

**THE MEASUREMENT OF THE IMPACT OF LEARNING
INTERVENTIONS: THREE METRICS**

ANDREW WICKS

A dissertation submitted in partial fulfilment of the
requirements of the University of Greenwich for the Degree
of Doctor of Philosophy

June 2020

I certify that this work has not been accepted in substance for any degree and is not concurrently being submitted for any degree other than that of Doctor of Philosophy being studied at the University of Greenwich. I also declare that this work is the result of my own investigations except where otherwise identified by references and that I have not plagiarised the work of others.

Andrew Wicks

Prof. Chris Walshaw

Prof. Constantinos Ierotheou

Prof. Liz Bacon

ACKNOWLEDGEMENTS

Prof. Lachlan MacKinnon was one of the supervisors until he moved on from the University of Greenwich, but has been as influential in the direction and scope of the work presented here as anyone. I feel very lucky to have had his support and encouragement.

All my colleagues at the University of Greenwich have been particularly supportive, especially Prof. Cornelia Boldyreff, Dr. Yasmine Arafa and Dr. Ralph Barthel.

I am lucky that my family (especially my wife, Jan) were willing to put up with such a lot so that I could complete this work. But my daughters, Amanda and Julie, their husbands and all my grandchildren all played their part too.

Geraldine Bradley created the logos used on the experimental website. Tony Weight was my sounding board and contributed to the ideas presented here. I could not have completed this without their support.

However, my supervisors, Prof. Chris Walshaw, Prof. Constantinos Ierotheou and Prof. Liz Bacon, were the rocks on which this work was built. They encouraged me every step of the way, put up with, “Oh, I’ve just had an idea ...” and then brought me back to reality. They were patient, knowledgeable and worked incredibly hard to get me to this point. These are the real A-Team.

ABSTRACT

This study aims to answer the question, "Can the differences between modules and their year groups be used to assess whether a change in pedagogy affected learning outcomes?" It is often assumed that the Student t-test and the Fisher F-test can be used to examine differences between successive year groups. However, this study suggests that any effects of a new technique in the classroom may be significantly influenced by a cohort effect (differences in the way successive year groups react).

The grades for 48 modules were analyzed over an 8-year period, from 2011/12 to 2018-19, to assess how well the Student t-test, the Fisher F-test, the Mann-Whitney U-test, the Levene test, and the Brown-Forsythe test performed at the 5% level. One would expect that approximately 5% of comparisons (approximately 16 cases) would be returned for each test. However, the actual numbers were 92, 51, 111, 75, and 62 respectively. This suggests the need for an alternative statistical technique.

This new technique ranks modules within a year by examining the results only for those students who had taken both modules. It then assesses the difference in ranking, indicating how much a module had moved relative to others, to identify where movements had occurred. Three measures were developed: the Learning Intervention Location Metric (LILM), the Learning Intervention Spread Metric using the mean ($LISM_{Mean}$), and the Learning Intervention Spread Metric using the median ($LISM_{Median}$). Their confidence intervals were estimated using bootstrapping with $n = 10,000$, returning 16, 15, and 15 results as significant at the 5% level respectively.

This led to a classification system in which changes in the LILM/ $LISM_{Mean}$ and LILM/ $LISM_{Median}$ are sorted into 9 categories, representing approximately the high, middle, and low one-thirds for each statistic, creating a 3 x 3 grid. This system can be used to classify year-on-year changes in student results.

This new system was then tested twice. The first test involved an experimental website that utilized accepted good teaching practices to design the student experience. The second test involved splitting a large module of 30 credits into two parts to assess whether that would improve results.

Unfortunately, neither experiment seemed to yield improvement. The experimental website remained unchanged, and splitting the large module appeared to have a negative effect. The reasons for these outcomes are discussed.

Therefore, this study aims to provide a new method for assessing the effectiveness of changes in the learning environment and classifying those changes. Additionally, it offers a prototype method for designing web-based learning environments based on perceived good practices.

Table of Contents

List of Figures.....	xiv
List of Tables	xvii
List of Equations.....	xix
Abbreviations.....	xx
Key Terms.....	xxi
Chapter 1 Introduction	1
1.1 Introduction.....	1
1.2 Motivation and Background.....	1
1.3 The Problem Definition.....	2
1.4 Aims and Objectives.....	3
1.4.1 Aims.....	3
1.4.2 Objectives	4
1.5 Definitions and Key Terms.....	4
1.6 Overview of the Chapters.....	5
1.6.1 Chapter 2 – The Literature Review.....	5
1.6.2 Chapter 3 – Methodology	6

1.6.3	Chapter 4 – The Effectiveness of Standard Statistical Tests	6
1.6.4	Chapter 5 – The TEL Experiment.....	6
1.6.5	Chapter 6 – The Learning Intervention Impact Metrics (LIIMs).....	7
1.6.6	Chapter 7 – LIIM Results and Discussions.....	7
1.6.7	Chapter 8 – Conclusions and Areas for Future Research	7
1.6.8	References, Appendices and Index	8
Chapter 2	The Literature Review and Analysis	9
2.1	Introduction.....	9
2.2	Theories of Learning.....	12
2.2.1	Introduction.....	12
2.2.2	The scope of the review.....	14
2.2.3	The stages of learning.....	14
2.2.4	Metacognition.....	15
2.2.5	Intelligence.....	16
2.2.6	Learning styles	17
2.2.7	Motivation and cognitive dissonance	19
2.2.8	Deep/Surface learning.....	21

2.2.9	Alternative education	21
2.3	Theories of Teaching.....	23
2.3.1	The zone of proximal development (ZPD)	23
2.3.2	Threshold concepts	24
2.3.3	Experiential learning.....	25
2.3.4	Connectivism and the Medici Effect.....	26
2.3.5	Self-paced learning	27
2.3.6	Role play.....	28
2.3.7	The flipped classroom	28
2.3.8	Computer-based immersive education	29
2.3.9	Adaptive learning.....	30
2.4	Statistical Controversy	32
2.4.1	Statistical tools	34
2.4.2	Two examples of controversies	36
2.5	A Discussion of Appropriate Statistical Techniques.....	38
2.5.1	Group comparisons.....	38
2.5.2	Parametric tests.....	39

2.5.3	Non-parametric tests.....	40
2.5.4	Analysing sets of modules	41
2.5.5	Using graphical methods	42
2.6	Conclusions.....	43
2.6.1	Learning and teaching	43
2.6.2	Analysis.....	43
2.6.3	Revised research question.....	45
Chapter 3	Methodology.....	46
3.1	Introduction.....	46
3.2	An Overview and Rationale for the Research Methodology.....	47
3.3	Obtaining the Data.....	48
3.3.1	Grades data.....	48
3.3.2	Experimental website data	49
3.4	Conducting the Experiments.....	49
3.4.1	Exploring five standard learning intervention measures.....	49
3.4.2	The development of the experimental website	50
3.5	Analysis of the Results.....	53

3.5.1	The analysis of student usage of the experimental site.....	53
3.6	Summary	54
Chapter 4	The Effectiveness of Standard Statistical Tests	55
4.1	Introduction.....	55
4.1.1	Samples for the descriptive statistics tables	55
4.1.2	Samples for the normality, kurtosis and skewness tables	56
4.1.3	Samples for the year-on-year comparison tables	56
4.2	Cleansing the Grades Data.....	57
4.3	Evaluating the Standard Statistical Techniques	60
4.3.1	Introduction.....	60
4.3.2	The 5% test.....	60
4.3.3	Student t-test (Parametric measure of location).....	61
4.3.4	Fisher F-test (Parametric measure of spread).....	62
4.3.5	Mann-Whitney U-test (Non-parametric measure of location).....	63
4.3.6	Levene test (Non-parametric measure of spread).....	63
4.3.7	Brown-Forsythe test (Non-parametric measure of spread).....	64
4.4	Conclusions.....	64

Chapter 5	The TEL Experiment.....	66
5.1	Introduction.....	66
5.2	The Design of the TEL Site.....	67
5.2.1	The educational practice – Theory.....	67
5.2.2	Title.....	68
5.2.3	Summary *.....	68
5.2.4	Useful sites.....	69
5.2.5	Contents.....	69
5.2.6	Alternative contents *.....	69
5.2.7	Associated materials.....	70
5.2.8	Pre-requisite knowledge *.....	71
5.2.9	External links *.....	71
5.2.10	Quiz and quiz results.....	72
5.2.11	Personal notes *	73
5.2.12	Approval rating and feedback	73
5.2.13	Social media links *	74
5.2.14	Coursework requirements and additional coursework materials.....	74

5.2.15	Usage statistics.....	75
5.2.16	A sample task	75
5.2.17	Summary of theories	77
5.3	The website – Practice.....	78
5.3.1	Site philosophy	78
5.4	Selecting the Experimental Group.....	79
5.5	The Running of the Experimental Site.....	81
5.5.1	The study system.....	81
5.6	The Tutorial System.....	83
5.6.1	Introduction.....	83
5.6.2	How to use the experimental website – A flowchart.....	85
5.7	Conclusions.....	86
Chapter 6	The Learning Intervention Impact Metrics (LIIMs)	87
6.1	Introduction and basic premise.....	87
6.2	What might be expected?	89
6.2.1	Initial thoughts.....	90
6.3	The Principle of “Like-with-Like”.....	90

6.4	The Positional Scores.....	93
6.5	The Learning Intervention Location Metric (LILM).....	95
6.6	Immediate potential problems with the LILM.....	98
6.7	Summary for the LILM	99
6.8	The Learning Intervention Spread Metrics.....	99
6.8.1	The measurement of spread	100
6.8.2	The Learning Intervention Spread Metric – Mean ($LISM_{Mean}$)	101
6.8.3	The Learning Intervention Spread Metric – Median ($LISM_{Median}$).....	103
6.9	Using Both the LILM and $LISM_{Mean}$ to Assess Impact	106
6.10	Combining the LILM and $LISM_{Mean}$.....	106
6.10.1	Positive LILM and Positive $LISM_{Mean}$ (Case 1).....	107
6.10.2	Stable LILM and Positive $LISM_{Mean}$ (Case 2).....	107
6.10.3	Negative LILM and Positive $LISM_{Mean}$ (Case 3).....	107
6.10.4	Positive LILM and Stable $LISM_{Mean}$ (Case 4).....	108
6.10.5	Stable LILM and Stable $LISM_{Mean}$ (Case 5).....	108
6.10.6	Negative LILM and Stable $LISM_{Mean}$ (Case 6).....	108
6.10.7	Positive LILM and Negative $LISM_{Mean}$ (Case 7).....	109

6.10.8	Stable LILM and Negative $LISM_{Mean}$ (Case 8)	109
6.10.9	Negative LILM and Negative $LISM_{Mean}$ (Case 9).....	109
6.11	Conclusion	110
Chapter 7	LIIM Results and Discussion.....	111
7.1	Introduction.....	111
7.2	Results from Historical Data.....	111
7.2.1	Introduction.....	111
7.2.2	The Learning Intervention Location Metric (LILM).....	112
7.2.3	Bootstrapping – Confidence Intervals for the LIIMs.....	113
7.2.4	Confidence intervals for the LILM.....	114
7.2.5	The Learning Intervention Spread Metric – Mean ($LISM_{Mean}$)	116
7.2.6	The Learning Intervention Spread Metric – Median ($LISM_{Median}$).....	119
7.2.7	Overview of the three tests.....	121
7.2.8	The case analyses.....	122
7.2.9	Summary.....	123
7.3	Alternative uses.....	124
7.4	Applying the Results	124

7.4.1	Examining the effectiveness of the experimental site.....	125
7.4.2	Examining other anomalies discovered by the LIIMs.....	128
7.4.3	The 16 other significant results	129
7.4.4	The reasons behind the LIIMs significant at the 5% level.....	131
7.4.5	Conclusions.....	132
7.4.6	Summary.....	133
7.5	Comparison to the Five Standard Tests.....	133
7.6	Conclusions.....	134
Chapter 8	Conclusions and Areas for Future Research	135
8.1	Introduction.....	135
8.2	Summary of the Study.....	135
8.3	How the Research Aims and Objectives were Met.....	139
8.4	Main Contributions of the Research.....	141
8.5	Limitations of the Study	142
8.6	Future Research.....	143
8.7	Final Reflections.....	143
References	145	

The raw and clean grade data sets overviews.....	152
The LIIM statistical overviews.....	154
The number of students.....	155
The mean score in each run.....	156
The median score for each run.....	157
The standard deviation of the scores for each run.....	158
The coefficient of skewness of the scores for each run.....	159
The kurtosis of the scores for each run	160
p of normality (D'Agostino-Pearson) for each run	161
Student t-test values by run.....	162
Student t-test probabilities by run.....	163
Fisher F-test values by run	165
Fisher F-test probabilities by run	166
Mann-Whitney U-test values by run	168
Mann-Whitney U-test probabilities by run.....	169
Levene test values by run.....	171

Levene test probabilities by run	172
Brown-Forsythe test values by run.....	174
Brown-Forsythe test probabilities by run.....	175
Positional scores for the LILMs for each run.....	177
Learning intervention location metrics (LILM) for each run.....	178
Positional scores for the LISM_{Mean} values for each run.....	180
Learning intervention spread metrics – mean for each run.....	181
Positional scores for the LISM_{Median} values for each run.....	183
Learning intervention spread metrics – median for each run	184
Screen shots from the LYW4.life website	186
Appendix 26.1 - Login/Registration.....	186
Appendix 26.2 - The Home Page.....	188
Appendix 26.3 - Enrolment	189
Appendix 26.4 - A Task.....	190
Appendix 26.5 - A Quiz	192
Appendix 26.6 - Run Administration.....	193
Appendix 26.7 - Creating and Editing Tasks.....	194

Appendix 26.8 - Checking Student Progress.....	199
Appendix 26.9 - View Progress.....	201
Appendix 26.10 - Personalisation.....	202
Appendix 26.11 - Usage Statistics.....	204
Ethical approval confirmation letter	205
Index.....	206

List of Figures

Figure 1 - Teaching and Learning Theories	13
Figure 2 - Kolb's Experiential Learning Loop (Kolb 1984).....	15
Figure 3 - Zone of Proximal Development (ZPD) (Vygotsky 1978).....	25
Figure 4 - Alternative view of the ZPD.....	25
Figure 5 - Histogram of the raw data set	58
Figure 6 - Histogram of the cleaned data set.....	59
Figure 7 - Examples of items within a task.....	76
Figure 8 - The value of t needed at the 5% significance level by degrees of freedom.....	80
Figure 9 - Experimental Site Flowchart of Usage	85
Figure 10 - Grade comparison in an ideal world.	87
Figure 11 - Histogram of the Learning Intervention Location Metrics (LILM).....	113
Figure 12 - The distribution of the LILM means.....	115
Figure 13 - The distribution of the LILM standard deviations	115
Figure 12 - The distribution of the $LISM_{MeanS}$	117
Figure 13 - The distribution of the $LISM_{MeanS}$ standard deviations.....	117
Figure 12 - The distribution of the $LISM_{MedianS}$	120

Figure 13 - The distribution of the LISM _{MedianS} standard deviations	120
Figure 16 - Opening screen	186
Figure 17 - Registering with LYW4.....	186
Figure 18 - Logging in to LYW4	187
Figure 19 - The home page once logged in.....	188
Figure 20 - Entering the code for the run.....	189
Figure 21 - Applying for a run without enrolling.....	189
Figure 22 - Opening a module run	190
Figure 23 - Selecting a task	190
Figure 24 - A task.....	191
Figure 25 - The quiz for a task	192
Figure 26 - Creating a run for a module.....	193
Figure 27 - Accepting a student onto a run.....	193
Figure 28 - Adding a tutor to a run.....	193
Figure 29 - Creating a task outline	194
Figure 30 - Editing the task.....	194
Figure 31 - Adding text and graphics.....	195
Figure 32 - Editing the task summary	196

Figure 33 - Adding video, sound and PDF.....	196
Figure 34 - Adding links to other sites.....	196
Figure 35 - Adding extra materials to a task	197
Figure 36 - Adding links to prerequisite tasks	197
Figure 37 - Creating quiz questions	197
Figure 38 - Seeing an overview of tasks created.....	198
Figure 39 - Monitoring the progress of a student.....	199
Figure 40 - The student report for a run.....	199
Figure 41 - An overview of all students on a run.....	200
Figure 42 - Progress report - The student view.....	201
Figure 43 - Progress so far - The student view.....	201
Figure 44 - Consent for research data	202
Figure 45 - Altering the email address.....	202
Figure 46 - Changing the background colours	202
Figure 47 - Changing the font size.....	203
Figure 48 - Feature request form.....	203
Figure 49 - Usage statistics.....	204

List of Tables

Table 1 - List of Abbreviations.....	xx
Table 2 - Statistical tests considered.....	50
Table 3 - Task items by inclusion.....	52
Table 4 - The number of instances where $p < 5\%$	64
Table 5 - Fictional initial year results.....	91
Table 6 - Fictional results analysis	92
Table 7 - The results for the year with a learning intervention.....	96
Table 8 - Paired scores in a year where student only take one module	97
Table 9 - Calculating the LILM.....	98
Table 10 - The squared deviations from the mean for year 0.	101
Table 11 - The squared deviations from the mean for year 1.	102
Table 12 – LISM _{Mean} results	103
Table 13 - The squared deviations from the median for year 0.	104
Table 14 - The squared deviations from the median for year 1.	105
Table 15 - LISM _{Median} results.....	105
Table 16 - LILM vs LISM _{Mean} Case Table	106
Table 17 - Bootstrap statistics for the LILMs	114

Table 17 - Bootstrap statistics for the $LISM_{MeanS}$	116
Table 17 - Bootstrap statistics for the $LISM_{MedianS}$	119
Table 18 - Summary of the LIIM Results	121
Table 19 - Case Matrix from Chapter 6	122
Table 20 - Case Matrix from using LILM (+/- 0.04) with $LISM_{Mean}$ (+/- 0.14).....	122
Table 21 - Case Matrix from using LILM (+/- 0.04) with $LISM_{Median}$ (+/- 0.14).....	123
Table 22 - Experiment (Module 5~04, 2018/19) – Descriptive Statistics.....	125
Table 23 - Module 4~01 (2017/18) – Descriptive Statistics	129
Table 24 - Runs whose LILM was significant at the 5% level.	131
Table 25 - A comparison of statistical tests at the 5% level.....	134

List of Equations

Equation 1 - Positional score calculation for module A	93
Equation 2 - Simplified equation for positional score.....	94
Equation 3 - The positional score for module B	94
Equation 4 - The positional score for module C	95
Equation 5 - Variance.....	100
Equation 6 - The formula for the positional scores for the $LISM_{Mean}$	102

Abbreviations

ANOVA	ANalysis Of VAriance
LILM	Learning intervention location metric
LISM _{Mean}	Learning intervention spread metric – mean
LISM _{Median}	Learning intervention spread metric – median
LIIM	Learning intervention impact metrics (LILM, LISM _{Mean} and LISM _{Median})
TEL	Technology enhanced learning

Table 1 - List of Abbreviations

Key Terms

The following definitions of key terms apply solely to this research study and they could be defined differently in other research studies.

Learning intervention: Any changes made to a teaching environment (academic or physical) which may or may not have influenced student achievement.

Learning intervention impact metrics (LIIMs): A group of three statistical techniques used to assess whether the changes made to a teaching environment (academic or physical) have influenced student achievement.

Learning intervention location metric (LILM): A statistical technique which assesses the probability of whether the changes made to a teaching environment (academic or physical) have influenced the “central point” of the student grades.

Learning intervention spread metrics using the mean (LISM_{Mean}): A statistical technique based on the mean grade which assesses the probability of whether the changes made to a teaching environment (academic or physical) have influenced the spread of the student grades.

Learning intervention spread metrics using the median (LISM_{Median}): A statistical technique based on the median grade which assesses the probability of whether the changes made to a teaching environment (academic or physical) have influenced the spread of the student grades.

Metacognition: The ability to understand how to optimise learning opportunities. Someone who is metacognitive would be a more self-sufficient learner and therefore more independent.

Null hypothesis significance testing (NHST): This is a statistical technique where two samples are compared to see if they come from the same population. It is usual to assume that they are from the same population (the null hypothesis) until the statistics prove otherwise. Rather than

give a clear cut “Yes” or “No”, the technique returns the probability that they are from the same population. It is then up to the researcher to decide whether this probability is significant.

Positional score ($PS_{Location}$, PS_{Mean} and PS_{Median}): A value awarded to a module which represents its position within a group of modules that have students in common. This merely looks at scores and does not take variations on marking, quality of teaching or environmental factors into account. There are three versions. The first tests the position of the location of the grades (the $PS_{Location}$) and another two for the position of the spread of grades (the PS_{Mean} and PS_{Median}).

Set: A group of modules which have students in common in an academic year. For example, it may be those modules taught to first year undergraduate on a programme.

Task: An atom of learning. It is a small piece of the syllabus that can be taught independently of the others. An example might be adding single digit numbers whose result is under 10, rather than addition. Addition requires several tasks which build on each other.

Technology enhanced learning (TEL): The use of ICT and Internet technologies to provide and/or enhance the learning of students. The creation of a new version of one of these formed part of the assessment of the effectiveness of the LILM, $LISM_{Mean}$ and the $LISM_{Median}$.

Chapter 1 Introduction

1.1 Introduction

This project started as an attempt to define a technique which would allow a TEL to be created, one based on a survey of perceived educational good practice. However, it was quickly realised that a statistical technique was needed which would allow successive years of a module to be compared because standard tests, such as the Student t-test, might not be appropriate.

1.2 Motivation and Background

There are many systems and websites that offer Technology Enhanced Learning (TEL). Yet there were no papers found which showed that a TEL site was based on a systematic review of the educational academic literature.

