12	0.05	0.32	-	0.26	0.13

Pre-test results not included

Table 3.8 Standardised Effect Sizes for each cohort for each semester

These results are from a meta-analysis of all the results for all the assessments in each semester for each cohort. The overall Effect Size for all the assessments in each semester is shown in Table 3.9

all sem a	0.51
all sem b	0.19

1999-2000 cohort results not included
Pre-test results not included

Table 3.9 Overall effect size for all cohorts for each semester

3.5.9 Overall effect size

Finally, a meta-analysis is performed on all the assessments over all cohorts, excluding the pre-test. This is shown in Table 3.10. Little can be read into this, as the metric for Semester B was significantly different to that for Semester A, as mentioned above. However, notwithstanding this, there is some noticeable effect present.

	1996-7	1997-8	1998-99	1999-2000	2000-01	2001-02	All
all	0.68	0.19	0.57		0.24	0.34	0.40

1999-2000 cohort results not included
Pre-test results not included

Table 3.10 Overall Standardised Effect Size for both semesters for each cohort, and for the whole period of the study.

3.6 Discussion

It was shown in Chapter 2 that there was close equivalence in the entrance qualifications of both groups - the control non-ITS group, and the experimental ITS group - see Appendix 2 for the statistical analysis. The data for the pre-test has also been analysed and the average effect size across all cohorts was found to be 0.156, indicating approximately a 10% non-overlap (See Introduction to Appendix 6). Alternatively, a t-test analysis of the pre-test data shows a *p* of 0.88, see Table 3.11 below. All this strongly suggests that the two groups are equivalent on entry.

Source of Variation	*F*	*P-value*	*F crit*
Between Groups	0.022161	0.88462	4.964603

Table 3.11 t-test analysis for pre-test marks, all cohorts

It is therefore instructive to consider the effect of the difference in teaching methodologies on the final assessments of the two groups. From Table 3.9 it can be seen that there is a significant effect in Semester A, with a lower effect in Semester B. Table 3.10 shows that the effect size for both semesters is 0.40, indicating a lower effect overall, but still important. This effect size means that the mean of the experimental group is at the 66th percentile of the control group. There is little doubt that the teaching methodology had an effect on the assessment results.

Further analysis of the data shown above does raise some interesting questions, however. Why does the effect seem to 'wear off' in the second semester? There also seems to be no simple relationship between the difference in pre-test results and the final result. For example, in 2001-02, the ITS group had significantly higher marks for the pre-test, and the assessment at the end of Semester A showed a similar effect, but at the end of Semester B the difference between the two groups was small! These questions will be addressed in the conclusions.

It is interesting to note the similarity between the results given here and those reported in two meta-analyses carried out on small group collaborative learning. The first, by Johnson, Johnson and Stanne (2000) considered 164 studies investigating eight cooperative learning methods. This covered schools as well as colleges. Consequently, we will not look at this study in detail, other than to comment that the authors state that the consistency of the results and the diversity of the cooperative learning methods provide strong validation for its effectiveness.

The other meta-analysis study was in 1998 by Springer, Stanne and Donovan (1998). This analysed 383 reports in literature related to small group learning in post-secondary science, maths, engineering and technology (SMET) courses from 1980 or later, 39 which met the inclusion criteria for the meta-analysis. These were, first that the undergraduates were on science, mathematics, engineering, or technology courses or degree programmes at accredited post-secondary institutions in North America. Secondly, studies must have incorporated small-group work inside or outside of the classroom. Thirdly, the study was conducted in an actual classroom or programmatic setting rather than under more controlled laboratory conditions. Fourthly, the research was published or reported in 1980 or later on the grounds that recent studies may be more relevant to the current global context in which students learn, and fifthly, the research reports enough statistical information to estimate effect sizes.

Of the 39 studies analysed, 37 (94.9%) presented data on achievement, 9 (23.1%) on persistence or retention, and 11 (28.2%) on attitudes. Most of the reports retrieved did not qualify for inclusion because they were not based on research.

According to Springer et al (ibid):

> "The main effect of small-group learning on achievement, persistence, and attitudes among undergraduates in SMET was significant and positive. Based on 49 independent samples, from 37 studies encompassing 116 separate findings, students who learned in small groups demonstrated greater achievement (*Effect Size, d* = 0.51) than students who were exposed to instruction without cooperative or collaborative grouping. Similarly, based on 12 independent samples, from 11 studies encompassing 40 findings, students in small groups expressed more favourable attitudes (*d* = 0.55) than their counterparts in other courses or programmes. Finally, based on 10 independent samples and findings from 9 studies, students who worked in small groups persisted through SMET courses or programmes to a greater extent (*d* = 0.46) than students who did not work cooperatively or collaboratively".

Ignoring the effects for gender, race and group size, which were not included in the study in this thesis, two significant effects were reported.

First, for the procedures used in small-group learning, Springer et al state:

> There was a higher average weighted effect for supplemental instruction (*d* = 0.65) - typically study sessions outside of class - than for in-class instruction (*d* = 0.44). The pattern of differences was reversed for attitudinal outcomes: more favourable effects on attitudes were evident for in-class instruction (*d* =0.59) than for supplemental instruction (*d* = 0.24). The data suggested that greater time spent working in groups had significantly more favourable effects on students' attitudes, with effect sizes of 0.77 for high group time, 0.26 for medium, and 0.37 for low. No significant association between time spent in groups and achievement was evident".

The other effect was that on the outcome measure:

> "The effects of small-group learning on achievement were significantly greater when measured with exams or grades (*d* = 0.59) than with the standardised instruments (*d* = 0.33). Although small-group work among students had significant and positive effects on students' attitudes toward learning the material (*d* = 0.56) and their self-esteem (*d* = 0.61), the effect on their motivation to achieve (*d* = 0.18) was one of only two nonsignificant results of small-group work that we report in this study".

It would be instructive to complete another meta-analysis of the published literature in this field with data published since 2000. Many papers have recently been published on studio teaching - see next chapter - and further data is now available that was not included in these two analyses. However, the data presented above does seem to support the findings of the study in this thesis which gave the overall effect size of 0.4 for both semesters or 0.51 for the first semester. These two effect sizes are within the range of those reported by Springer et al above. They found an effect size of 0.51 for greater achievement for those learning in small groups in a collaborative or cooperative manner. They also found that those who worked in small groups persisted in their courses or programmes to a greater extent - an effect size of 0.46, as reported above. This would seem to be confirmed from attendance data taken during the duration of the study reported in this thesis. For the non-ITS group attendance at lectures and tutorials started off high (around 95%) at the beginning of the semester, but dropped to around 50% by the end. For the ITS group attendance has remained at around 95% throughout the course, a figure which continues to this day.

Also, from the data presented in this study in Section 3.5 above, it is clear that the highest consistent effect sizes were for the final examination in Semester A. This would seem to corroborate the findings of Springer et al reported above, that The effects of small-group learning on achievement were significantly greater when measured with exams or grades (*d* = 0.59). This compares with an effect size of 0.58 in the current study.

Having established that, at least in the first semester, there is a difference in educational performance between the two groups based on their assessment results, it is now useful to consider the students' reactions when presented with the studio environment. As the two groups had no choice in this matter, as the groups were delineated by the degree programmes the students enrolled in, the reactions of the studio-based group is important for our understanding of the methodology.

At the same time, students at Rensselaer Polytechnic Institute chose to take the studio course, and also had previous experience of studio based teaching. It is instructive to compare their reactions to those from CityU.