John Hattie (Hattie 2009), in his statistical survey of educational practices, found that TEL had a marginal impact on learning. However, most materials uploaded to these sites are “shovelware” (Chou & Liu 2005), i.e. just pre-existing materials which have been made available online. It is therefore not surprising if the TEL results are then similar to pre-TEL results since the materials being presented are the same as before, just on a new medium (computer versus paper).

TEL offers a greater opportunity for learning because the learner could have greater control over the mode, time, place and speed of study. Most educationalists would see the development by the student of strategies that work for them as positive because they increase independence

and therefore lead to greater student metacognition (Festinger 1957). If the learner only sees TEL as an alternative to paper then the status quo would be maintained.

One would imagine that the use of standard statistical tools to compare year-on-years differences that have been used by the academic community for a long time, would not be controversial. Conventionally, tests such as the Student t-test would be used to compare the experimental year with its predecessor. But that assumes that the students come from similar populations. That would be unlikely in this case because year group profiles differ year-on-year. Standard techniques such as the Student t-Test (Student 1908) and the Fisher F-Test (Snedecor 1934) are standard techniques for assessing differences between groups. These parametric tests were augmented by non-parametric tests such as the Mann-Whitney U-Test (Mann & Whitney 1947), the Levene Test (Levene 1960) and the Brown-Forsythe Test (Brown & Forsythe 1974).

However, there has been controversy about their use for some time as Raymond Nickerson (Nickerson 2000) pointed out. The criticism does not centre around whether these techniques are mathematically robust – they are – but rather, the circumstances under which they are used. If these tests are used under circumstances for which they were not designed then their validity is, at best, questionable.

1.3 The Problem Definition

The problem is how a change in the learning environment, e.g. different teaching styles, alternative materials or changing the rooms/times of classes, can be assessed. Do the standard statistical tests work in this situation, and if not, could new tests be devised which performed

better? The development of the TEL site would then provide an opportunity to test both the usefulness of the TEL and whichever statistical tests proved most appropriate.

1.4 Aims and Objectives

Initially, the research question is, *“Can a set of learning elements be synthesised from existing educational research and experimentally analysed for effectiveness, to form the basis of an online pedagogy?”*

This research question will be revised after the literature review and can be found in section 2.6.3.

1.4.1 Aims

- 1) To develop a TEL experience which uses educational theory in its design as an example of a learning intervention.
- 2) To assess the appropriateness of standard statistical techniques, such as the t- and F-test, U-test, Levene test and Brown-Forsythe test because these are taught and used as the standard parametric and non-parametric statistical tests.
- 3) To offer and test an alternative statistical technique if the standard measures were shown to be inappropriate.
- 4) To use each of the tests, both old and new, to determine whether the new TEL aided student achievement.

1.4.2 Objectives

- 1) To create and use an on-line learning environment based on academic literature and assess its impact on student achievement.
- 2) To test whether standard statistical tools of the Student t-test, the Fisher F-test, the Mann-Whitney U-test, the Levene test and the Brown-Forsythe test should be used when assessing the effect of a learning intervention.
- 3) To provide alternative, non-parametric tests for changes in location and spread if the five tools mentioned above are inappropriate.

1.5 Definitions and Key Terms

This study uses the term learning intervention to mean any activity which might impact on the ability of students to achieve grades. It uses grades as a proxy variable for the quantity of learning and understanding that has taken place. Poorly designed assessment strategies might affect the outcomes, but they would then become another learning intervention.

Another example of a learning intervention would be a TEL learning environment which is a website used on a module to aid the student. This might be software such as Moodle or Blackboard, but also includes sites used for reference (StackOverflow, Wikipedia, Google Earth, etc.) as well as those aimed at teaching such as the Khan Academy (<https://www.khanacademy.org>) or Udacity (<https://www.udacity.com>).

The pages for these sites are broken down in this study into items. An item is an element such as the title, a summary, the content being delivered or a quiz. This study started as an attempt

to provide an initial list of potential items based on the academic literature and then create a method to test which items helped students the most.

1.6 Overview of the Chapters

1.6.1 Chapter 2 – The Literature Review

This provides the context for this study. It started as an overview of the areas of education which may have an impact on the design of educational websites. The intention was to create a list of features which go to make “good education” as well as a general definition of “good education”. This list was not intended to be exhaustive. Others may consider that factors they feel are important have been omitted. The methodology used would allow the new factors to be added and, if appropriate and then the design of the TEL site to be amended accordingly. This area has now been scaled back in the literature review.

However, the statistical side of the project became ever more prominent as it continued. The controversy over the use of null hypothesis significance testing (NHST) kept cropping up in the literature (Herrera-Bennett et al. 2020)(Nickerson 2000). Even though this section of the literature review is smaller in terms of the number of words, its effect on the setting of the context for the work is greater. Therefore, the research question was adapted for the new situation.

1.6.2 Chapter 3 – Methodology

The research methodology had to be revisited in the light of the changed research question. A list is a linear object and the list for the project plan in the methodology implied a fixed order for completion. That was not the case here (and probably rarely is). The chapters 3 to 7 were part of a development cycle which evolved as new opportunities for insights became available.

1.6.3 Chapter 4 – The Effectiveness of Standard Statistical Tests

After gaining ethical approval, the grades have been taken from the university system and organised into modules and years.

The means and variances of the grades in modules in successive years were compared using the t-test, F-test, Mann-Whitney U-test, Levene test and the Brown-Forsythe test to assess whether these tests are appropriate in this situation. These tests are used because they are, in the main, taught in statistics classes for those who intend to use statistics rather than those who will be statisticians, as the way to analyse differences between groups.

1.6.4 Chapter 5 – The TEL Experiment

The design of the site (lyw4.life) is based on the good practice identified during the literature review. However, it is recognised that this design could not be fixed since ideas on good practice and other good practices are being discovered. The design of the TEL site is therefore modular so that the elements described in this chapter can be added, amended or deleted as changes become necessary.

1.6.5 Chapter 6 – The Learning Intervention Impact Metrics (LIIMs)

Having created a TEL design, it would then need to be tested. The students who used this system were on a second-year module covering the design and development of relational databases.

It is necessary to assess whether the TEL has an impact on the outcomes for this module. This chapter examines the design of alternative statistical tests which could be used to assess the impact of a learning intervention.

1.6.6 Chapter 7 – LIIM Results and Discussions

The practical implementation of the new statistical tests are examined in this chapter. The new tests are compared to the five standard statistical tests for their applicability when analysing the effect of a learning intervention. The new techniques are then used to also find areas of education which caused anomalies and to classify outcomes from learning interventions.

1.6.7 Chapter 8 – Conclusions and Areas for Future Research

This chapter establishes which contributions to research have been made. It does this by evaluating the outcomes from the research questions and assessing the implications of these findings. It is recognised that the statistical analyses presented only represent the outcomes for the programmes and year groups on which they are based, therefore caution and a wider range of programmes and institutions is urged before firm conclusions are drawn.

This discussion of the results from the previous chapters allowed the research questions, aims

and objectives to be revisited to evaluate the effectiveness of the research. A list of areas for future research will be developed since not everything could be included in a project of this size.

1.6.8 References, Appendices and Index

References to all the works cited are listed.

This project has produced a lot of data. Putting all of that into the text would have spoilt the flow by providing detail where the overview was more useful. Most of the data has therefore been placed in the appendices. These are listed in the table of contents.

The letter granting ethical approval has also been included.

Chapter 2 The Literature Review and Analysis

2.1 Introduction

John Carroll (Carroll 1963) wrote a paper which put forward the idea that, given enough time, most people could learn anything. The paper put forward that:-

$$\text{Quantity of Learning} = \frac{\text{Time needed to learn}}{\text{Time available to learn}}$$

Whilst this seems obvious because it allows space for people to fill gaps in underlying knowledge, his formulation of the concept led Benjamin Bloom (Bloom 1968) to create a system he called Mastery Learning. Bloom argued that teachers know about a third of a new class will do well, a third will pass and a third will fail (or be on the verge of failing) before they even start teaching. He therefore put forward that the syllabus should be split into small sections of knowledge and that students should progress from one item to the next only when they can demonstrate that they have mastered the current item. Students would therefore, with help from the teacher, take the time needed to complete each item. In other words, the time each student takes is not fixed. Some will progress more quickly because they already have a more solid foundation, some will complete all the syllabus despite having to fill in some gaps on the way whilst the final third will not complete the syllabus but can do enough to pass. Bloom therefore rearranged the Carroll equation above to:-

$$\text{Time needed} = f(\text{Time available}, \text{Quantity of learning})$$

His experiments gave a ratio of 1:6 between the most and least able, i.e. if the most able required one hour then the least able required six hours.

CHAPTER 2

If the original set of expectations is a normal curve for the end-of-year results then, Bloom argued, this method would skew the curve to the upper end allowing more students to obtain secure knowledge.

At the same time Fred Keller (Keller 1968), a behavioural psychologist and a follower of B.F. Skinner, developed his Personalised System of Instruction. This was essentially the same system as Mastery Learning, except that students peer taught and peer assessed.

Coincidentally, Robert Gagné (Gagne 1968) also put forward a similar, but rather more rigorous, system which he called Conditions of Learning. Gagné broke his system into three main areas; planning, instruction and evaluation. The start of the planning phase requires the identification of all the learning outcomes and then proceeds step-by-step through the rest of planning, instruction and evaluation for each outcome.

The idea of these non-linear approaches to learning appealed to some educators in the 1960s such as Holt (Holt 1966) and Illich (Illich 1971). They advocated abandoning a set syllabus in favour of allowing students to modify their own curricula. This was shown not to work because students were being asked to create a road map to a place they had never been. The students therefore wandered aimlessly around the countryside of learning as their whims took them. They could not know how to get to a place that they had not been. This caused a backlash against this form of education and the techniques developed by Bloom, Keller and Gagné became less accepted.

To counter the liberal education ideals of the 1960s, successive Secretaries of State for Education in the United Kingdom have presided over increasingly proscriptive educational systems. The introduction of league tables, a national curriculum and targets all run counter to the ideal of individualised learning. All are expected to learn at the same pace.

CHAPTER 2

Whilst much of liberalised learning has been abandoned, the work of Carroll (Carroll 1963), Bloom (Bloom 1968), Keller (Keller 1968) and Gagné (Gagne 1988) on how an individualised system could function still exists in pockets. The syllabus is broken down into small tasks. The student are then allowed to progress through the tasks at their own speed. Crucially, there is supervision included which makes it difficult for the student to just slack off. It is therefore proposed that two points be learned from this:-

1. The syllabus should be broken into small tasks so that all aspects of the syllabus are covered.
2. A monitoring system should exist to ensure that students are progressing and will achieve.

These two points form the context in which the other aspects identified in the next section exist.

2.2 Theories of Learning

2.2.1 Introduction

This review provides an overview of educational theory. This is a vast area and it is obviously not possible to cover everything let alone provide detail. The area has been split into two main branches, learning and teaching, each of which is the split into themes.

The purpose is to provide a list of themes, which may be altered by subsequent researchers, upon which a TEL experience could be built. The system is intended to be adaptable so that new research or alternative views of the topic can be accommodated. This is not intended to be the taxonomy of education but rather a first attempt.

On top of the list of techniques used in the Bloom, Keller and Gagné and systems (peer involvement, small tasks, regular monitoring and the standardised presentation of materials), there are many other tricks in the armoury of the educationalists.

The diagram on the next page shows which general areas have been identified during the review of the academic literature. These ideas presented are representative of the areas of academic research and do not form a comprehensive review.



Figure 1 - Teaching and Learning Theories

Each of the main areas shown above is examined to see what, if anything, can be learned from each area.

2.2.2 The scope of the review

The socio-political environment is not within the scope of this study. This means that a review of the legislation and social circumstances surrounding education has been omitted. These have been left out because they do not impact learning directly, the micro-education, but set the scene for the context in which learning occurs, the macro-education.

2.2.3 The stages of learning

Bloom's Taxonomy (Bloom & Krathwohl 1956) was amongst the first attempts at classifying the stages people would go through when assimilating new areas of learning. The list was split into three domains; cognitive (knowledge-based), affective (emotional-based) and psychomotor (physical activity-based). The cognitive domain is the one used by educators when designing curricula or training materials.

The experiments Piaget (Piaget et al. 1952) conducted on his own children might be seen as unethical now, but produced a series of four stages (sensorimotor, pre-operational, concrete operational and formal operational) through which children and adolescents were seen to progress. The message here was that the form of educational materials presented had to match the stage of development.

David Kolb and Ron Fry (Kolb 1984) developed these ideas into a feedback loop which they called experiential learning. They argued that people cycled through this loop gaining a higher level of understanding each time around. People could start in any position on the cycle, but the educational value was gained by the process of looping.

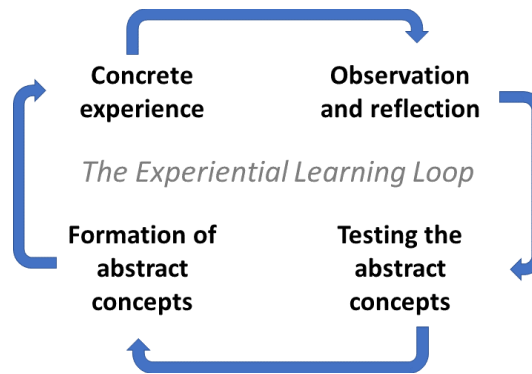


Figure 2 - Kolb's Experiential Learning Loop (Kolb 1984)

2.2.4 Metacognition

According to John Flavell (Flavell 1979), an independent learner will progress more efficiently than one who relies on “Miss/Sir” all the time. They will find ways of solving problems in ways that work for them. These independent learners develop strategies to cope with unknown situations. This definition of metacognition was expanded upon by Kruger and Dunning (Kruger & Dunning 1999) as well as Chabris and Simons (Chabris & Simons 2010). The latter two sets of researchers sought to explain the limiting factors on metacognition, ignorance of the scope of a subject and missing crucial items of learning respectively.

The Kruger-Dunning Effect, as it is known, describes what happens when an individual misunderstands their scope of an area of knowledge. Students were asked to estimate the grade they expected. The Kruger-Dunning study showed that those whose work is in the upper quartile underestimate their grade and vice versa. Someone who feels they have a better grasp of a subject than they enjoy, will underestimate the effort required to pass and overestimate their expected grade.

CHAPTER 2

Chabris and Simons created a short video of people passing basket balls between each other. Viewers were asked to count the number of passes during the video. It is therefore surprising that half the people who watched the video (https://www.youtube.com/watch?v=IGQmdoK_ZfY) did not notice another person, dressed in a gorilla costume, walked in front of those passing the ball and wave its arms about, beats his chest and then walks off. They came to the conclusion that students can miss the obvious when concentrating on something else. For example, a student who is struggling to understand something may miss the very piece of information that makes it clear. The teacher knows that all the information has been given, but the student does not.

2.2.5 Intelligence

Whilst there were others before them, both Alfred Binet and Charles Spearman published work on intelligence in the early 1900s.

In the case of Binet and Theodore Simon (Binet & Simon 1916), the work was used to devise a scale to assess the relative abilities of children within their age group. A list of age specific tasks was created, and children were assessed to see how many of these tasks they could perform. A child who could complete more tasks than their age had a higher “intellectual” age than their chronological age and vice versa.

Spearman (Spearman 1904) noted that pupil scores across a variety subjects were not randomly distributed. Someone with a high score in one subject was likely to score well in others. Spearman called this *g* for general intelligence. This contrasts with the views of Howard Gardner (Gardner 1995) who claimed that there were eight different areas of intelligence (later

amended to nine); musical-rhythmic, visual-spatial, verbal-linguistic, logical-mathematical, bodily-kinaesthetic, interpersonal, intrapersonal, naturalistic existential.

A further twist was added by James Flynn (Flynn 1987) and Reuven Feuerstein (Feuerstein et al. 1985) whose work implies that there is an element of being able to learn to be intelligent. Both argue that environment and educational opportunities shape the intelligence of an individual. Flynn examined current and historical IQ test scores across 14 countries and found that IQs had increased by 21 points (more than one standard deviation) between 1952 and 1981/2 – just one generation. Feuerstein also noted significant gains in IQ when his method of Instrumental Enrichment was used to “teach” intelligence. This would be the interface between intelligence and metacognition.

2.2.6 Learning styles

Learning styles were very popular in the 1980s and ‘90s. The main protagonists, Dunn and Dunn (Dunn et al. 1995), Myers and Briggs (Myers et al. 1985), Honey and Mumford (Honey & Mumford 1989) and Fleming and Mills (Fleming & Mills 1992), saw great benefit in fitting the method of teaching to preferred style of the student. However, the paper by Riener and Willingham (Riener & Willingham 2010) made the case that learning styles do not exist. The position taken here is somewhere between the two extremes. It seems intuitively obvious that people have preferred learning styles, but the longitudinal studies that show these are fixed and immutable appear to be missing. A student may prefer to be taught today but perform better with videos or text tomorrow. One could argue that their preferred style moved with their mood or how they were being taught, etc.

CHAPTER 2

If preferences exist, a class will contain all of them. The person teaching cannot perform in the optimal way for everyone. Even if that were possible, the teacher could not know which is applicable to each student today. This is where technology can step in by allowing each student to select their own preference.

Pask (Pask et al. 1975) postulated a split between serialists and holists. Serialists prefer to see the parts and then construct a view from those, whilst holists prefer an overview into which that parts can be slotted. This causes major problems for educationalists who then must decide which way they want to present in a class. Compounding this problem is that for some subjects we may prefer a serialist approach because we have a background in the subject, whilst new subjects may benefit from holist teaching. Therefore, even these categorisations would appear fluid. The research which backed up this view was provided by John Clark (Clarke 1993) who found that the serialist/holist approach was appropriate for TEL.

Fleming (Fleming 1995) suggested that there were four learning styles; visual, aural, reading/writing and kinaesthetic. Whilst these may or may not exist, some students will prefer each approach. Again, whoever is teaching cannot deliver each approach simultaneously. TEL could give each student the opportunity to select the mode which suits them. Muesser Cemal Nat (Cemal Nat & others 2012) described how a system which analysed learning styles might also be used to help student select appropriate learning modes.

Gregorc (Gregorc 1982), Dunn and Dunn (Dunn et al. 1995), Myers and Briggs (Myers et al. 1985), Felder and Silverman (Felder & Silverman 1988), Honey and Mumford (ibid.) and Hermann (Herrmann 1991) all produced inventories of learning styles which are statistically unique, but integrated in the works of Pask and Fleming as cited above.

2.2.7 Motivation and cognitive dissonance

There are many theories of motivation, but most, such as Maslow (Maslow 1943), McGregor (McGregor 1960) and those involved in conditioning, deal with the general human condition. Those are not relevant to this study because they are not applicable when learning a curriculum.

The idea of positive mental attitude (PMA) underlies this whole project. Whilst Napoleon Hill and W. Clement Stone (Hill & Stone 1991) did not invent the phrase, they did describe it in their book and are credited with having been the originators of the concept in the 1960s. Most school education concentrates on what the pupil cannot do and assumes that “Miss/Sir” knows best. There is some justification for this attitude in that teachers have been there and seen it, therefore have some idea of the way to address this issue.

This generalisation masks the role of the individual in the process as well as ignoring students as co-creators of knowledge, thus hindering their metacognition. The assumption here is that if a student wants to tackle “harder” tasks then there are two possible outcomes; they succeed and feel good or that they don’t and realise that they need underpinning knowledge. The second option is usually called failure and carries a negative connotation. But a computer system is human-agnostic – there are no value judgements attached, so failure turns into experience. Seligman (Seligman et al. 2009) called this positive education and argued that their study based in Geelong Grammar School showed how this had a positive effect on achievement.

The research on how the brain works by Carol Dweck (Dweck 2017) came to the same conclusion. She showed that an attitude of “not yet” yielded far better results than “fail”. She called this the *growth mindset*, i.e. growth assumes a level of trial and error which in turn implies that some learning experiments will fail. Rather than be identified as a problem, these become steps on the way to success. Whilst this approach harks back to the liberal teaching of

CHAPTER 2

the 1960s which was discussed above, the difference lies in the degree of teacher supervision. It is still clear that the teacher is watching to ensure an acceptable rate and quality of work.

The research undertaken by Shen and Hsee (Shen & Hsee 2017) looked at whether awarding scores that were not tied to a specific reward increased performance if the points awarded went up as more was achieved. This seems counter-intuitive at first. One might expect people to try harder if gaining points is made more difficult as one climbs the subject ladder. Shen and Hsee found the opposite to be true and hypothesised that greater reward induced greater effort. Therefore getting 1 point for the first achievement, 2 for the second and so on was found to work well.

According to John Festinger (Festinger 1957), humans will seek to minimise their discomfort by rationalising their situation. An oft quoted example is that of the fox in one of the fables by Æsop in which a fox sees a bunch of juicy grapes that are just out of reach. The fox rationalises not being able to get any grapes by saying that they are probably sour anyway. Students do much the same thing when a topic appears out of reach. They can be tempted to rationalise this as the module not being one which interests them or is not relevant to their career. Some teachers prefer to stretch their students, which is good if the support is there, but can lead to cognitive dissonance if it is not (Harrison et al. 2006).

Alex Michalos (Zumbo & Michalos 2014; Michalos 1985) expanded on this theme by suggesting that dissatisfaction is the distance between what one has and what one wants. These are relative to what others have, the best one has had in the past, what one expects to have in the medium term, what one feels one deserves and what one needs.

2.2.8 Deep/Surface learning

Marton and Säljö (Marton & Säljö 1976) proposed that learners had different aims when studying. Some were genuinely interested in the subject and saw the qualification as secondary – these are deep learners. Others saw the subject matter as a means to an end, e.g. gaining a qualification which would secure them a particular job. All the second category wanted was to pass – these are surface learners.

This categorisation was extended by Entwistle (Entwistle 2000)(Entwistle et al. 1979). He argued that there was a third category, strategic learners. These students swapped between deep and surface learners to optimise their grades and/or prospects.

It is tempting to link this theory with that of Kruger and Dunning. Is there a relationship between deep learners and those who underestimate their grades? Is the converse true? Do surface learners overestimate their grades?

2.2.9 Alternative education

Learning is seen as an individual activity, because it is the individual who is awarded the qualification. The learner is the only one who acquires the learning, therefore, an individual approach to learning was deemed by many to be the most effective. The Socratic Method sought to question an individual until they reached understanding. This is a theme picked up by 1762 story, *Emile – or on Education*, by Jean-Jacque Rousseau in which a boy is brought up through several stages to think for himself.

However, it was the practical educationalists of the 19th century, Pestalozzi (Pestalozzi 1894) and Fröbel (Fröbel 1885) who created, for example, practical “toys” (which Fröbel called gifts)

CHAPTER 2

with which children could learn. These techniques are credited with virtually eliminating illiteracy on Switzerland by the 1830s (Tate 1914).

The next step was the schools set up by Maria Montessori from 1907 in which children were encouraged to explore and enjoy learning (Montessori 1913). These schools still flourish all over the world.

However, the regimented thinking of the first half of the 20th century made this form of education uncommon. It was the advent of the more relaxed social culture of the 1960s that made people like John Holt (Holt 1966), Neil Postman (Postman 2009) and Ivan Illich (Illich 1971) question the teaching methods. This ushered in a more relaxed era in education which went out of fashion again as quickly as it came in. Students were being asked to learn with ever less guidance. Some will have suffered from cognitive dissonance brought on by not knowing what they are meant to be doing. Others will have succumbed to the fun of freedom. However, it was generally agreed that educational standards started to fall and that the gap between private and public education was becoming larger. Therefore, successive governments brought in ever more restrictive and proscriptive educational legislation.

Whilst some still lobby for a more individual educational system , it has become increasingly rare.

However, it is interesting to note that paper-based pockets of Mastery Learning (Benjamin Bloom) and Personalised System of Instruction (Fred Keller) still exist in enough places to make it necessary for John Hattie to include them in his list of meta-analysis studies.

2.3 Theories of Teaching

Section 2.2 looked at theories around learning, how knowledge and skills are acquired. This section investigates the flip side, how can educators influence learning.

2.3.1 The zone of proximal development (ZPD)

This concept was put forward by Lev Vygotsky in 1934, but only came to international attention when his work was translated in 1978 (Vygotsky 1978). It states that people will learn topics most easily if they are close to things they know already. Whilst this sounds obvious, it has a great effect on how curricula should be structured. The difference between this and threshold concepts is that it works in the opposite direction, from known to unknown. But the problem areas are just the same – students may be missing assumed areas of knowledge. Whilst the topic will have been covered in previous classes, there is no guarantee that it has been mastered.

Davis Ausubel (Ausubel 1963) added the idea of teachers having to take the time to find out where the ZPD is for each student. This is an impossible task unless each student is tested on each underlying topic. Even if that were possible, the effort required to bring each student up to speed for their areas of weakness would be enormous. This is an area where TEL comes in its own. However, students can fill in any gaps themselves if they are metacognitive and are allowed to revisit troublesome concepts.

2.3.2 Threshold concepts

The idea was explored originally by David Perkins (Perkins 1999), but was popularised by Meyer and Land (Land et al. 2008). It postulates that students may find a topic difficult because their understanding of a lower level topic is missing. The missing topic must be revisited (or even visited for the first time) before progress can be made. There are many reasons why this might be the case. For example, the sequencing of the topics might be wrong, the student may have been absent when that topic was covered or the concept was not mastered on the first visit. Whatever the cause, progress is being hampered and a way must be found to overcome the hurdle.

If an area like this is left unresolved then either the student may catch up, if the area is relatively small or it could mean that permanent damage is caused. In either case, the educational process is hampered. This makes identifying such areas rather important.

The zone of proximal development (ZPD), as conceived by Vygotsky, could be summarised as:-

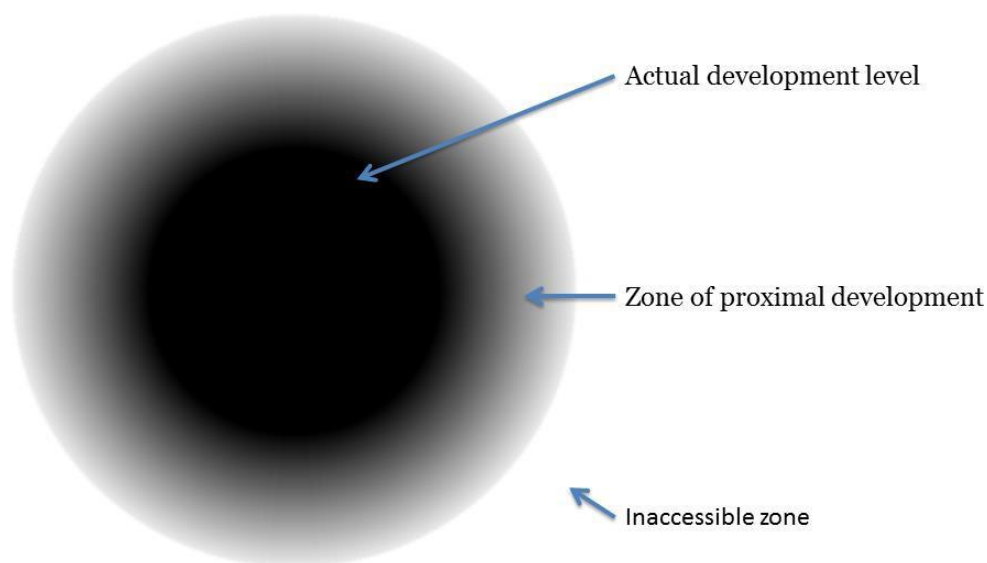


Figure 3 - Zone of Proximal Development (ZPD) (Vygotsky 1978)

However, a more accurate depiction might be:-



Figure 4 - Alternative view of the ZPD

Here, the knowledge is irregularly shaped and includes holes which also have the fuzzy edges of figure 3. These holes may be the threshold concepts needed to progress, but which may not be covered again in later modules. This is an area in which TEL could help if non-linear learning can be accommodated and the materials are easily found.

2.3.3 Experiential learning

This is sometimes confused with hands-on learning. The difference is that experiential learning relies on a level of reflection about the processes being experienced. John Dewey (Dewey 2007) was an early advocate of this approach in the 1930s. It was meant as a riposte to those who only considered a classical education as being worthwhile (Jovanovic 2016). Jerome

Bruner (Bruner 1961) saw experiential learning as an opportunity to allow students to construct their own world view, constructivism.

David Kolb (Kolb 1984) added the idea of a cycle of stages through which experiential learning should go; active experimentation, concrete experience, reflective observation and abstract conceptualisation. It did not matter where the learner started or how often they went around, merely that they cycled through these stages until the scaffold had been constructed.

Bonwell and Eison (Bonwell & Eison 1991) defined the concept of active learning in which students are presented with new concepts which they explore in a tactile way. For example, they might design a database using cards which have tables they need to link “correctly” using pieces of string and paper clips to show the “many” ends.

Situated Learning was described by Lave and Wenger (Lave & Wenger 1991) who felt that learning was primarily a social process where participants in a group help each other. They applied their theory to a wide range of groups such as midwives, butchers and recovering alcoholics. This view has echoes of connectivism which is discussed in 2.3.4 below.

2.3.4 Connectivism and the Medici Effect

Frans Johansson (Johansson 2004) was intrigued by the way new ideas developed rapidly in places where new ideas were already growing. This critical mass of disparate intellectuals who fed off each other creating new ways of looking at problems. He called this the Medici Effect because it was particularly visible in the early renaissance Venice. This free flow of ideas caused others to reconsider their view of their subject area. This is active metacognition and therefore a desirable aspect of education. A student working with another student is good

because they are better able to understand the difficulties involved. However, even better would be if students could connect with others of different disciplines.

The connectivist approach suggested by Siemens and Downes (Siemens 2004; Downes 2008) introduces the use of social media to education. They suggest that just relying on “Miss/Sir” closes out other sources of potential instruction. A peer explaining a concept is much closer to the problem of acquiring the concept than the teacher and social media or built in peer-to-peer communication in TEL can help with that.

MacKinnon and Bacon (MacKinnon & Bacon 2016; Bacon & MacKinnon 2014; MacKinnon & Bacon 2015) improved the concept of virtual scaffolding from their experience of social constructivism. They refined the argument that technology is not and should not be a solitary experience of just interacting with a machine and that a social element needs to be added to make on-line learning a more acceptable experience.

2.3.5 Self-paced learning

There were two experiments in Chicago between the wars which broke the mould of large-class formalised teaching. The first was the Winnetka Plan (Corcoran 1927; Shrock 1995) which was inspired by the work of John Dewey. The other was the work of H.C. Morrison of the University of Chicago (Morrison 1926) who wrote about the need to teach the individual over the curriculum. Both sought to break the subject matter into smaller pieces and required the students to master these pieces. Whilst both were successful, they disappeared without trace after their originators retired during World War II.

As explained in section 2.1, after this war the baton was taken up by Benjamin Bloom (Bloom 1968), Robert Gagné (Gagne 1988; Gagne 1968) and Fred Keller (Keller 1968) and gained some traction in the 1970s. However, the anti-educational liberalism of the 1980s made this approach less acceptable and it only survived in pockets.

2.3.6 Role play

Role play, as described by Edward de Bono (De Bono 1989), requires active student participation and would be difficult to reproduce on a computer. Video conferencing could be used, but why use that when classrooms are available, and students can interact directly? There might be a need for video conferencing if students are widely dispersed geographically, but that is not usually the case. Whilst people world-wide have to study using video conferencing, the recent pandemic may bring it into the mainstream. This aspect of good practice is not being incorporated in this project because the effect of face-to-face learning is outside its scope and would be impractical to implement in a project of this size.

2.3.7 The flipped classroom

The flipped classroom has become very popular recently. Salman Khan (Khan 2015) created the Khan Academy video channel on YouTube in 2006. The idea developed by accident when trying to help his nephews learn mathematics. Rather than teach first and then set homework, the homework is to watch the video and then discuss the contents in class, thus flipping the usual way of teaching. This gives students several bites at the cherry and can allow previous misunderstandings, misconceptions or gaps to be worked on before they become a problem.

Khan has intentionally styled the presentation of materials using the principles outlined by Bloom with Mastery Learning.

Johnathan Bergman and Aaron Sams (Bergmann & Sams 2012), chemistry teachers in Colorado, pointed out that students do not need teachers to impart knowledge, rather they need them when they get stuck. They saw that students felt less threatened by education when they could review a concept or skill and then get help on the sections that concerned them. They found they spent more time with pupils who were struggling and less time explaining concepts with pupils who already understood them (Tucker 2012).

2.3.8 Computer-based immersive education

There have been several attempts at gamifying learning. Sebastian Deterding (Deterding et al. 2011) defined gamification as “*the use of game design elements in non-game contexts*”. A major initiative saw the massively multiplayer on-line game (MMOG) being used to teach key skills in English and mathematics in the late 1990s and early 2000s. Whilst some, such as Paul Gee (Gee 2003), Thomas Connolly (Connolly et al. 2012) and Cagin Kazimoglu (Kazimoglu 2013), found great benefits.

Neverwinter Nights was a game developed by BioWare in 2002. It is set in a fantasy world and follows the rules of Dungeons and Dragons. However, modules could be added. West Nottinghamshire College released a version in which Key Skills in Literacy and IT had been added. It became popular in some colleges for a while but was dropped when the gaming element was found to detract from the education (Sparrowhawk 2007).

Others such as Sebastian Deterding (Deterding 2015) a short while later found 62 reasons why gamification did not work. Bradley Wiggins (Wiggins 2016) found current gaming techniques to be similar to traditional learning materials. Hattie (Hattie 2009) did not have a category for gamification when analysing teaching strategies, but included it under simulations. This category scored 0.34 (where 0.4 is the threshold for an effective strategy), meaning that its usefulness was, at best, minimal. However, this may have changed in the intervening years since the technology has improved and become more accepted.

Increased research in this area may yield strategies that enhance the effectiveness of gamification in the same way that this research aims to enhance TEL.

Therefore, gamification is being excluded from this study for three reasons; its effectiveness seems to be marginal, it takes a lot of resources to develop and not all students enjoy playing games (some may see that as frivolous) all of which means that it is not worth the amount of effort involved by including it in this study.

2.3.9 Adaptive learning

There was some excitement in the 1990s and early 2000s when it was realised that the paths students took through TEL materials could be analysed and predictions made about what was to be presented next. Peter Brusilovsky (Brusilovsky et al. 2004; Brusilovsky 1998; Brusilovsky 2001) and Paul de Bra (Paul De Bra et al. 1999; P. De Bra et al. 1999) led the way in defining adaptive learning as it came to be known. However, there are two problems with this approach.

CHAPTER 2

The first problem is that people change their mind about what suits them, have different areas of expertise or just reject an automated offering. For example, a student studying, say, business studies might see something that excites them about nuclear physics. They do not want to be nuclear physicists, but they are motivated to get a general overview. TEL is not at the stage where this level of telepathy is realistic yet, but the increasing use of artificial intelligence may make this practical in the near future.

The second problem is that a system which decides a learning path, no matter how benign, hinders their metacognitive skills by making decisions for the learner. Any advantage gained by better sequencing could be mitigated out by the loss of metacognitive skills.

Adaptive learning is also complex and time consuming to author. For all these reasons, it has been omitted from the lessons learned.

2.4 Statistical Controversy

One would imagine that the use of standard statistical techniques, ones that have been used by the academic community for a long time, would not be controversial. Standard techniques such as the Student t-Test (Student 1908) and the Fisher F-Test (Snedecor 1934) are standard techniques for assessing differences between groups. These parametric tests were augmented by non-parametric tests such as the Mann-Whitney U-Test (Mann & Whitney 1947), the Levene Test (Levene 1960) and the Brown-Forsythe test (Brown & Forsythe 1974).

However, there has been controversy about their use for some considerable time as Raymond Nickerson pointed out (Nickerson 2000). The criticism does not centre around whether these techniques are mathematically robust – they are – but rather, the circumstances under which they are used. Calculating the t-values is just a matter of arithmetic and there are many pieces of software (such as Excel) or websites that will calculate their values. They can be applied in situations which do not meet the criteria for their use. The effect of this is demonstrated in chapter 4.

Educational studies compare one group with another to assess whether a new intervention (any change made intentionally to the teaching or learning environment) has had an effect. The researcher is therefore looking for any statistically significant differences either between the means or is the magnitude of the variance. The definition of the level of significance required to dispute the null hypothesis depends on the researcher and their needs. In education, if the new mean of the scores is significantly higher than the old mean, then the intervention has been a success. If the variance in the new results is greater than the old variance then the intervention has caused a greater polarisation of results. The criteria, taken from <https://researchbasics.education.uconn.edu/t-test/>, are:-

“ASSUMPTIONS UNDERLYING THE T TEST

1. The samples have been randomly drawn from their respective populations
2. The scores in the population are normally distributed
3. The scores in the populations have the same variance ($s_1=s_2$) **Note: We use a different calculation for the standard error if they are not.**

These assumptions can each be tested.

Assumption1 – *Is the sample random?* : Educational studies are rarely that big that a sample is taken. It is more usual that the results of one group are compared with those of a different one. This assumption is therefore impractical in most educational research.

Assumption 2 – *Are the populations are normally distributed?* : Educational studies deal with human behaviour. The normal distribution was designed to model human behaviour, therefore grades ought to be more or less normally distributed. Tests such as Anderson-Darling, Shapiro-Wilk and D’Agostino-Pearson can be used to test this, but there are conditions applied to each of these too.

Assumption 3 – *Are the variances the same?* : This can be tested using the F-test. It ought to be safe to assume that different groups of students in the same area would have results with more or less the same variance. This assumption may therefore be assumed away by researchers.

A learning intervention is any change made to a teaching environment (academic or physical) which may have influenced student achievement. The measurement of the effectiveness of a learning intervention can be viewed as controversial because cohorts vary. A good year (one where the general attitude is positive) can be followed by one that is poorer – despite having the same entry criteria, being taught by the same people, using the same materials and even in the same rooms.

CHAPTER 2

This makes statistical comparisons between years difficult because there is no guarantee of relating like-with-like. Conventional ANOVA techniques assume that the samples (in this case, the grades of the students) came from the same population. This is not the case here because any measurements compared would have to take two, possibly competing, factors into account; the unknown but natural variation between years and the impact of the learning intervention being studied.

The following thought experiment hypothesises that there are two successive years with the first having a generally positive attitude whilst the second one is generally negative. If an intervention, such as trialling the flipped classroom technique, took place with the negative group, then it may appear less effective because the comparison cohort was more positive. If the years were reversed with the first one being more negative, then an intervention could appear more effective since the group on which the experiment is performed has an inbuilt head start. This is on top of any variation produced by the Hawthorne Effect (Landsberger 1958) which happens when the subjects of the experiment behave differently because they know that they are being watched. This implies two questions; first, do such variations between years exist and second, do any variations between years affect the outcome of experiments?

2.4.1 Statistical tools

There are four statistical techniques which are often used to compare groups of results; parametric tests, non-parametric tests, ANOVA (ANalysis Of VAriance) and graphical methods.

Parametric tests assume that the shape of the distribution is known. This is often assumed to be a normal distribution. There are two techniques used here, the Student t-test (Student 1908)

to assess the probability that the means are come from the same population and the F-test (Snedecor 1934) to assess the relative spread.

The non-parametric equivalent to the Student t-test would be the Mann-Whitney U-test (Mann & Whitney 1947) in situations where the shape of the distribution is unknown. Alternatives to the F-test would be a Levene test (Levene 1960), which is based on the mean of the sample or the Brown-Forsythe test which uses the median.

These would work in a statistical sense, but they any lack of knowledge about an underlying distribution. This makes them less sensitive to changes because they have to take the shape of any distribution into account. Some of the changes due to learning interventions which do have an impact may not be detected using these techniques.

The ANOVA techniques are used where several sets of results are being compared which is the case for the new techniques being proposed. These ANOVA techniques rely on the F-test which has already been discounted. However, a more serious flaw in using ANOVA is that a learning intervention may make the test case behave normally. For example, a mathematics module on a computing programme may be unpopular (they have come for computing and might see the numerical side as less relevant). This could therefore lead to a lower mean score as well as a greater spread of marks. A successful learning intervention (e.g. workbooks as part of the assessment process) may make the mathematics module now appear “normal”, i.e. not significantly different to other modules taken by the same students. It could then be argued that a comparison with the year before would show a change in status from abnormal to normal (for a given level of significance) but quantifying that change and assessing the probability of such a change would not be possible because the students are behaving differently to the way they would normally. The same problems hold true for the Tukey-tests (Tukey 1949) which are also used to examine differences within a group of sets of data.

Finally, graphical techniques, such as Box-Whisker diagrams (Wickham & Stryjewski 2011) give a visual representation how groups differ. This is useful for a first look at the data to see whether further tests are needed, but do not give a quantitative measure of these differences.

2.4.2 Two examples of controversies

There have been many controversies over theories in education, far too many to be explored here, where the use of statistics has been questioned. Therefore, two particular areas of contention already outlined will be explored; learning styles (section 2.2.6) and the work of John Hattie (section 2.1).

Learning styles became fashionable in the 1980s and 1990s, but its origins go back further. Educationalists such as Anthony Gregorc (Gregorc 1982), Rita and Kenneth Dunn (Dunn et al. 1995), Peter Honey and Alan Mumford (Honey & Mumford 1989) and Neil Fleming (Fleming & Baume 2006) showed that their theories worked by using statistical analysis.

The backlash came in the 2010s. It was then that people like Cedar Riener and Daniel Willingham (Riener & Willingham 2010) started to question their use (or even usefulness). This was followed by a closer look at the statistics by people such as Harold Pashler et al. (Pashler et al. 2008) and Joshua Cuevas (Cuevas 2015) each of whom found problems.

A second example of statistical controversy in education would be the work of John Hattie who, as discussed above, used meta-analysis (the statistical amalgamation of several studies) to calculate the impact of a range of strategies used in education. The Times Educational Supplement (TES) describe the first book by John Hattie (Hattie 2009) as “teaching’s Holy Grail” (TES – 21st Nov. 2008). The idea that greater insight into education could be gained by

CHAPTER 2

grouping studies and then analysing each group seemed obvious in hindsight. Over 52,637 studies were grouped under 138 categories. Hattie then used the standard statistical technique called meta-analysis to “average” the effects in each category. The categories were also grouped into six areas.

This quantified the effect that each category had on learning and therefore gave the relative weight that those who were teaching should give to each aspect of their craft.

However, doubts about the way that this technique had been applied began to surface. Ewald Terhart (Terhart 2011) took exception to the work. This has since been backed up by others, including Pierre-Jérôme Bergeron and Lysanne Rivard (Bergeron & Rivard 2017) who called Hattie’s work “pseudoscience”.

There are two issues highlighted here; in some cases, the tool used may not be the best tool as with learning strategies and also sometimes a tool might be used incorrectly as is being claimed for the work of John Hattie.

2.5 A Discussion of Appropriate Statistical Techniques

This research assumes that splitting a group when running an educational experiment is both unethical and impractical. It is unethical if the new technique is trialled on group A and is successful then those in group B are being disadvantaged unless that group have access to the new technique later. That would mean that the groups would have to swap the order of teaching and that may not be feasible. It is impractical since students in one group will have contacts in the other and therefore leakage of any useful on-line materials will occur which would affect the outcome of the experiment. This means that a new learning intervention has to be trialled on the whole year group. As explained in section 3.3.2 below, the results for the group used for the learning intervention will be compared to those of the preceding year. The null hypothesis would be that the experiment had failed and thus the mean (or median) and variance of the score would not be significantly different – in this case at the 5% level. This could be tested using null hypothesis significance testing techniques taught on all introductory statistics courses.

2.5.1 Group comparisons

Almost all statistical textbooks which deal with sample comparisons cover this topic. Statistical tests such as the t-test or F-test are standard tools taught for those dealing with comparisons. The question then becomes one of the effectiveness of differing statistical techniques. In general, there are two types of techniques for comparing groups; parametric and non-parametric.

2.5.2 Parametric tests

Parametric tests assume that the shape of the distribution is known, for example a normal distribution. Many human undertakings are normally distributed which means that researchers may assume that parametric tests are appropriate. There are two tests which are used in this situation; the Student t-test to assess the probability that means from two samples have come from the same population and the Fisher F-test to estimate the probability that the variances of the two samples are from the same population. The t-test measures differences in the location of the two samples, whilst the F-test measures the differences in their spread.

The Student t-Test was created by William S. Gosset (Student 1908) to assess the probability that two samples were drawn from the same population (or that a sample should have a specific mean). If a learning intervention takes place, then the researcher would expect the mean score to alter from the one in its preceding year. However, the mean scores will vary anyway. The problem is to assess whether the level of change is significant. The values obtained from the t-test can then be used to calculate the probability that the variations between the two means is just the normal random fluctuation.

It is possible that the means stay the same, but the learning intervention affected the spread of marks achieved. George W. Snedecor (Snedecor 1934) created the F-test, naming it after Sir Ronald Fisher who had developed the general principles used to calculate the probability that the variance of two samples came from the same population. Again, there will always be random variations in the spread of grades, but the F-statistic can be used to calculate the likelihood that this is just another random occurrence.

Both of these tests assume that the data on which they are based is normally distributed and that the populations have similar characteristics, e.g. age distribution, motivation, previous

educational experiences, etc.

2.5.3 Non-parametric tests

Assuming a normal distribution for every pair of samples would be unreasonable, therefore, non-parametric tests such as the Mann-Whitney U-statistic (developed by Henry B. Mann and Donald R. Whitney (Mann & Whitney 1947)) would be more appropriate. The common assumption is that this compares medians, but their statistic is based on ranking the data and therefore one could have two samples with the same medians returning a statistically significant U-value.

This technique does not assume any particular distribution and is less discriminatory than the t-test. This makes the Mann-Whitney U-test a blunt tool for analysing, what may be, fine differences between the location of the scores of year groups.

Not all changes have to be to the location of a sample. It may be of interest that the spread has altered significantly. For example, a greater spread of grades for a module would mean that some students are achieving higher grades relative to those in the previous year even though the scores overall stayed the same. However, the converse is also true. Those towards the bottom would have to score less well for the location of the grades to remain the same. This may move some into a fail who might have passed the year before.

There are two non-parametric tests being analysed in this study which seek to assess the probability that the spread within two samples is different. The Levene test (Levene 1960) was created by Howard Levene who worked as a statistician in the field of genetics at the University of Columbia. It uses the mean of the samples to calculate the probability that the amount of

spread within two or more samples is the same.

A similar test was described by Morton Brown and Alan Forsythe (Brown & Forsythe 1974). This works in the same way as the Levene test, except that it uses the median in place of the mean. This test has been included because of its similarity to the Levene test.

There are many other tests that could have been used in this study, such as Barlett's test (used to assess the variance and/or the normality of the samples) and the Kruscal-Wallace test (a non-parametric test of difference in variance). Tests such as these have not been included because it is being assumed that the five tests that are included are representative of the other tests.

2.5.4 Analysing sets of modules

Analysis of variance (ANOVA) is a group of techniques used to compare two or more data samples. This technique could be used to examine the scores for each module within a set of modules at a given level and calculate the probability that they all come from the same population using the F-test.

A Tukey analysis (Tukey 1949) works in a similar way. It returns a table of results showing how well each sample relates to each of the other samples. It is then possible to see whether any sample stands out.

These techniques have been road tested since the second world war and are therefore considered robust.

The problem lies in the detail. These tools cannot differentiate between samples moving from an anomalous position to a normal one. They only look to see what is abnormal in the group being assessed now. For example, suppose module A has been taught brilliantly in the past

with students achieving higher than expected results. A new lecturer takes over and module A now performs equally well as the others. Module A would now have become “normal” and would therefore not be flagged as significant by either technique.

It would, of course, be possible to see the movement between years. These techniques would show that module A was significant in one academic year and not in the next. However, there would be no scale. Has module A moved from just above to just below the significances level? How much movement is needed before a statistically significant difference is assumed?

2.5.5 Using graphical methods

There are also more qualitative methods used when discussing the differences between groups. For example, a Box-Whisker plots (Wickham & Stryjewski 2011) will show the relative positions of medians, inter-quartile range and range of sets of data. Plots such as these have been adapted to show relative frequency as well. However, the problem is still the interpretation of the results. What appears significant to one may not to another. This works well where an overview is needed to see whether further analysis would be worthwhile, but less well where an objective measure of difference is required.

2.6 Conclusions

2.6.1 Learning and teaching

Different TEL sites, such as the Khan Academy (<https://www.khanacademy.org/>) or PearlTrees (<https://www.pearltrees.com/>), quote academic educational theories to back up their design. However, no sites have been found which work in the opposite direction, from a survey of educational theories to site design. There is therefore a gap in the literature which either ties theories to design elements or explains why they cannot be converted into practice. For example, a recent report from Athabasca University (Cleveland-Innes et al. 2019) on the design of MOOCs does not mention educational theories.

For the purposes of this study a third category has been added, those theories that could be put into practice but have been excluded because of the time or effort needed to implement them.

As shown in the literature review above, there are too many theories of learning and teaching for a study of this size to cover them all (even if that were possible). However, this study does seek to take a representative sub-set which is large enough to demonstrate how these and other theories may be fitted into a TEL framework.

2.6.2 Analysis

Ulrich Dirnagl (Dirnagl 2019), encountered the same problem with the inappropriate use of statistical tools whilst writing about biology. He cited the experiment run by Craig M. Bennett, Abigail A. Baird, Michael B. Miller, and George L. Wolford and presented as a poster at 15th Annual Meeting of the Organization for Human Brain Mapping, San Francisco, CA, June 2009.

CHAPTER 2

As their Ig Nobel Prize citation read; “*Craig Bennett, Abigail Baird, Michael Miller, and George Wolford [USA], for demonstrating that brain researchers, by using complicated instruments and simple statistics, can see meaningful brain activity anywhere — even in a dead salmon.*” (Craig M. Bennett & Wolford 2010). The point being that the inappropriate use of standard statistical techniques can be used to demonstrate misleading results.

A recent paper by Mathew Makel et al. (Makel et al. 2019) makes this point even more clearly. They asked current educational researchers about their use of a range of dubious statistical practices such as excluding non-significant studies or data or filling in missing data. 14,228 questionnaires were distributed via email and 1,488 valid responses were returned. They found, for example, that 23% of researchers reported filling in missing data and 37% excluded non-significant studies or data. Whilst those who reported that they were relatively early stage researchers reported higher incidences of dubious practice than those claiming to be middle career or late career, the differences were not great. For example, using their table S7, the filling in of missing data was reported as 24.52% for early stage researchers, 21.78% for those who were mid-career and 21.39% for those reporting to be in the late stages of their educational research career. Interestingly, this study used an adaptation of the t-test without any discussion about its appropriateness to this data.

The use of parametric tools is questionable in situations where empirical data is used because the shape of the distribution has not been proven.

The non-parametric tests would work in that situation but are not very sensitive and would therefore give type II errors, false negatives.

ANOVA and the Tukey technique only work in situations where all samples are normally distributed. The interpretation of whether a change shown by these is significant makes them quite subjective.

Finally, as discussed earlier, graphical comparisons are useful for gaining general insights but cannot yield an objective and definitive guide.

This makes the area of the analysis of the effect of learning interventions a gap that needs to be explored.

2.6.3 Revised research question

Regardless of how effective the TEL would be, there would always be a doubt hanging over any analysis of the results because of the question marks hanging over the normality of the distribution of grades within a results set. However, developing a set of statistical tools which even out the differences between years would be a useful contribution. The research question for this study has therefore evolved to be:-

Can the differences between modules and their year groups be used to assess whether a change in pedagogy affected learning outcomes?

Chapter 3 Methodology

3.1 Introduction

The original focus of the research was to examine the link between perceived good educational practice and the design of a site to deliver that good practice. However, the new research question, identified at the end of the previous chapter, shifted the research towards how to test whether the site had an impact. The literature review (Nickerson 2000) (Cohen 2016) flagged up that there was a controversy over how this could be achieved using standard statistical techniques where null hypothesis significance testing is to be used.

The first step would be to design a learning intervention so that any new statistical technique could be tested against a known outcome. Now that candidate educational theories have been found, their applicability to the design of a TEL website can be assessed. These would fall into three areas;

1. those that were appropriate to TEL and capable of being implemented in a research project of this size
2. those that were appropriate but impractical to implement (although they could be included in later versions of a TEL site)
3. those that were not appropriate to being incorporated into a TEL site.

In order to test the effectiveness of a TEL site, a statistical test would have to be designed which counteracted the effect of changes in the make-up of year groups. This statistic could be trialled on a range of modules where the data is available. For the purposes of this research, grade data for computing modules at the University of Greenwich for the academic years between 2011/12 and 2018/19 was available (the ethical approval letter is shown in appendix 27). The statistical test could then be used to assess the effectiveness of the website.

3.2 An Overview and Rationale for the Research Methodology

Research does not follow a linear path. Plans are made, altered in the light of new information and conclusions drawn from the journey.

The original focus of the literature review was on educational good practice. An overview of theories surrounding good educational practice was created. No study of this size could hope to cover them all but the sample, shown in the literature review above (sections 2.2 and 2.3), needed to be extensive enough to provide a solid basis for the recommendations of the design of the TEL website.

The website needs to be tested and statistical tests used to determine its effectiveness. This led to the literature on the effectiveness of some of the standard tests used to assess the impact of experiments in education. Some, such as Arianne Herrera-Bennett (Herrera-Bennett et al. 2020) wrote papers which questioned whether the null hypothesis significance testing (NHST) was being used as intended. Cohorts differ in ability and attitude which may make the use of standard tests inappropriate because they were not designed for such situations. What works one year may not be effective the next and seemingly minor changes, such as moving to a different room, may alter the way the module functions. This understanding has been well documented and will be examined further in Chapter 4. Therefore, the standard statistical tests needed to be assessed for their appropriateness to the changes in grades. As outlined above, it was decided to obtain grade data for 48 modules over 8 academic years and then evaluate each tool to see whether an approximately random set of results was obtained. It was decided to use a 5% significance test for each test. If the test worked then approximately 5% of the inter-year module comparisons should be significant at the 5% level. 5% was chosen because it was big enough to return a reasonable number of results, thus minimising the risk of one or two extreme outliers influencing the outcomes. It is also a level often used in research to distinguish

between “normal” variations and ones that are unusual. Were that shown not to be the case then alternative tests would have to be developed to assess the effectiveness of the experiment.

In either case, the effectiveness of the TEL site could be analysed using each test so that conclusions could be drawn.

3.3 Obtaining the Data

3.3.1 Grades data

As stated above, the data needed to answer the research question was obtained from the grades system of the university after obtaining ethical approval (see appendix 27) for its use. There is no system available for downloading the data directly, so the grades were scraped from the module results pages for each run of each of the modules. A computer program was then written to transform this data into a dataset which allowed an easy analysis of the grades. Two data sets were created, the original full data set and one which excluded all those who were recorded as having scored zero. Scoring zero for submitted work is such a rare occurrence that those managing this were lumped with those who did not submit. This second set was the one used since it excluded those who did not engage with the module. They had less opportunity to be affected by any learning interventions. The generation of these datasets is explained in greater detail in sections 4.2 and 4.3 below.

3.3.2 Experimental website data

The statistical analysis of the grades for the module on which the site could be assessed could then be augmented by data from the use of the website. There were only two possible sources for the data on site usage, Google Analytics and page tracking from within the site. However, the change in the research question means that this data would be redundant for this study.

3.4 Conducting the Experiments

3.4.1 Exploring five standard learning intervention measures

There are other statistical tests which could be used to assess whether a learning intervention was successful, but it would be impractical to evaluate them all. By convention, these tools are broken into two categories, parametric and non-parametric tests. The parametric tests assume that the distribution is known, whilst the non-parametric tests are used on unknown distributions.

A further distinction could be made between tests to assess differences in the location of the data (e.g. its mean or median) and differences in the spread of the data (e.g. its standard deviation or inter-quartile range). Knowing that the average has moved significantly would help to determine whether the learning intervention had a positive or negative effect. Likewise, a determination that the spread of grades has altered significantly shows whether those at either end of the grades spectrum have been affected.

Common tests in each of these categories could then be used as a trial, as shown in the table below. The reasons each test was chosen are explained in their sections in chapter 4.

	Parametric tests	Non-parametric tests
Location	Student t-test	Mann-Whitney U-test
Spread	Fisher F-test	Levene test Brown-Forsythe test

Table 2 - Statistical tests considered.

Each of these tests will be used on the historical grade data gathered from the university system for the 48 modules over 8 years to see which were significant at the 5% level. It would be expected that each of the tests in table 2 would return approximately 5% of the inter-year module comparisons as being unusual. If 5% of the sample is “unusual” then the probability would be that these oddities that need to be investigated. This would be an average so, say, 4% or 6% could happen by chance, but 8%+ (based on the binomial theorem) being returned on a sample of hundreds would indicate a likely problem with the test.

3.4.2 The development of the experimental website

Whilst the website had been the original focus, it became a secondary issue once the more urgent problem of assessing its effectiveness became apparent. Sections 2.3 and 2.4 above describe areas of educational good practice or theories of how the process of education could be improved. There were seven areas of learning and nine areas of teaching described. These areas of educational interest would need to be sorted into three categories:-

- 1) Those which were of no interest when developing a TEL site of this kind.
- 2) Those which were of interest but were either impossible or impractical to for a project of this size.

3) Those which were helpful and could be implemented in this TEL site.

It would also be necessary to take the general principles of universal design (see section 5.1) into account because it is a standard technique to enhance usability, not just for TEL sites, but also for other areas of human/object interaction.

The site was designed to be used as an extra learning resource for students when teaching a module covering the design and development of an Oracle relational database from January 2019 to March 2019. Students would be able to use this resource to augment the lecture and tutorial materials. It therefore qualified as an experimental learning intervention on which the statistical tests could be assessed.

The syllabus will be split into 36 tasks which were ordered according to the expected scheme of work. However, changes often occur with the scheme of work because a topic may be found easier or more difficult by the group than in previous years or a student asks a question which is relevant, but takes the discussion in a different direction, etc. There was no requirement for students to complete the tasks in order. This flexibility would mean that new topics could be added, others could become “for information only” and that students would be able to skip topics with which they were familiar or dwell on those they found more complicated.

The intention was to use the student dashboard to see who was keeping up and who would need to be encouraged. The dashboard is described in more detail in section 5.4 and a screenshot of an example is given in appendix 26.

A task would be defined as being completed when the student achieved 100% on the multi-choice quiz for that task. This was in line with the criteria established by both Benjamin Bloom (Bloom 1968) and Fred Keller (Keller 1968) when they created paper-based equivalent systems. In order to analyse which sections of the web page are useful to the students, the tasks

CHAPTER 3

will be split into different items, each of which is explained in section 6.2. Some of these, such as the title, content and quiz, would have to be included in every task because excluding, say, the title will make the page more difficult to follow or excluding the content will make the page pointless. However, other items, such as the summary and the alternative materials could be used to assess their usefulness by leaving them out occasionally. These items could be hidden on the site by having a variable which indicated their inclusion status, either on or off. Since this was boolean, 50% of the tasks would include, say, the summary (or any other of the switchable items). These switchable items were independent of each other, thus a particular task might exclude the summary but include the alternative materials. The number of extra attempts students needed to pass the task would show the effect this item had on student performance. Any items to be removed would be removed for all students using that task. The plan was to achieve this by assigning each task a random binary number. If the digit corresponding to the item were to be a zero, then that item would not be shown. The items would be treated in this manner:-

Items always included in a task	Items to be missing in 50% of tasks
Title	Summary
Useful sites	Alternative contents
Contents	Underpinning tasks (called reliesons)
Associated materials	External links
The quiz and quiz results	Student generated contents
Approval rating and feedback	Personal notes
Coursework requirements	Links to social media
Usage statistics	

Table 3 - Task items by inclusion

3.5 Analysis of the Results

Five standard tests (the Student t-test, the Fisher F-test, the Mann-Whitney U-test, the Levene test and the Brown-Forsythe test) were to be compared against the proportion of expected outcomes at the 5% significance level. These tests were chosen because they appear in introductory statistics courses and are therefore commonly used in situations where groups are being compared, for example, Cagin Kazimoglu (Kazimoglu 2013).

The controversy (Nickerson 2000) over the application of the standard statistical tests identified implied that alternative statistical tests may need to be found. The five standard tests identified in table 3 would be used as examples which could be applied in the sphere of educational research. Once these had been developed, the next step would be to compare a similar analysis at a 5% significance level using these new statistical tests developed. Again, a significance level of 5% ought to return approximately 5% of the inter-year module comparisons as being significant.

3.5.1 The analysis of student usage of the experimental site

In the short-term, percentage usage and completion would provide an insight into the effectiveness of this experiment because one would assume that more learning would equate to a greater understanding and thus improve grades. Common sense suggests that lower scores would be obtained if were a student were expected to hand in the coursework before the whole of the module has been taught, therefore, more materials would be expected to improve the average grade.

The final phase will be to draw conclusions about the effectiveness of the different statistical tools based on the evidence from both the inter-year comparisons and their ability to identify significant learning interventions.

3.6 Summary

Plans would have to be adapted and the research direction amended to follow new and more promising routes. The study started by looking at the development of a TEL site based on good educational practice but shifted emphasis when it was realised that the analysis of the effectiveness of changes in an educational environment would need to be developed too. Thus the website and the new statistical tests would have to go hand-in-hand. The first stage of this, the assessment of the standard statistical tools, is described in the following chapter.

Chapter 4 The Effectiveness of Standard Statistical Tests

4.1 Introduction

This chapter considers the effectiveness of standard statistical techniques by evaluating the module data using a number of standard tests. As outlined at the end of chapter 3, the effectiveness of the five common statistical tests needs to be established. This chapter evaluates each test against the 5% significance level.

The identifying code for each module has been anonymised and is not the one used at the university. The first digit shows the level of the module (either 4 – first year undergraduate, 5 – second year undergraduate, 6 – final year undergraduate or 7 – Masters). The final two digits are a number assigned to a module. This numbering reflects the order in which the modules were retrieved from the system and not their alphabetical or module code order.

4.1.1 Samples for the descriptive statistics tables

The data covered 48 modules over 8 academic years. This would have given 384 runs if every module had run in every year. Of these, 11 runs did not happen. They were:-

Module	Academic Year(s)
4~02	2015/16, 2016/17, 2017/18, 2018/19
4~03	2018/19
4~04	2018/19
4~05	2018/19
5~09	2018/19
6~01	2017/18
7~01	2017/18, 2018/19

Table 4: Those modules which did not run.

On top of these, module 6~02 in 2017-18 ran with only 3 students – too few generate descriptive statistics was also excluded. It was decided to make the cut-off at four students since this is the minimum needed to calculate the descriptive statistics. This left 372 runs of modules which could be used in this exercise. There were another six runs which had a single figure number of students (see appendix 3) but were included.

4.1.2 Samples for the normality, kurtosis and skewness tables

Some of the normality and kurtosis tests were only applicable if there were at least 20 data items so any run with fewer than 20 students (54 of the 372) were also excluded in these analytical statistical test tables, leaving a trial sample size for these tests of 318 module runs.

4.1.3 Samples for the year-on-year comparison tables

The tests which compared the grades in a learning intervention year with its predecessor to assess whether a significant movement had taken place, meant that a base year was needed when comparing runs (in this case 2011/12), the maximum number of comparisons was 336, i.e. the 384 less the 48 modules in the first year of results, 2011/12. The modules which had less than 4 students were all in later years, so those were discounted too because they would be too small for the descriptive statistics to be helpful. This meant that the maximum number of comparisons was 324. There were a further runs with 5 to 9 students which were kept for the descriptive statistics. Finally, a missing run causes a subsequent comparison to be missing as well, so another 6 were lost bringing the number of usable comparisons down to 318 module

runs – still a significantly large sample for any of the tests. The number of students in each run is shown in appendix 3.

4.2 Cleansing the Grades Data

The student grades system is held on an intranet to which only academic staff (and the appropriate administrators) have access. The results can be displayed by for a particular run of a module. Therefore the 384 runs had to be requested individually and, if available, the scrape was downloaded as plain text. This plain text version had the grades and other associated student information spread over nine lines of text for each result, for example:-

```
000999999
Bloggs, Josephina ..... P11111 BSc ARTIFICIAL INTELLIGENCE
1
30
RE
60
Yes
OE
EPSEP
```

Therefore, the extraneous information, such as programme and fee status needed to be eliminated and the file reduced to a single line of information for each grade awarded (student id, academic year, module name and grade only).

CHAPTER 4

The module 4~01 was split in two for the years 2017/18 and 2018/19. The contents of these two modules together were the same as the original. It was therefore decided to average out the scores for these two modules and treat them as if the original module had continued. This could then be handled as if it were another learning intervention to be tested.

The next step was to aggregate all the files into one dataset to make analysis easier. This produced a dataset containing 30,975 records. As explained in chapter 3, this raw dataset was cleaned by eliminating those awarded zero because of non-engagement. This eliminated 2,365 scores to make a clean dataset of 28,610 records. The following two graphs illustrate the difference between the two data sets:-

Figure 5 - Histogram of the raw data set

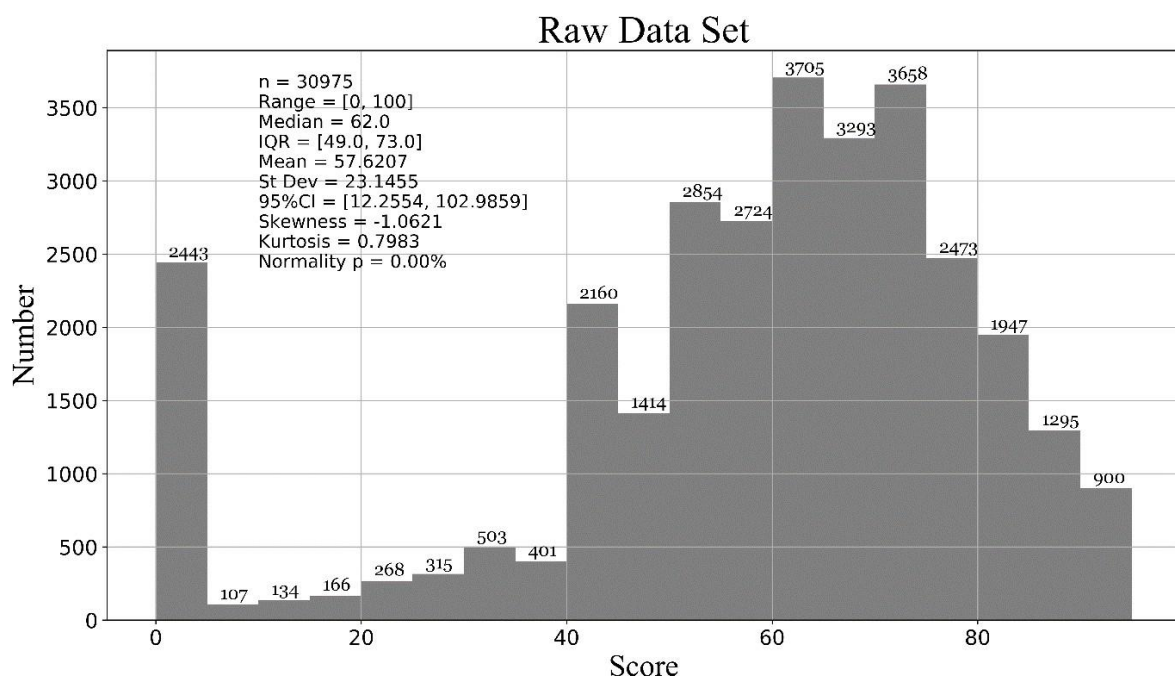
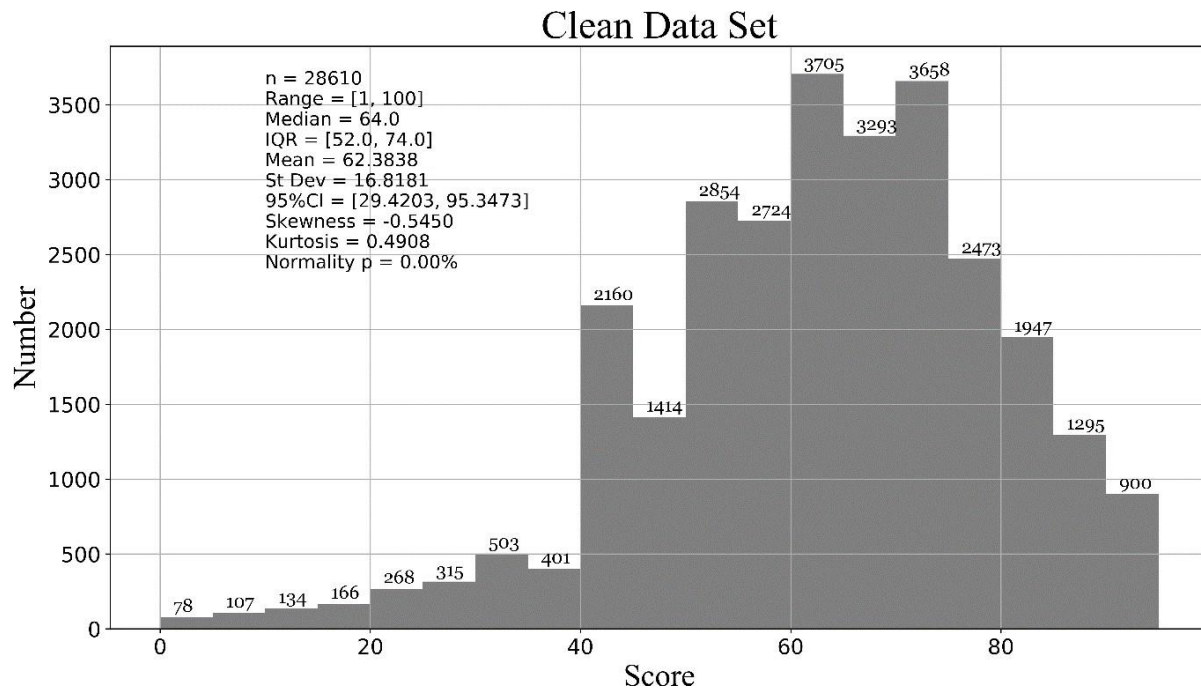


Figure 6 - Histogram of the cleaned data set



These graphs illustrate the skewing effect of including those who did not participate.

Finally, the problem of student anonymity needed to be addressed. A program was written which created randomised student identifiers and altered the module identifiers. The datasets were then sorted into the newly randomised student identifier order so that all name order and/or module order was lost. The sorting routine in the NotePad++ text editor was used to ensure that the related information on each line of data stayed together. This reshuffling of the records meant that the grade-student-module links were retained without a student or module being identifiable.

4.3 Evaluating the Standard Statistical Techniques

4.3.1 Introduction

Statisticians define statistical tests as being either parametric or non-parametric. Parametric tests are those used on samples where the distribution of the data is known. This happens when the data sample is assumed to come from a population modelled by a probability distribution that has a fixed set of parameters and is large enough to be representative of that population.

By contrast, non-parametric tests focus on situations where the probability distribution cannot be guaranteed.

The problem identified by Nickerson (Nickerson 2000) was that the teaching of the null hypothesis significance testing (NHST) tended to imply that most problems could be analysed using parametric tests when non-parametric ones might be more appropriate. Two common statistical problems are analysed here; changes in location and changes in spread.

4.3.2 The 5% test

A learning intervention is usually trialled with the expectation of some benefit for the student or some efficiency in the educational system. In either case, the effect on the grades of the students needs to be quantified. Are the grades going up, down or has the change had a neutral impact? These are changes in the location of the grades of the students.

The learning intervention might have a neutral effect on the location but cause those at the top stretch further ahead whilst the weaker ones regress further. Alternatively, are the stronger

ones reined in whilst the weaker ones catch up? This is a problem of assessing changes in the spread of grades.

It is possible that a learning intervention could affect both location and spread, in which case both types of test could show a significant change.

Common tests in these two dimensions, parametric/non-parametric and location/spread, are analysed below. These are standard tests, the newest of which was created in 1974 and have been chosen because they are well tried and frequently used. Many other tests exist but were not analysed because they are not as commonly taught and therefore as commonly used. For example, neither Bartlett's, Welch's nor Tukey's ANOVA tests have been included.

An experiment was devised to assess whether these tests were appropriate to situations such as evaluating learning interventions. As explained in section 3.2, the acceptance level of 5% was set because it was small enough to be useful (NHSTs are often set at 5%) but large enough to be expected to return a reasonable number of significant values. If the tests were appropriate then they should return approximately 5% of the 318 comparisons as being significant, i.e. around 16. If $n = 318$ and $p = 0.05$ then the 95% confidence interval for the number of tests returning a result is [9, 22], whilst the 99% confidence interval is [7, 25] using the binomial probability distribution.

4.3.3 Student t-test (Parametric measure of location)

The Student t-test was created in its current form by William Gossett in 1908 (Student 1908). He had to use the pseudonym Student because his employer, the Guinness Brewery, did not like its employees publishing under their own names. The t-test is used either to assess the

probability that a sample conforms to a particular value or, as in the cases presented here, whether two means could have been derived from samples from the same distribution.

It is now possible to compare each run with its successor to see whether the means differ significantly at the 5% level.

Appendix 10 shows the table of t-values and appendix 11 shows their associated p-values. The appendix with the p-values table has those comparisons which were significant at the 5% level highlighted. These yield 92 instances (33% of comparisons) as significant at the 5% level which is a considerable way outside the upper 99% confidence interval limit of 25.

4.3.4 Fisher F-test (Parametric measure of spread)

This is the equivalent of the t-test, except that it measures the probability that the spread of two samples come from the same population. George Snedcor published his paper (Snedecor 1934) in 1934 on the F-test which he named after Sir Ronald Fisher who had originally described the concept.

Each run was compared with its predecessor and the F-values obtained are shown in appendix 12. The p-values associated with the F-values are shown in appendix 13. Again, those instances where the p-values are below 5% have been highlighted.

This yielded 51 instances (18%) of comparisons as significant at the 5% level. Again, this was considerably higher than the upper 99% confidence interval limit of 25.

4.3.5 Mann-Whitney U-test (Non-parametric measure of location)

Henry Mann and Donald Whitney created their statistical test and published it in 1947 (Mann & Whitney 1947). Its purpose was to assess the probability of an observation from one population exceeding an observation from another population (either larger or smaller). The observations had to be independent (it is assumed that the student grades are independent of each other) and can be ordered, but no distribution was assumed.

The analysis shown in appendices 15 and 16 show that there were 111 instances (40% of comparisons) as significant at the 5% level. Yet again, this is well above the upper 99% confidence interval limit of 25.

4.3.6 Levene test (Non-parametric measure of spread)

The Levene test is less used in the area of education but is the equivalent of the F-test except that it does not assume an underlying distribution. Howard Levene published a paper in 1960 (Levene 1960) which described how the variances within two samples could be compared. It was designed to be used in situations where there is doubt over the range of spread in two samples being similar. The test uses the variances from the means within each sample as a measure.

The 75 instances (27% of comparisons) where the p-value are significant at the 5% level have been highlighted in appendix 17. The Levene values are shown in appendix 16. Once more, this is a long way above the upper 99% confidence interval limit of 25 for the expected number of significant results.

4.3.7 Brown-Forsythe test (Non-parametric measure of spread)

This test is similar to the Levene test but uses the median instead of the mean in its calculations and is used to calculate the probability that the spread within two samples shows that they have come from the same population. The technique was described by Morton Brown and Alan Forsythe in a paper in 1974 (Brown & Forsythe 1974).

Appendix 18 shows the Brown-Forsythe statistics and appendix 19 their p-values. The highlighted cells show the 62 instances (22% of comparisons) where the p-value was significant at the 5% level.

4.4 Conclusions

These results can be summarised as:-

	Type	Measuring ...	99% Upp. CI	Total
Student t-test	Parametric	Location	25 (8%)	92 (33%)
Fisher F-test	Parametric	Spread	25 (8%)	51 (18%)
Mann-Whitney U-test	Non-parametric	Location	25 (8%)	111 (40%)
Levene test	Non-parametric	Spread	25 (8%)	75 (27%)
Brown-Forsythe test	Non-parametric	Spread	25 (8%)	62 (22%)

Table 4 - The number of instances where $p < 5\%$

As can be seen, none of the standard tests come close to the expected number of instances of significant comparisons at the 99% level of 25. The best results are for the Fisher F-test test which assumes a parametric spread which gave over twice the maximum number of expected significant comparisons. The Mann-Whitney U-test gave over four times the maximum expected number of significant comparisons.

CHAPTER 4

There is clearly something else going on. These are standard statistical tests that have been used extensively in educational research for decades. It seems that these tests are all picking up the variations in ability, motivation, etc. between successive student groups and at a level that makes judging the effect of a learning intervention problematic. Genuine changes caused by a learning intervention can be masked or over-stated by the effect of group changes.

It seems that straight comparisons between groups using standard statistical techniques may be not work as expected in the area of education. This implies that a different set of tools might be needed.

Chapter 5 The TEL Experiment

This chapter discusses the educational design of the experiment to be used when assessing the usefulness of both the current statistical tests described in chapter 4 and the proposed new tests described in chapter 6.

5.1 Introduction

There were two guiding principles for the design of the site to be used by students taking a database module in the experiment:-

- 1) **Universal Design** – This was used to make the site as accessible as possible to the greatest number of students. For example, the design of the site allows students to control the font size, foreground colour and font colour. They can control their speed through the materials and the order in which they work. However, it was not intended that the “ideal” colour co-ordination or layout be a part of this study because it would not aid answering the research question.
- 2) **User anonymity** – It was important that all users be allowed to remain anonymous in case the site were to be hacked. If it were, then little identifying information would be available beyond that email address. Signing up for the website does not require the name, age or location of the user. Users just need to supply a valid email address. Anyone with an Internet link can access the site – that is a design feature. However, a tutor would know the unique numbers of their students when they are accepted onto a run and so would only see the results for these students. The email address people give could be one that has been set up only for the use of this site. An email address is required to send confirmations that tasks have been completed, for student-student and student-teacher interactions, etc. It is therefore important that the mailbox used, be monitored regularly by the user. Each person who registers with the site is given a unique number. This number can be shared with other students and staff in the same way that email addresses are shared – by consent – in situations where interaction needs to occur such a group work or

tutorials. This allows contact via the website between individuals without any personal information being shared.

5.2 The Design of the TEL Site

The term “design” is used here to describe the educational design of the site rather than its physical appearance. Students would be working with a site based upon a broad spectrum of perceived good practice in education.

5.2.1 The educational practice – Theory

The design of the tasks was driven by the literature review. The TEL site is based on the concepts used to create paper-based self-paced learning by Fred Keller (Keller 1968), Benjamin Bloom (Bloom 1968) and Robert Gagné (Gagne 1988) as discussed in chapter 2.

To define some terms in the context of this TEL site, a task is a single atom of learning together with its presentation. It covers a single concept or learning activity. For example, there may be a task covering a many-to-many relationship in a relational database. The presentation is broken into items derived from the literature review. The work on this task would be presented with a title, a brief summary, etc. Each of these parts that go to make up what the student sees is an item. An example of a task with the items annotated is shown in 5.2.17 below.

Items below which end with an asterisk (*) are intended to be part of an experiment to assess their usefulness in on a TEL site. There would be a 50% chance that an item would not appear for a task. If an item was excluded, then it would then be excluded for all students looking at that task since it would be unethical to disadvantage some learners. These items are:-

- Summary (1) - A single sentence describing the contents of the task.
- Alternative contents (2) - Video, audio and PDF formats for the content of this task.
- Reliesons (3) - Links to tasks which are the immediate prerequisites for this task.
- External links (4) - Links to alternative materials for the content of this task.
- Quiz (5) - A self-assessment quiz used to check whether the task has been understood.
- Student generated content (6) - Additional links/comments to help others.
- Personal notes (7) - Notes made by the student to help their understanding/retention.
- Social media links (8) - Used to connect with others.

5.2.2 Title

The title is the headline for a task as it helps with orientation. It is a scaffolding item which helps the user to orientate their thinking by giving a context to the work. The term scaffolding was coined by David Wood (Wood et al. 1976) who was working with Jerome Bruner on the role of tutoring. The term was used to describe the act of providing direction without giving the solution, in line with earlier work by Jerome Bruner (Bruner 1961). This was later linked to the concept of the zone of proximal development (Vygotsky 1978) by many others, for example by C Addison Stone (Stone 1993). Later, Lachlan MacKinnon and Liz Bacon (MacKinnon & Bacon 2016) brought the concept into the digital age and it is in the spirit of this definition that the term scaffolding is used here.

5.2.3 Summary *

The summary is a short phrase, single sentence or brief overview which describes the contents of the task. As with the title, this is a scaffolding item which leads the user into thinking about the subject matter in general. However, it can also be seen as help for holist learners as described by Donald Pask (Pask et al. 1975) because it provides the bigger picture before the detail is revealed.

5.2.4 Useful sites

This item includes links to sites such as Evernote, search engines, on-line graphics, etc. These links are included not just to make the life of the student easier, but also to help with a feeling of community. The intention is to allow students feel that they can stay on-site and still be in contact with others. This is in-line with the connectivist ideas of Siemens and Downes (Siemens 2005) as well as MacKinnon and Bacon (MacKinnon & Bacon 2016).

5.2.5 Contents

The contents comprises the teaching materials, which may be composed of text and graphics. This is the exploration of a single self-contained atom of learning. It is intended that it should be self-sufficient, i.e. could stand on its own. This conforms to the ideas proposed in The Personalised System of Instruction (Keller 1968), Mastery Learning (Bloom 1968) and Conditions of Learning (Gagne 1988) as described above which guide this experiment.

5.2.6 Alternative contents *

Whilst learning styles were dismissed by Riener and Willingham (Riener & Willingham 2010), it may be that learning preferences are not fixed. Neil Fleming (Fleming 1995) argued that there were three learning styles, visual, aural and kinaesthetic. However, he overcame the categorisation problem by claiming that people like a mixture of these. One learning activity can only be used in one way (a video must be watched, an audio file played, etc.) and these are

unlikely to be used simultaneously. Therefore, the materials are presented as video, audio and PDF formats for the content of a task which acknowledges the fact that different people will approach a task differently. Presenting the materials in video for those who want a visual presentation, MP3 for those preferring to listen and PDF for those who want something in their hand rather than just on-screen is helpful and encourages the student to take control of their learning. These metacognitive skills (Flavell 1979) can be developed and then transferred outside TEL. Making students more metacognitive is thought to be a useful skill. For example, Musser Cemal Nat (Cemal Nat et al. 2011) argued that the development of these skills could be enhanced by the use of technologies which allowed users to control the speed and direction of their studies.

5.2.7 Associated materials

The associated materials are handouts in text, sound or video format that go along with a task. These are extra content to that described in 5.2.5 and 5.2.6 that the students might find useful but are not part of the explanation of the topic being considered. For example, a task on a topic would have its own video attached (see the section above), but the tutor may have found other videos which could also be helpful because they explain the topic from an alternative angle. There might be materials, such as alternative websites or pointers to sections within a library, which could also be useful.

5.2.8 Pre-requisite knowledge *

Pre-requisite knowledge includes links to tasks which are the immediate prerequisites for this task. The Vygotskian view of learning as being linked to the zone of proximal development (Vygotsky 1978) makes it sound as if the knowledge is on firm foundations. However, we all have gaps in our understanding. These gaps make acquiring new knowledge harder. These gaps can then become threshold concepts (Land et al. 2008; Meyer & Land 2003) which either slow the learner down or even stop them in their tracks. In either case, the experience is a demotivating one for the student. The pre-requisite knowledge item aims to provide links to this missing underpinning knowledge without the need for searching or admitting a deficiency. The links might be to other tasks in this syllabus or to, say, websites providing that service.

The term used for this here is a relieson (relies on). If over time enough relieson links could be developed by different students and authors of tasks, then a map of underpinning knowledge could be produced. These links could be developed over time as tutors find students having different problems.

5.2.9 External links *

This item includes links to alternative materials for the content of this task created by the author or others. This empowers students (as well as staff) to help other students by providing links, materials or alternative explanations (possibly in other languages). Whilst there will always be those who want to add inappropriate materials. This could be policed in a similar way to comments on FaceBook or YouTube, by peer moderation. These interactions with the site encourage a feeling of community (connectivism (Downes 2012)), the ability to grow knowledge collaboratively (the Medici Effect (Johansson 2004)) and metacognition (Flavell

1979). As MacKinnon and Bacon (MacKinnon & Bacon 2016) show, this constructivist approach helps all learners because explaining something consolidates it for the person explaining who also has greater empathy for misunderstandings.

5.2.10 Quiz and quiz results

It has been hypothesised that there are three different kinds of learners (Lublin 2003); deep (those who are genuinely interested in the materials), strategic (those who want to pass because the qualification is important to them) and surface (those who will do the minimum). The self-assessment quiz is used as a check by the student whether they have understood a task. The quiz is short (typically three or four multichoice questions) and the results are returned immediately. The student is able to take the quiz as often as they want and they are deemed to have “passed” the task when they achieve 100%. This does not necessarily imply that they have achieved any marks in their coursework. The gamification of learning is not fashionable, as Sebastian Deterding (Deterding 2015) pointed out. However, the tutor would be able to see results of these quizzes, both the number of attempts and the results of each, and thus help the student with any troublesome content during any one-to-one tutorials.

As Deterding (*ibid*) shows, scoring is not always helpful. However, research by Luxi Shen and Christopher Hsee (Shen & Hsee 2017) came to a different conclusion. They described a system they called numerical nudging. Instead of the user getting easy points at the start of a game to encourage them to go further, they showed that a system that increased the points awarded as you went up the scale improved the motivation to complete more. This system would award 1 point for the first task completed, 2 for the second successful completion, 3 for the third and so

on. The module used for this experiment has 36 tasks which gives a maximum point score of 666.

5.2.11 Personal notes *

This item consists of notes made by the student to help their understanding/retention of the material. Repetition has been shown to aid retention (Topping et al. 2007), therefore a student making notes could be expected to remember information better making future tasks easier. However, the article provides evidence this concept is ambiguous since both the quality and quantity of the repetition is important. The quality of the retention is governed by the quality of the feedback received each time a task is repeated and the quantity is the number of repetitions. Whilst there would be no feedback on these notes, the repetition would be useful.

5.2.12 Approval rating and feedback

The approval rating is a star system where the student selects between 1 and 5 stars (a semantic differential scale) depending on how useful they found this task together with feedback to the author on what was good and what can be improved. This looks as if it is for the content creator rather than the consumer of that content and feedback on the perceived quality of the content is always helpful. However, encouraging the student to evaluate their learning experience is one of the metacognitive skills (Flavell 1979) and the act of communicating that experience to others may also be beneficial. Another benefit is engagement which helps to form a two-way relationship between the teacher and the taught. This helps to lower barriers and reduce the authoritarian aspects of teaching. Writers from Socrates (Stonehouse et al. 2011) to Sir Ken

Robinson (Robinson 2008) have argued for this to happen. The site then encourages the teacher to become a guide rather than an enforcer.

5.2.13 Social media links *

The use of social media may sometimes be seen as a distraction, but it can be used to connect with others with the intention of finding alternative sources and techniques. This connectivist approach has been shown by Stephen Downes and George Siemens (Downes 2012) to have advantages. Students working together (as opposed to one doing the work and the others “borrowing” it) encourages students to come at a problem from different angles thus arriving at alternative solutions or using different methods of solving the problem. These networks exist regardless of whether teachers like them or not, so harnessing the positive aspects could be expected to help. This Medici Effect, as described by Frans Johansson (Johansson 2004), has a firm historical foundation. Using a TEL merely brings the process into the age of the Internet.

5.2.14 Coursework requirements and additional coursework materials

This is a small item of work set by the tutor to assess as an exercise or as part of a piece of coursework. The feedback is given by the tutor on whether the requirement set has been completed and if not, suggestions for improvement. This item on the site should include a statement encouraging positive feedback as encouraged by Carol Dweck (Dweck 2017; Dweck 2010) in her work on a growth mindset. This also provides an incentive for a more supportive learning environment, a positive emotional contagion as outlined by Amit Goldenberg and

James Gross (Goldenberg & Gross 2020). A negative approach to feedback used to be thought of as encouraging. However, the work of Leon Festinger (Festinger 1957) on cognitive dissonance – the feeling that effort was pointless since criticism was coming anyway – showed the opposite to be the case.

5.2.15 Usage statistics

Students tend to be social animals and therefore like to see how their position in the system compares to others (Shen & Hsee 2017). The overview of the site usage aims to provide both that place in the hierarchy (without the associated value judgements) and a feeling of community (Bacon & MacKinnon 2014). Only the student would know what their current score is, but they can see how that score compares to the group as a whole. The Kruger-Dunning Effect (Kruger & Dunning 1999), which explains why some students feel they have achieved more than they have whilst others underestimate their work, is mitigated by displaying the reality of their relative position. This item therefore is aimed at aiding the development of metacognitive skills which are transferable from subject to subject.

5.2.16 A sample task

Below is an annotated screen shot of a task. The items have been numbered and the description corresponding to each number is listed on the next page.

The screenshot shows a web interface for a course titled "An Introduction to COMP1556 (2)". The page has a blue header with navigation links: MY OFFICE, MATERIALS, QUIZ, COURSEWORK, LINKS, and FEEDBACK. On the left, there is a sidebar with a "My Office" section containing links to Materials, Main contents, Videos, MP3 and PDF, Student generated content, Related tasks, and External links. Below this is a "Knowledge Check" section with links to Quiz, Coursework, and Additional materials, followed by a "Feedback" button. Further down are sections for "Searching" (with icons for Google, Bing, and Yahoo), "Resources" (with icons for various social media and sharing options), "File Storage" (with icons for various file storage services), and "Social Media" (with icons for various social media platforms). The main content area is titled "An Introduction to COMP1556 (2)" and contains a description of the course: "This is a level 4 (2nd year undergraduate) course in database design. Here you will learn about why a database has to be normalised, how to design a sustainable database and how to use SQL to query and manipulate a database." Below the description are three tabs: Video, MP3, and PDF. The main content area is divided into several sections: "Introduction" (welcoming students and listing three areas of focus), "The design of sustainable databases" (discussing object-oriented relational database design and normalisation), "The creation of databases filled with dummy data" (discussing the use of LucidChart and data generation packages), and "The use of real-world SQL" (discussing the use of CREATE TABLE and SELECT commands). The page also includes a "How will you be assessed?" section and a footer with copyright information: (c) 2018, 2019 LYVA Life. The page is annotated with six numbered arrows: 1 points to the course title, 2 points to the course description, 3 points to the Video, MP3, and PDF tabs, 4 points to the main content area, 5 points to the sidebar, and 6 points to the Social Media section.

5

1

2

3

4

6

MY OFFICE MATERIALS QUIZ COURSEWORK LINKS FEEDBACK

My Office

Materials

Main contents

Videos MP3 and PDF

Student generated content

Related tasks

External links

Knowledge Check

Quiz

Coursework

Additional materials

Feedback

Searching

Resources

File Storage

Social Media

An Introduction to COMP1556 (2)

A different task? View this task

This is a level 4 (2nd year undergraduate) course in database design.

Here you will learn about why a database has to be normalised, how to design a sustainable database and how to use SQL to query and manipulate a database.

Video MP3 PDF

Introduction

Welcome to COMP1556. This is a level 4 (second-year undergraduate) course which will cover the following three areas:-

1. The design of sustainable databases
2. The creation of databases filled with dummy data
3. The use of real-world SQL

The design of sustainable databases

We will be using object-oriented relational database design (OORDBD) which has some significant advantages; you do not need to go through the three levels of normalisation, your database will be in fourth normal form (4NF), writing the SQL is easier and making changes when the client says, "I've had an idea ..." does not require renormalising your database.

There is a six-step process, the last of which is adding the fields which would be the last step in conventional normalisation.

The creation of databases filled with dummy data

You will see how you can create all the tables and their fields without typing a single line of code! You can do this by creating your ERDs in an on-line package called [LucidChart](#). There is a link which, when clicked, shows all the SQL needed to create the tables. You just copy that those commands into your DBMS, run it and the tables are created. This is free for academic users.

You can then use a data generation package, such as [GenerateData.com](#) or [Mockaroo.com](#), to populate these tables with hundreds of rows of dummy data. This makes testing your queries MUCH easier. But (and there is always a "but"), you do have to work out the order in which the tables have to be populated because the PKs and FKs must be synchronised - and that requires planning.

The use of real-world SQL

Most people on this level, of course, will have used the CREATE TABLE and SELECT commands. However, there is much more that SQL can do. This course takes you through the more complex areas that are used in real situations.

How will you be assessed?

You have a coursework to complete. Most of the coursework is done in pairs - you find a buddy and team up. You register with whom you are working by telling your tutor.

The coursework presents you with a scenario. The first task is to create a conceptual ERD (along with any sub-classes). You then create the physical ERD which you then use to create the tables in MySQL. The next step is to create a plan of how the tables are to be populated and get the data in there. These steps form the bulk of the coursework and are completed with your pair.

The final steps are to create some SQL queries (there are two sets and you decide who will do which set). Finally, you evaluate the work you and your pair have produced in an academic way.

(c) 2018, 2019 LYVA Life

Safe browsing Exploit mitigation Courseware Encryption

Figure 7 - Examples of items within a task

Key:

1. The title item
2. The summary item
3. The alternative media item
4. The contents item
5. Links to the quiz, coursework and additional media items
6. Links to the social media and resources items

5.2.17 Summary of theories

The work above shows that technology enhanced learning using good educational practice based on the literature is possible. It is, however, unrealistic to assume that a) all theories have been covered here and b) that there will not be any new ones in the future which affect the TEL design. Therefore, the list of items presented above has to be flexible, allowing new items to be added and existing ones to be merged or dropped if necessary. This list is a starting point and not the definitive version.

Theories such as the role of intelligence and stages of learning have not been considered here even though they are relevant. Carol Dweck (Dweck 2017), Reuven Feuerstein (Feuerstein et al. 1985) and Daniel Willingham (Willingham 2009) all see intelligence as at least partly learned and therefore aided by “good” TEL. The stages of learning described by Rudolf Steiner (Steiner 1989), Jerome Bloom (Bloom & Krathwohl 1956), Jean Piaget (Piaget et al. 1952) and David Kolb (Kolb 1984) are treated as implicit here – the tasks should be presented at a level appropriate to the learner.

Finally, only the theory of adaptive hypermedia, the directing of student studies based on an algorithm which was championed by Paul Brusilovski (Brusilovsky 1998; Brusilovsky 2001) and Paul de Bra (Paul De Bra et al. 1999) have been discounted because they run counter to

the objective of making students metacognitive. A student who can direct their own learning is assumed to be at an advantage over one who has that direction provided for them.

5.3 The website – Practice

Having established the theoretical background to the design, attention now moves to its implementation. This section discusses how the site would work for both students and staff.

5.3.1 Site philosophy

The general concept for this web site started in 2005, several years before the start of this research. Virtual learning environments in use at the time, for example Blackboard and Moodle, did not offer students the opportunity to explore, travel at their own pace or have their progress tracked automatically. Students were bound into a learning pattern determined by the lecturer which made students reliant on them and encouraged the “Miss/Sir” mentality of dependence rather than the metacognition advocated by John Flavell (Flavell 1979). The aim of this experimental site is to encourage the opposite, independence and metacognition.

There was no expectation that students should take a linear approach to their learning. A student studying, say, computing might find it helpful to look up colour theory when designing a user interface. Someone may want to browse a topic, say, archaeology or nanotechnology out of interest. These “minor” achievements are not normally recorded. However, an employer may prefer someone who can demonstrate a wider range of understanding.

A basic principle was privacy and anonymity. This has obvious benefits for the user, but also for the site. Users would not be required to give any personal details beyond a valid email address. Those using the site for the experiment were advised to use either their university

email address or to create one which they only used when accessing the site. Those studying a module can use their institutional email address if they wish, but that would not be as secure since someone knowing the institutional email address would know who they are. Whilst this is unlikely in general, the possibility still exists. Using an anonymised email address means that the educational achievements cannot be tied to an individual. Students at an institution just give their teacher/lecturer their unique site identification number. This number is created by the site and is the automatic id generated when someone applies to use the site. This can then be used by that institution in the same way as, say, a personal email address.

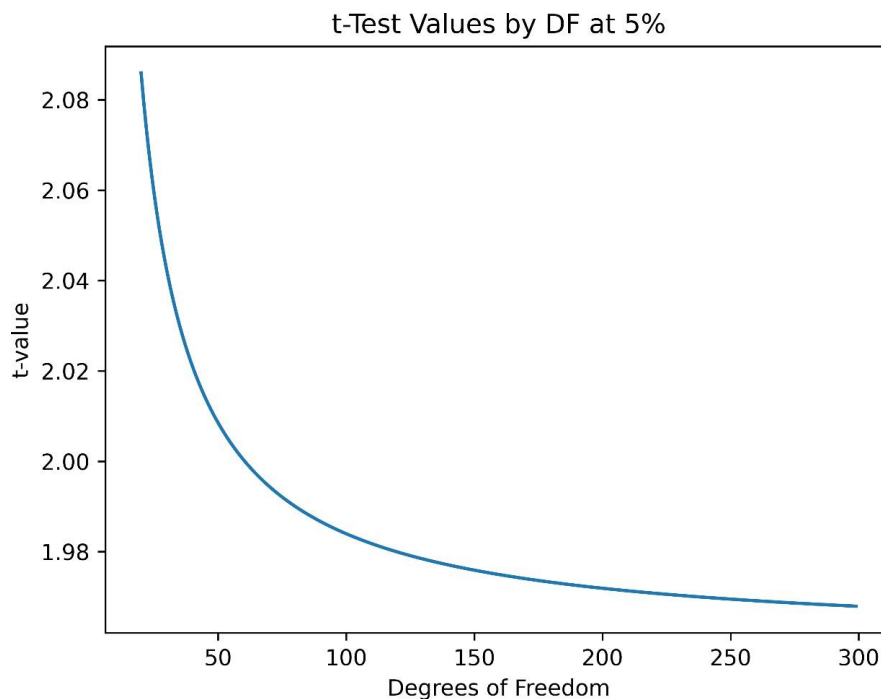
It was always intended that the site will be opened up for general use once this project is completed and so the same advice will be given.

5.4 Selecting the Experimental Group

The cohort to be used had to be big enough for statistically significant comparisons to be made, but the larger the group, the more accurate would be the results. However, this advantage tails off as the sample size gets bigger. The jump from a group of, say, 20 to 25 is likely to be more significant than moving from 100 to 105 because the accuracy should not alter much with large samples. The statistical tests analysed in chapter 4 need a sample size of at least 20, therefore this would be the minimum acceptable cohort size. The groups being taught by the author also drove the group selection. Initially it was intended to use those taking a first-year module in programming for the experiment, but there were some problems. Whilst the group would have been large ($n > 240$), the assessment method, the module team, and the running of the module, varied year-on-year. It was therefore decided to use a second-year database design group ($n = 109$) which had been stable for both content and teaching team for some years. It was felt that

the differences in the size of these groups would make less difference to the outcomes than the changes in assessment and teaching team. A sample size of 109 is statistically similar to one of 240 because the standard statistical tables are very similar over the range of 108 and 239 degrees of freedom. For example, the t-test statistic moves from 0.676 to 0.675 over this range of degrees of freedom (Wetzel Wolfgang 1967) at a 5% significance level. The difference in sample size would only become apparent in the third decimal place in the value of t. This is illustrated by the tapering off of the tail in the diagram below:-

Figure 8 - The value of t needed at the 5% significance level by degrees of freedom



As was shown in chapter 4, comparing one year to another using, say, a t-test would not reflect the reality of teaching. The outlook of groups varies significantly from one academic year to the next without any changes in teaching. This may seem like hyperbole, but the analysis in chapter 4 demonstrates that this is the case for the dataset of grades in this study.

The analysis in chapter 4 also demonstrates that comparing a “good” year to a less positive year, or vice versa, would skew the results. For example, if a really good year, in terms of

results, was to be followed by a poorer year than any positive effects of a learning intervention could be hidden by the differences in year groups. If the roles were reversed, a poor year is followed by a good year, then a learning intervention would seem more positive than it was.

5.5 The Running of the Experimental Site

The system proposed contains two parts, the study part and the tutorial part. The study part is where the student works through the TEL materials and is used to give/take instruction. The tutorial part is used by the tutor to monitor and record student progress. This allows one-to-one (or one-to-a-few) interaction making the process more personal. It is the area that seemed to be missing during the liberal educational practices of the 1960s.

5.5.1 The study system

A run is a module taught in a single academic year. The students on that run are a single cohort. The fact that these students are assigned to different classes is not relevant since each class will get the same input. However, runs cannot span academic years. A tutor must set up a run before anything else can happen. The run has a name, start and end date, etc. plus any other information that the tutor deems necessary in a comment box. The tutor would then need to decide which tasks are needed to cover the syllabus.

To be valid and viewable by a student, a task must have a title, a summary, contents, video/audio/PDF versions and a quiz of at least three questions. Other learning items described above may be added as needed. It is envisaged that these tasks could be reused year-on-year or included in other runs, making them reusable. The tasks could be updated and typos/errors

removed on the fly without students having to re-download newer versions. No tasks were added during the run of the experiment, but some typographical errors were corrected.

Students can be enrolled once the tutor is confident that there is sufficient material available for the students to be allowed in. The needs of the students may not always be those predicted, so new tasks may be added or current ones tweaked throughout a run.

The students can work through the tasks in any order. Some may take longer on one task but need less time on another, some may progress faster and others slower. Each student would be able to move at their own pace rather than the more able becoming bored because they have finished or the weaker ones getting left behind which is what can happen when a cohort is required to learn at the same pace. Students are likely to feel more secure if they are confident that they have understood the material that they have covered, even if that means that they have not completed the syllabus, rather than if they have rushed through a syllabus without gaining understanding.

A student selects a task, works through the materials using whichever format(s) suits them best (on-screen, video, audio or printed) and then takes the quiz. They can retake the quiz as often as they need and the feedback on their performance is immediate. They will be told which questions they got right. Mastery is achieved when they score 100% on the quiz. Each student will be able to monitor their own progress via their home page. They will know their on-site score. They can then compare their score to a table shown on the home page. This league table provides a second level of comparison since they can compare their score against the generality of their peers and against the world in general. The scoring system used is that developed by Shen and Hsee as explained in section 5.2.11 above. This score is not necessarily the mark for the module. That mark could be determined by a coursework based on the syllabus and marking scheme.

This process is repeated for each task (or until the run ends). Tutors can give help and advice on a one-to-one basis or might see students having a similar problem in a small group as is explained in the next section.

5.6 The Tutorial System

The discussion in section 5.5 covered the learning aspects of the site. This section looks at how the tutor would interact with the website.

5.6.1 Introduction

The materials presented on the site would not replace the teaching or materials currently in place. They would, however, provide the reinforcement of the teaching as well as the ability to monitor student progress in the tutorials. The website would not be the coursework, but rather provide feedback and encouragement.

Once the module is underway, the tutorial work can begin. Every student is given a unique number when they enrol on the site. They pass on to this unique number to their tutor. The tutor would also be registered on the site and have a unique number which they can give to their students. This ensures confidentiality because those outside this module would not know these identifiers. The students and tutors would know that any comments or results cannot be linked to them, but only by those included in that run.

Where possible, each student would be seen on a regular basis, say, each week. During that meeting the tutor would review the progress of the student and give help where necessary.

CHAPTER 5

They would then record their comments on the system so that the student could see them too and a brief record is kept of the meeting. For example, the tutor might set a target for the work to be completed by the following week or suggest avenues the student might explore to overcome a problem.

These two systems have been summarised in the flowchart on the following page.

5.6.2 How to use the experimental website – A flowchart

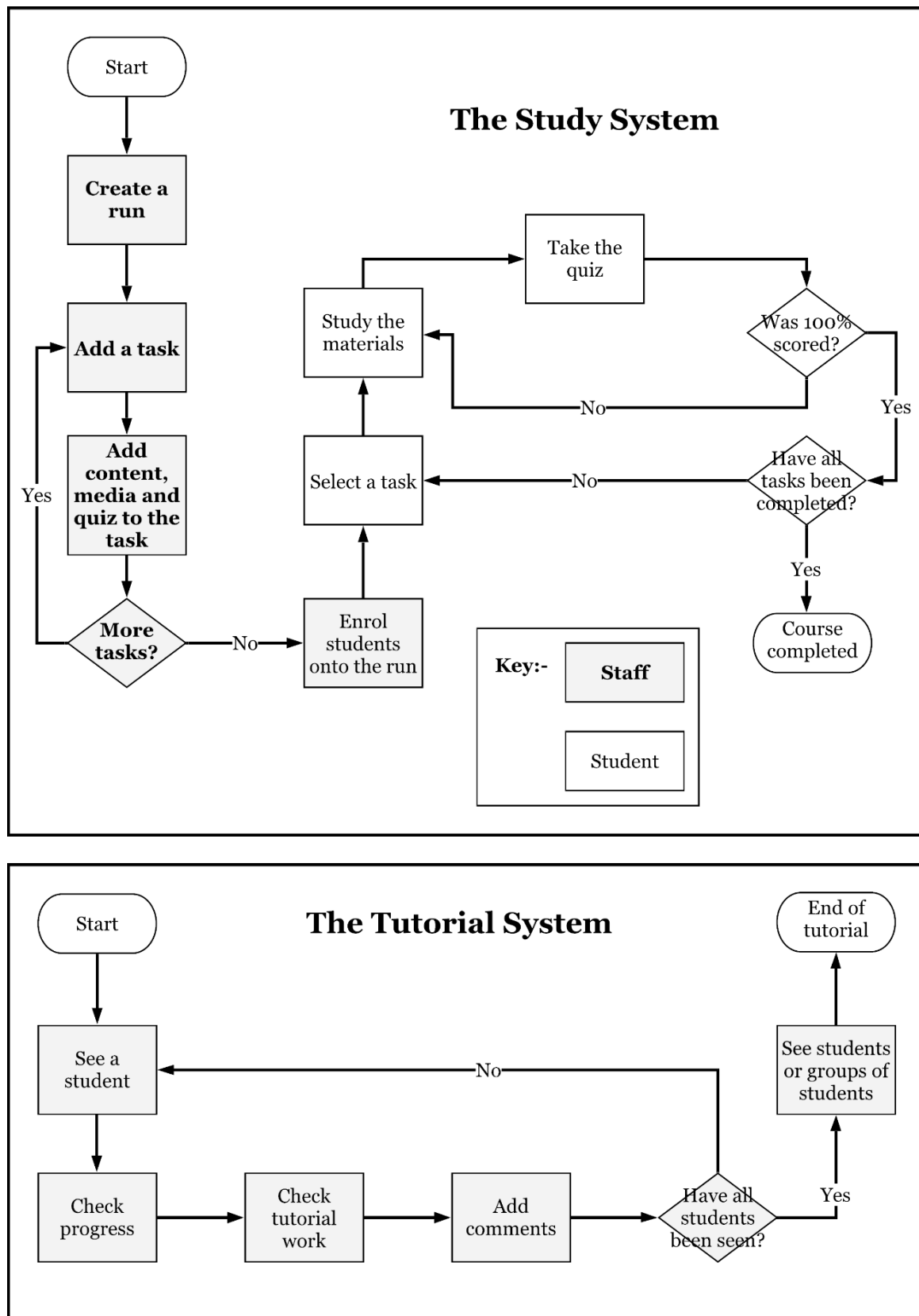


Figure 9 - Experimental Site Flowchart of Usage

5.7 Conclusions

It has been established in chapter 4 that the five standard statistical tests may provide misleading results and this chapter explores an experiment which might be used to test the new statistics explained in the next chapter.

This chapter has examined both the theory and practice of the experiment which will be used as part of the answer to the research question. The theories examined in the literature review were linked to items on the experimental TEL site and the way that the site would be used was explained.

Chapter 6 The Learning Intervention Impact Metrics (LIIMs)

6.1 Introduction and basic premise

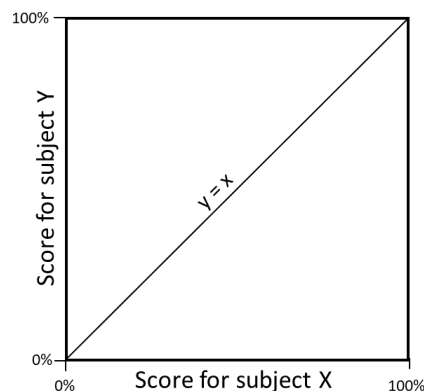
Chapter 4 established that finding a possible alternative to the five standard statistical tests for location and spread might be helpful. This chapter turns its attention to how these alternative tests might be derived. The lesson from chapter 4 is that any comparisons made, have to be like-with-like, which does not happen when different cohorts are compared.

Let us assume that a cohort of students takes two modules and finds these to be:-

1. equally difficult
2. assessed to an equal standard
3. taught to the same standard
4. equally provided with access to learning materials and resources.

In the unlikely event that these four criteria were to be met, it would be expected that the students would get the same results in both modules. A plot of the two sets of results against each for a cohort should form a straight line with the equation $y = x$. The intercept would be 0 and the gradient would be 1. The correlation coefficient would also be 1.

Figure 10 - Grade comparison in an ideal world.



CHAPTER 6

In an ideal world, the scores for all modules would match, i.e. someone scoring 60% in one module would score 60% in the other module.

This thought experiment does not assume that students have:-

- the same starting points for each module
- the same levels of ability for each module.

In practice, since the initial four criteria are unlikely to be met, the pairs of scores will form a cloud of points and the linear correlation coefficient will be < 1 , whilst the slope and intercept are unlikely to be 0 and 1 respectively.

Each of the initial four assumptions will influence the amount of scatter generated. For example, assumption 1 assumes that all students find a module equally difficult. However, previous educational experiences will affect that greatly. Someone who has been taught to understand mathematical techniques is likely to find a higher-level mathematics module easier than someone who has been taught to rote learn techniques. This will cause scatter in the diagram.

Similar arguments can be made for each of the other three assumptions. Assumption 2 states that the work in both modules is to be assessed to the same standard – which is also unlikely since tutors vary in their expectations. Assumption 3 requires all modules to be taught to the same standard, which would require the ability to teach to be the same for both tutors. Finally, assumption 4 states that the students on each module would have equal access to learning materials and resources. This would require all presentations, tutorial materials and library resources to be of the same standard – another impossibility. The cumulative effect of these scatters could be quite unpredictable because one could counteract the other.

6.2 What might be expected?

If larger quantities of a new statistical test were calculated, then an empirical pattern might emerge and tend towards the population distribution. The question then becomes which type of distribution to expect.

The expectation of a teacher/lecturer on entering a classroom at the beginning of a new academic year is that the ability of the students will vary. Both Carroll (Carroll 1963) and Bloom argued that these abilities are likely to be normally distributed. To quote Bloom:

“Each teacher begins a new term (or module) with the expectation that about a third of [their] students will adequately learn what he has to teach. He expects about a third of his students to fail or to just “get by”. Finally, [she/he] expects another third to learn a good deal of what he has to teach, but not enough to be regarded as “good students”.”

(Source: (Bloom 1968))

Whilst a normal distribution might be expected for grades, using a D’Agostino-Pearson test for normality returned 140 runs with $p < 0.05$ (see appendix 9) out of 318. Runs with $n < 20$ were discounted because this test would not be valid in that range (D’Agostino & Pearson 1973). Therefore, 44% of the valid runs were deemed not to be normally distributed at a 5% significance level. This implies that the underlying distribution is unlikely to be normal.

If the distributions of the grades within runs are not normally distributed, then both the Student t-test and the Fisher F-test can be ruled out because they assume that the underlying data is normally distributed, at least for smaller samples – which is the case for the majority of modules. Any new statistical tests created will be compared against the non-parametric tests,

the Mann-Whitney U-test for location and the Levene and Brown-Forsythe tests for spread as explained in chapter 4.

6.2.1 Initial thoughts

Comparing the average score in one year to the average score in the next may not give a true reflection of how well that new learning intervention performed. Of the many possibilities for how things could turn out, the least likely to happen is that the new strategy is evaluated accurately because both the original and the trial group have exactly the same characteristics (academic background, motivation, etc.) and therefore the researcher is comparing like with like.

The problem of inter-year differences between successive years could be overcome by comparing within a year before comparing between years. If a module, say, B is the one on which the learning intervention is trialled, then its position relative to the other modules should alter. If the trial had a positive impact and module B had a mean score which was less than the other modules before the trial, it could now perform better and so its position relative to the other modules will increase. This positional score (PS) is the average score for those students who took module B when compared to their average scores in the other modules, i.e. the same students are compared in both modules. This principle is the subject of the next section.

6.3 The Principle of “Like-with-Like”

If all students take the same combination of modules then the following analysis is not relevant and can be ignored. This principle is only needed where students have a choice of subjects.

CHAPTER 6

Below are the results for a fictional programme which is being used to create a worked example. Each student has a unique identification number which is shown in the first column in the table below. In this year students could take just 1 module, 2 modules or all 3 modules. For the sake of this example, each student has taken either 2 or 3 modules, but they could have taken just 1 if they wished (as discussed in section 6.5). For example, student 001 took modules B and C, whilst student 005 took all three modules.

Table 5 - Fictional initial year results

Stud ID	Module A	Module B	Module C
001		23	56
002	65	66	
003	37	50	68
004		81	75
005	79	42	65
006	71	59	
007	59	44	
008	46	29	52
009		62	55
010	83		61
011	69		52
012	93	91	
n	9	10	8
Total	602	547	484
Mean	66.9	54.7	60.5

The mean scores for modules A, B and C were 66.9, 54.7 and 60.5 respectively. However, these could be biased by the students taking them. Therefore, the mean score for module A, for example, should not be compared with that of module B directly. Only those taking both modules A and B should be included. In the case above, it would only be students 002, 003, 005, 006, 007, 008 and 012 who are represented in both modules. This would give new means of 64.3 for module A and 54.4 for module B – but only for those students who took both modules.

CHAPTER 6

One could make the same comparisons for modules A and C as well as for B and C. In table 6 below, each pair of modules is shown. The first letter is the score that the student got in that module when the student also took the other module.

For example, student 001 got 23 in module B and 56 in module C. This module B grade is shown in column BC, whilst the module C grade is in column CB. The first letter represents the grade for that module. The order for the letters is not transitive therefore, BC₀₀₁ is not the same as CB₀₀₁. Student 001 scored 23 in module B and 56 in module C.

Table 6 - Fictional results analysis

Stud ID	AB	AC	BA	BC	CA	CB
001				23		56
002	65		66			
003	37	37	50	50	68	68
004				81		75
005	79	79	42	42	65	65
006	71		59			
007	59		44			
008	46	46	29	29	52	52
009				62		55
010		83			61	
011		69			52	
012	93		91			
n	7	5	7	6	5	6
Total	450	314	381	287	298	371
Mean	64.29	62.8	54.43	47.83	59.6	61.83
Module $\bar{x}$	66.9		54.7		60.5	

The mean score for module A for those students who also took module B is 64.29. The mean score for module A for those students who also took module C is 62.8. There seems to have been a slight difference between those picking module B and those opting for module C. In neither case is the module average of 66.9 appropriate because some students who did not take both modules are included.

The final row of table 6 shows the original module means as a comparison. As can be seen, none of the like-with-like means are the same as the module mean.

Whilst interesting, this does not help directly. The tactic used here is to find the position of each module relative to each of the others. For this, it is necessary to compare the module A grades with those for modules B and C, but only using those module A students who took the other modules. In this instance, we would not be interested in those who just took B and C.

6.4 The Positional Scores

The means of the six columns in the example above are; AB – 64.29, AC – 62.8, BA – 54.43, BC – 47.83, CA – 59.6 and CB – 61.83.

Obviously, the number of students taking a combination must be taken into account, because a module with more students should weigh more heavily on the statistic than one with only a few. The positional score of location for module A could therefore be calculated by using a weighted average:-

Equation 1 - Positional score calculation for module A

$$\begin{aligned}
 PS_L &= \frac{\sum(n\bar{x}_a)}{\sum(n\bar{x}_a)} \\
 &= \frac{64.29 \times 7 + 62.8 \times 5}{54.43 \times 7 + 59.6 \times 5} \\
 &= \frac{63.67}{58.58} \\
 &= \underline{1.1252}
 \end{aligned}$$

(where a is an individual score in module A and $\bar{a}$ is the corresponding score in the other module.)

However, the means are being multiplied by n and that just gives $\sum x$, therefore the calculation above could be rewritten as:

Equation 2 - Simplified equation for positional score

$$\begin{aligned}
 PS_{LocationA} &= \frac{\sum(x_a)}{\sum(x_{\bar{a}})} \\
 &= \frac{\sum AB + \sum AC}{\sum BA + \sum CA} \\
 &= \frac{450 + 314}{381 + 298} \\
 &= \frac{63.67}{58.58} \\
 &= \underline{1.1252}
 \end{aligned}$$

(where $\bar{a}$ denotes the score for those modules that are not from module a but were taken by students who did.)

The tactic is to place the total of the scores for any column starting with the module under consideration on the top of the fraction, in this case A. The denominator is the sum of the column which end with A. There is a symmetry to the second line in the equations above.

The positional location scores can be calculated in the same way for the other positional location score for module B:-

Equation 3 - The positional score for module B

$$PS_{LocationB} = \frac{\sum(x_b)}{\sum(x_{\bar{b}})}$$

$$\begin{aligned}
&= \frac{\sum BA + \sum BC}{\sum AB + \sum CB} \\
&= \frac{381 + 287}{450 + 371} \\
&= \underline{0.8136}
\end{aligned}$$

Positional location score for module C:-

Equation 4 - The positional score for module C

$$\begin{aligned}
PS_{LocationC} &= \frac{\sum(x_C)}{\sum(x_{\bar{C}})} \\
&= \frac{\sum CA + \sum CB}{\sum AC + \sum BC} \\
&= \frac{298 + 371}{314 + 287} \\
&= \underline{1.1131}
\end{aligned}$$

The relative positions of each of the three modules has been calculated using the principle of like-with-like. Any discrepancies caused by different types of students taking different modules has been eliminated.

6.5 The Learning Intervention Location Metric (LILM)

The positional scores are interesting because they show the relative position of each of the module. Module B appears to be lagging behind modules A and C. This could have many causes; tougher marking by the lecture responsible for module B, superb teaching by those tutoring modules A and C or something as mundane as the teaching for module B starting at 9am whilst A and C start at 1pm, for example. Measures could therefore be put in place to

CHAPTER 6

remedy whichever discrepancy is identified. This learning intervention would need to be monitored to assess its effectiveness since it may be that the perceived problem was not the cause. The problem then becomes one of measuring these relative changes between modules over successive years.

If a learning intervention took place in the following year (year 1) in module B one might get results such as:-

Table 7 - The results for the year with a learning intervention

Stud ID	Module A	Module B	Module C
101	87	85	82
102	56	57	
103	29	42	54
104	63	56	43
105	45	55	48
106	76	63	68
107	80		
108			32
109	55	62	59
110	59		61
111		63	43
112	67	54	
113		63	57
114	46		
115	43	58	
n	12	11	10
Total	706	658	547
Mean	58.8	59.8	54.7

Whilst the number of students has increased, this academic year seems to have performed less well. The raw mean scores for each module would be 58.8, 59.8 and 54.7 for modules A, B and C respectively. This decrease in grades could mask a successful experiment, making the effect look less than it really is. The next step is to calculate the positional location scores for each module.

CHAPTER 6

Students 107, 108 and 114 only took one module, therefore, they cannot be compared to other modules and their scores drop out.

Table 8 - Paired scores in a year where student only take one module

Stud ID	AB	AC	BA	BC	CA	CB
101	87	87	85	85	82	82
102	56		57			
103	29	29	42	42	54	54
104	63	63	56	56	43	43
105	45	45	55	55	48	48
106	76	76	63	63	68	68
107						
108						
109	55	55	62	62	59	59
110		59			61	
111				63		43
112	67		54			
113				63		57
114						
115	43		58			
n	9	7	9	8	7	8
Total	521	414	532	489	415	454
Mean	57.89	59.14	59.11	61.13	59.29	56.75

The positional location scores can be re-calculated for these modules for this academic intake.

The results are 0.9873, 1.0472 and 0.9623 for each of the modules A, B and C respectively.

It can now be seen that the positional scores for modules A and C have gone down whilst that for module B has gone up. Module B has performed better than would be expected relative to the other two modules.

The final step is to quantify these movements. If the year 0 positional location scores are subtracted from the year 1 scores, then the Learning Intervention Location Metric for each module is obtained for year 1. In these cases, the LILMs are:-

Table 9 - Calculating the LILM

$$LILM_A = 0.9873 - 1.1252 = \underline{-0.1379}$$

$$LILM_B = 1.0472 - 0.8136 = \underline{0.2335}$$

$$LILM_C = 0.9623 - 1.1131 = \underline{-0.1508}$$

The question of whether these changes are significant is discussed chapter 7 below, but it seems that the location of module B has changed for the better. If one module goes up, then the other modules have to go down because they all live in the same set. However, if one of the modules goes up, the downward trend for the others will be spread between them. This sharing out of the negative effect would mean that a significant change caused by a learning intervention in one module need not make the movements on the other modules significant.

6.6 Immediate potential problems with the LILM

The most obvious problem with this technique is that it only works in situations where numerical grades are awarded. The use of classifications such as pass, merit and fail could not be analysed using this statistic because they are not fine-grained enough. For example, one extra merit instead of a pass could skew the results.

This is a post hoc calculation. It assumes that the sample LILMs generated by the grade data is representative of all LILMs, which may not be the case. However, the sample does provide a starting point which can be validated, amended or disproved by using the technique in other institutions and for other subject areas.

6.7 Summary for the LILM

This technique to find the LILM for a set of modules for an academic year would be:-

1. Find the relative position of each module in the year before the learning intervention (the positional score for each module in that year).
2. Carry out a learning intervention.
3. Find the relative position of each module in the year of the learning intervention (the positional score for each module in the latest year).
4. Calculate the quantity of movement in the positional scores between the two years (the learning intervention location metric for each module in that year).

The results of this technique when applied to grades data used in this research are discussed in detail in chapter 7. However, it is also possible to use the like-with-like principle for assessing whether a learning intervention has also made a significant difference to the spread.

6.8 The Learning Intervention Spread Metrics

The LILM shows changes in the relative location of modules. It is directly comparable to the t-test and the U-test for parametric and non-parametric distributions respectively because they are also measures of location. However, one could have a situation where the location remains the same, but the spread of grades is altered. This can be analysed using a similar technique by creating the Learning Intervention Spread Metric (LISM) which is comparable to the standard deviation for normally distributed data and the Levene and Brown-Forsythe tests for non-parametric data as they all measure spread.

The same two sets of student grades for years 0 and 1 will be used to demonstrate how these metrics are calculated.

6.8.1 The measurement of spread

It is usual to measure spread as the squared variance from the mean. This eliminates positive and negative deviations cancelling each other out. For example, the variance of a set of data is given by:-

Equation 5 - Variance

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

As discussed in chapter 4, the Levene test uses the mean as the basis for calculating the W-statistic which can then in turn be used to estimate probability. However, the Brown-Forsythe test uses either the median or the trimmed-mean, say, (5th percentile + 95th percentile) ÷ 2. It would, therefore, be possible to mimic both these tests using a similar technique to the one described for the LILM by selecting the grades where the modules have students in common. However, the squared deviation from the mean would mimic the Levene test whilst the squared deviations from the median would shadow the Brown-Forsythe test.

The LISMs will test the level of significance of any changes in the spread of grades within the module being examined. This is useful when assessing whether a change in the learning environment has caused the range of grades to narrow or increase.

6.8.2 The Learning Intervention Spread Metric – Mean (LISM_{Mean})

The same two hypothetical sets of results as used in sections 6.3 and 6.5 will be used here.

A similar analysis can be undertaken for the spread. However, instead of taking the score, the squared deviance from the mean is used since this eliminates negative numbers and is a convention in statistics. This is the equivalent of the Levene test metric.

For example, student 001 had a score of 23 in the BC column in table 2. The mean score was 47.83 for that column. This would give a squared deviance of:-

$$(23 - 47.83)^2 = -24.83^2 = \underline{616.53}$$

The table below shows a slightly different figure because a spreadsheet was used to produce those figures and therefore there are fewer rounding differences.

Table 10 - The squared deviations from the mean for year 0.

Stud ID	AB	AC	BA	BC	CA	CB
001				616.69		34.03
002	0.51		133.90			
003	744.51	665.64	19.61	4.69	70.56	38.03
004				1100.03		173.36
005	216.51	262.44	154.47	34.03	29.16	10.03
006	45.08		20.90			
007	27.94		108.76			
008	334.37	282.24	646.61	354.69	57.76	96.69
009				200.69		46.69
010		408.04			1.96	
011		38.44			57.76	
012	824.51		1337.47			
Total	2193.43	1656.80	2421.71	2310.83	217.20	398.83

Table 11 - The squared deviations from the mean for year 1.

Stud ID	AB	AC	BA	BC	CA	CB
101	847.46	776.02	670.23	570.02	515.94	637.56
102	3.57		4.46			
103	834.57	908.59	292.79	365.77	27.94	7.56
104	26.12	14.88	9.68	26.27	265.22	189.06
105	166.12	200.02	16.90	37.52	127.37	76.56
106	328.01	284.16	15.12	3.52	75.94	126.56
107						
108						
109	8.35	17.16	8.35	0.77	0.08	5.06
110		0.02			2.94	
111				3.52		189.06
112	83.01		26.12			
113				3.52		0.06
114						
115	221.68		1.23			
Total	2518.89	2200.86	1044.89	1010.88	1015.43	1231.5

The same technique can now be used to create the positional scores for the mean spread metric.

The positional spread score for module A for year 0 would be given by:-

Equation 6 - The formula for the positional scores for the LISM_{Mean}

$$\begin{aligned}
 PS_{SpreadA} &= \sqrt{\frac{\sum((x_A - \bar{x}_A)^2)}{\sum(x_A - \bar{x}_A)^2}} \\
 &= \sqrt{\frac{\sum AB + \sum AC}{\sum BA + \sum CA}} \\
 &= \sqrt{\frac{2193.43 + 1656.8}{2421.71 + 217.2}} \\
 &= \sqrt{\frac{3850.23}{2638.91}} \\
 &= \underline{1.2079}
 \end{aligned}$$

The other positional spread scores for modules B and C would be 1.3512 and 0.394 respectively.

The equivalent values for year 1 are 1.5135, 0.7404 and 0.8364 for each of the three modules.

The movement in spread between year 0 and year 1 can then be calculated, the Learning Intervention Spread Metric, using the mean ($LISM_{Mean}$) as:-

$$LISM_{MeanA} = 1.5135 - 1.2079 = \underline{0.3056}$$

$$LISM_{MeanB} = 0.7404 - 1.3512 = \underline{-0.6108}$$

$$LISM_{MeanC} = 0.8364 - 0.394 = \underline{0.4424}$$

It would therefore seem that the hypothetical learning intervention resulted in a decreased spread for module B relative to the other two modules. Again, the question is whether the change in spread for module B is significant or not, and that is also addressed in chapter 7.

The results can be summarised as:-

Table 12 – $LISM_{Mean}$ results

	Module A	Module B	Module C
Pos. Scr. Year 0	1.2079	1.3512	0.3940
Pos. Scr. Year 1	1.5135	0.7404	0.8364
$LISM_{Mean}$	0.3056	-0.6108	0.4424

6.8.3 The Learning Intervention Spread Metric – Median ($LISM_{Median}$)

The same two hypothetical sets of results as used in sections 6.3 and 6.5 will be used here.

CHAPTER 6

The same calculations can be performed using the median instead of the mean, making it the counterpart to the Brown-Forsythe test metric. The calculations for both years for the dummy data are shown in the tables below:

Table 13 - The squared deviations from the median for year 0.

Stud ID	AB	AC	BA	BC	CA	CB
001				529.00		20.25
002	0.00		256.00			
003	784.00	1024.00	0.00	16.00	49.00	56.25
004				1225.00		210.25
005	196.00	100.00	64.00	16.00	16.00	20.25
006	36.00		81.00			
007	36.00		36.00			
008	361.00	529.00	441.00	289.00	81.00	72.25
009				256.00		30.25
010		196.00			0.00	
011		0.00			81.00	
012	784.00		1681.00			
Total	2197.0	1849.0	2559.0	2331.0	227.0	409.5

Table 14 - The squared deviations from the median for year 1.

Stud ID	AB	AC	BA	BC	CA	CB
101	961.00	784.00	784.00	506.25	529.00	702.25
102	0.00		0.00			
103	729.00	900.00	225.00	420.25	25.00	2.25
104	49.00	16.00	1.00	42.25	256.00	156.25
105	121.00	196.00	4.00	56.25	121.00	56.25
106	400.00	289.00	36.00	0.25	81.00	156.25
107						
108						
109	1.00	16.00	25.00	0.25	0.00	12.25
110		0.00			4.00	
111				0.25		156.25
112	121.00		9.00			
113				0.25		2.25
114						
115	169.00		1.00			
Total	2551.00	2201.00	1085.00	1026.00	1016.00	1244.00

Using the same technique to calculate the positional scores and the learning intervention metric

– median values gives:-

Table 15 - LISM_{Median} results

	Module A	Module B	Module C
Pos. Scr. Year 0	1.2051	1.3697	0.3902
Pos. Scr. Year 1	1.5039	0.7458	0.8369
LISM_{Median}	0.2988	-0.6239	0.4466

These are virtually the same results as when using the mean and is included here to show that this is the case. The LISM_{Median} is therefore being retired for the remainder of this chapter.

6.9 Using Both the LILM and $LISM_{Mean}$ to Assess Impact

It would therefore seem that the spread in module B has also decreased relative to the other modules. Again, the increase in learning intervention spread values for modules A and C may not show an increase in spread in those modules because they are being measured relative to module B.

The next question is whether the changes in relative spread help the students? The table below shows the possibilities:-

Table 16 - LILM vs $LISM_{Mean}$ Case Table

	Positive LILM	Stable LILM	Negative LILM
Positive $LISM_{Mean}$	Case 1	Case 2	Case 3
Stable $LISM_{Mean}$	Case 4	Case 5	Case 6
Negative $LISM_{Mean}$	Case 7	Case 8	Case 9

The terms “positive”, “stable” and “negative” do not have an exact meaning here. Their meaning would depend upon the value the analyst places on the learning intervention. For example, a learning intervention such as employing a new teacher or moving the start time of classes forward an hour are not trivial. The term “stable” might be drawn more narrowly than if, say, a class was moved to a different room. The final step in the analysis would be to examine each of the nine cases.

6.10 Combining the LILM and $LISM_{Mean}$

The value of a learning intervention might depend on both the LILM, how the learning intervention moved the scores, as well as its $LISM_{Mean}$, how it changed the spread of scores. Each case can now be described.

6.10.1 Positive LILM and Positive LISM_{Mean} (Case 1)

Here both the location and spread have increased significantly. This is an ambiguous case. It could be good news if the rise in grades offsets the number of students whose grades are towards the bottom of the group. Alternatively, it could be bad news if some at the lower end are pushed below the pass mark. It would therefore be necessary to examine the effect on students before a decision is made on the need for remedial action.

6.10.2 Stable LILM and Positive LISM_{Mean} (Case 2)

Here the grades have stayed stable but the spread has increased. This would be fine for those at the top, their grades are increasing. However, a larger proportion could fail if those at the bottom are being pushed below the pass mark. It may be that the spread of grades in the original year was too small (too little differentiation) and that the grading has now normalised. The detail would have to be examined.

6.10.3 Negative LILM and Positive LISM_{Mean} (Case 3)

In case 3 the grades have decreased significantly and spread has increased significantly. This would indicate a greater number of students who fail both because the relative grades have gone down and the increased spread causes more students to move into the lower regions. This could be the worst possible situation.

6.10.4 Positive LILM and Stable LISM_{Mean} (Case 4)

The location of the grades has increased whilst the spread has remained the same. This is a good situation. The number of those achieving lower grades will have gone down relatively and the grades in general, including those at the top, will have increased. This is a good case from the point of view of the students.

6.10.5 Stable LILM and Stable LISM_{Mean} (Case 5)

Both the location and spread have stayed stable. This shows that the learning intervention had no significant effect. It neither increased the average score appreciably nor saved some from failing. Changes here are tinkering rather than radical changes. This is the default situation where the same people are teaching using the same materials year on year. It is the case that one would expect to happen most often.

6.10.6 Negative LILM and Stable LISM_{Mean} (Case 6)

The grades have decreased significantly whilst the spread remained stable. This is unlikely to be good for the pass rate. The whole group seems to have drifted downwards relative to other modules in the same set and there is no compensation of those at the top increasing achievement. There may be a reason for this, such as someone new to teaching taking over a module or, say, a significant level of sickness. These reasons may not re-occur, but the source of the problem may need to be found.

6.10.7 Positive LILM and Negative LISM_{Mean} (Case 7)

Both the location of the grades has gone up significantly and spread has decreased significantly. This is the ideal situation, if the grading system allows for differentiation. Those at the bottom are likely to have been shifted up significantly whilst those towards the top have gone up even more.

6.10.8 Stable LILM and Negative LISM_{Mean} (Case 8)

Here the location has remained constant and the spread has decreased significantly. This may have moved some out of the danger zone, but the cost is to those at the top who may have moved down. It is therefore possible to argue this case both ways. This could happen if the method of assessment has been altered. Alternatively, it may be that new materials have become available which have brought down the spread.

6.10.9 Negative LILM and Negative LISM_{Mean} (Case 9)

Here both the location and spread have decreased significantly. This is an ambiguous situation and the detail would need to be examined. It could be that someone new has taken over and that they are marking more strictly. It could be that a learning intervention has back-fired – unfortunate, but a lesson has been learned.

The clear message here is that in most cases further analysis would be needed. The module monitoring reports giving the overview of how the module ran would be a rich source of research material if it was available.

6.11 Conclusion

This chapter explains the thinking behind the statistical tests which might be used in an educational experiment. It used the principle of like-with-like, i.e. modules are only compared directly for those students that they have in common whilst year groups are only compared where they have modules in common.

The practice of implementing these tests is described in the following chapter. There the LIIMs ($LILM$, $LISM_{Mean}$ and $LISM_{Median}$) are calculated for each module in the data set drawn down from the university system and an analysis of the outcomes is carried out.

Chapter 7 LIIM Results and Discussion

7.1 Introduction

The previous chapter looked at the theory behind three possible statistical tests to measure the effectiveness of a learning intervention. Together these three tests are called learning intervention impact metrics (LIIM). One metric, the learning intervention location metric (LILM) is intended to test whether a learning intervention has resulted in a movement of the scores of a cohort.

There are then two other metrics which look at changes in the spread of scores between the two cohorts involved in the learning intervention; the learning intervention spread metric using the mean (LISM_{Mean}) and the learning intervention spread metric using the median (LISM_{Median}).

This chapter looks at their implementation and analyses their performance when applied to the same data set used when examining the five standard tests discussed in chapter 4. The implementation and analysis will answer the research question for this data set.

7.2 Results from Historical Data

7.2.1 Introduction

The original data set was described in detail in chapter 3. This data set covers 48 modules in the 8 academic years from 2011/12 to 2018/19. The clean data set being used contains 28,610 grades. It is this data set that is being used for the calculations of the LIIMs below.

7.2.2 The Learning Intervention Location Metric (LILM)

The method to calculate the LILMs was described in chapter 6. The stages were:-

1. Group the grades by module run (all those taking the same module in the same academic year).
2. Group the modules by set (all those modules at the same level in the same academic year).
3. Calculate the positional score for each module in each set.
4. Subtract the positional score for each module in the previous academic year from the one in the year under consideration to get the learning intervention location metric (LILM).

Appendix 20 of positional scores for the LILM calculations for each module in each academic year can be found in table 1, whilst the LILMs are in appendix 21.

The mean of the LILMs would be expected to be approximately 0.0 since the positive and negative values should cancel out.

It might be expected that the distributions of the LILMs would form a normal distribution because, 1) they are the result of human activity and 2) the central limit theorem suggests that they might. The graph of the distribution of the LILMs also resembles an imprecise normal distribution as shown in the histogram below:-

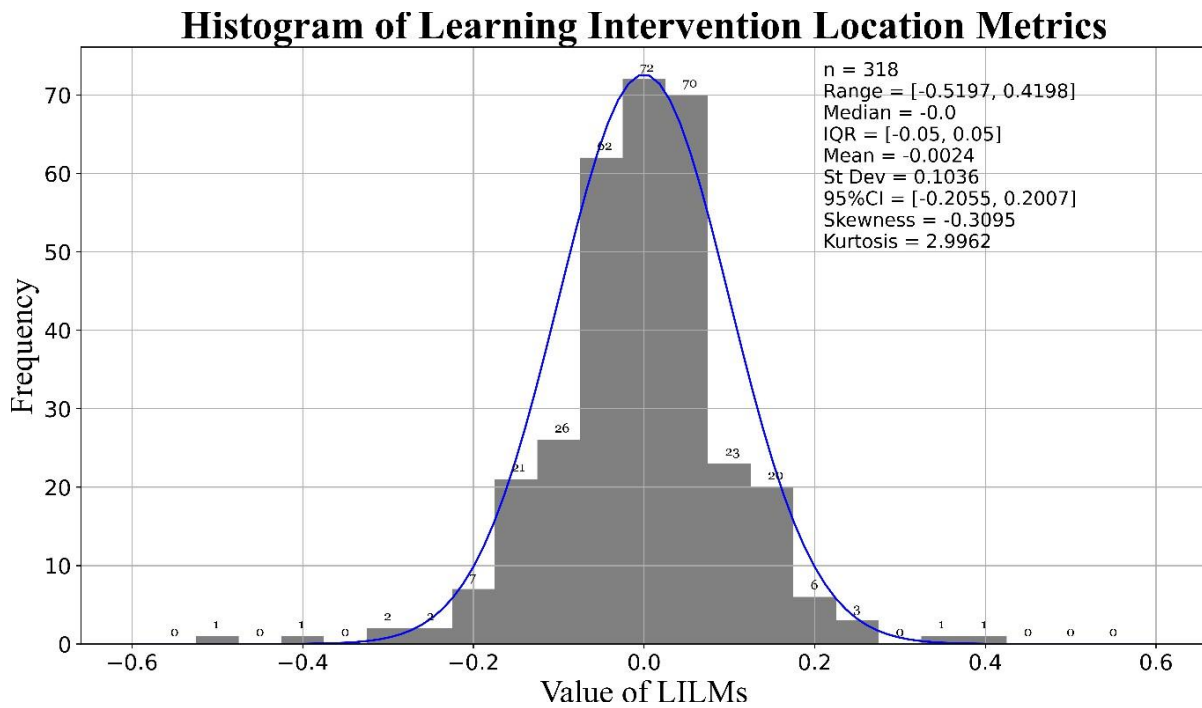


Figure 11 - Histogram of the Learning Intervention Location Metrics (LILM)

However, the D'Agostino-Pearson test (D'Agostino & Pearson 1973) strongly suggests that this is not the case. The graph above shows the descriptive statistics of this LILM distribution. The probability that the distribution is normal was calculated as 0.0% - the actual value is 1.3802×10^{-7} as calculated by the *SciPy* package in Python. On the positive side, the kurtosis is calculated as 2.9962 which is close to the ideal value of 3.0 for the normal distribution.

7.2.3 Bootstrapping – Confidence Intervals for the LIIMs

Bootstrapping (Efron 1992) is a statistical technique adapted from the Jackknife technique by Bradley Efron in 1992. It is used to estimate the mean and standard deviation of a large sample by randomly sampling the dataset to get a metasample of the same size as the dataset. The mean and the standard deviation of the random sample is calculated and the process repeated a

large number of times (here, 10,000 samples were taken each time). The mean and standard deviations of these samples are then calculated to get a better approximation of the population mean and standard deviation.

7.2.4 Confidence intervals for the LILM

The 10,000 samples of 318 randomly selected LILMs was taken and their descriptive statistics were calculated for both the the mean and the standard deviation:-

Table 17 - Bootstrap statistics for the LILMs

Statistic	The Means	The St. Dev.s
n	10,000	10,000
Min	-0.0262	0.0825
Max	0.0189	0.1302
Range	0.0451	0.0477
Mean	-0.0023	0.1033
Q1	-0.0062	0.0988
Median	-0.0023	0.1031
Q3	0.0016	0.1075
IQR	0.0078	0.0087
St. Dev.	0.0058	0.0065
67% CI ($\pm 0.431\sigma$) (Hawthorne)	[-0.0048, 0.0002]	[0.1005, 0.1061]
95% CI ($\pm 1.96\sigma$)	[-0.0137, 0.0091]	[0.0905, 0.116]
99% CI ($\pm 2.576\sigma$)	[-0.0173, 0.0127]	[0.0865, 0.12]
Skewness	-0.0467	0.1825
Kurtosis	0.0372	0.042
D'Agostino-Pearson p of normality	0.1191	0.0

The graphs below show the distributions for each of these samples:-

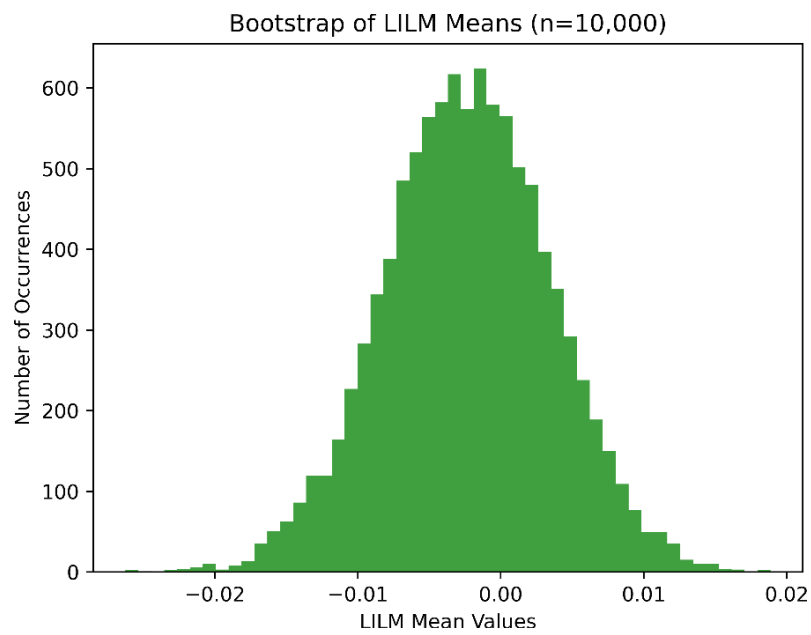


Figure 12 - The distribution of the LILM means

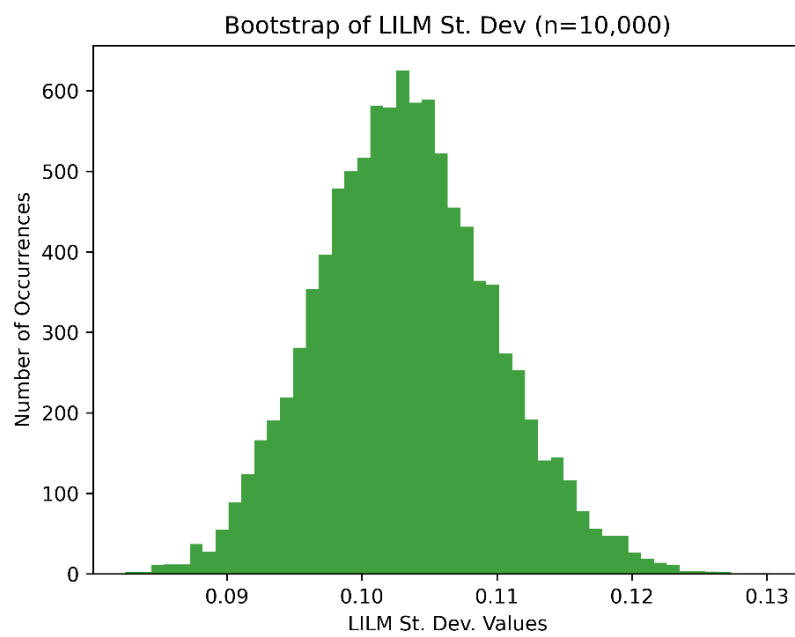


Figure 13 - The distribution of the LILM standard deviations

It was expected that the mean would be 0.0. The actual value of -0.0023 (to 4 d.p.) is 0.023σ from 0.0. It would therefore seem safe to assume a mean of 0.0 until further studies have been completed.

The mean of the standard deviations is 0.0988 which, given the size of its standard deviation (0.0065), seems a reasonable approximation.

The Hawthorne Effect (Mayo 2009; Landsberger 1958) has been estimated by Hattie (Hattie 2009) as being around 0.4σ (31%). Rounding this a little beyond that to $\frac{1}{3}$, the 33%CI (0.431σ) for the LILMs would be [-0.0426, 0.0426] and the 95%CI would be [-0.1936, 0.1936]. Given that these are approximations, using ± 0.04 and ± 0.2 respectively would not alter the outcomes and make the results easier to understand. Were an educational experiment to be conducted then these values could be used as a starting point to gauge its effectiveness at moving the mean score until more data comes in.

7.2.5 The Learning Intervention Spread Metric – Mean (LISM_{Mean})

The 10,000 samples of 318 randomly selected LISM_{Mean}s was taken and their descriptive statistics were calculated for both the the mean and the standard deviation:-

Table 18 - Bootstrap statistics for the LISM_{Means}

Statistic	The Means	The St. Dev.s
n	10,000	10,000
Min	-0.07	0.2286
Max	0.0715	0.4495
Range	0.1415	0.2209
Mean	0.0009	0.3282
Q1	-0.0115	0.3064
Median	0.001	0.3274
Q3	0.0135	0.3493
IQR	0.025	0.0429

St. Dev.	0.0187	0.0316
67% CI ($\pm 0.431\sigma$) (Hawthorne)	[-0.0071, 0.009]	[0.3146, 0.3419]
95% CI ($\pm 1.96\sigma$)	[-0.0357, 0.0376]	[0.2664, 0.3901]
99% CI ($\pm 2.576\sigma$)	[-0.0473, 0.0491]	[0.2469, 0.4096]
Skewness	0.0076	0.1562
Kurtosis	0.0362	-0.1575
D'Agostino-Pearson p of normality	0.7123	0.0

The graphs below show the distributions for each of these samples:-

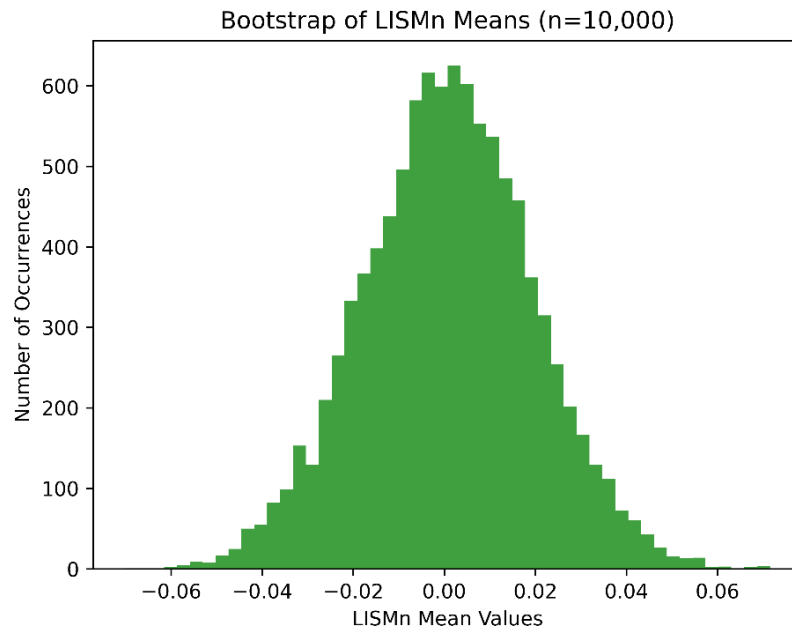


Figure 14 - The distribution of the LISM_{MeanS}

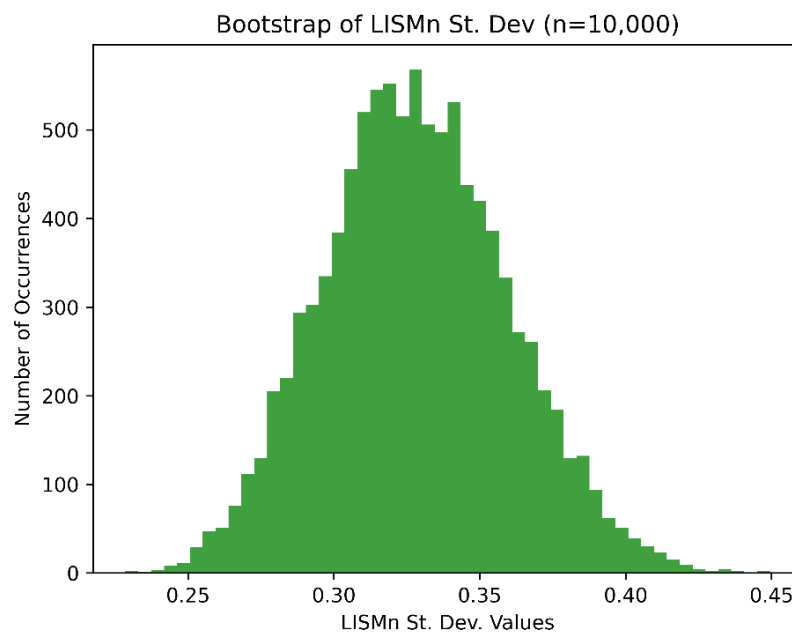


Figure 15 - The distribution of the LISM_{MeanS} standard deviations

CHAPTER 7

It was expected that the mean would be 0.0. The actual value of 0.0009 (to 4 d.p.) is 0.0027σ from 0.0. It would therefore seem safe to assume a mean of 0.0 until further studies have been completed.

The mean of the standard deviations is 0.3282 which, given the size of its standard deviation (0.0316), seems a reasonable approximation.

As stated above, the 33%CI (0.431σ) for the $LISM_{Means}$ would be $[-0.1414, 0.1414]$ and the 95%CI would be $[-0.6433, 0.6433]$. Given that these are approximations, using ± 0.14 and ± 0.64 respectively would not alter the outcomes and make the results easier to understand. These values could be used to gauge its effectiveness at moving the spread of the scores until more data comes in.

7.2.6 The Learning Intervention Spread Metric – Median (LISM_{Median})

The 10,000 samples of 318 randomly selected LISM_{Median}s was taken and their descriptive statistics were calculated for both the the mean and the standard deviation:-

Table 19 - Bootstrap statistics for the LISM_{Median}s

Statistic	The Means	The St. Dev.s
n	10,000	10,000
Min	-0.072	0.2356
Max	0.0705	0.4788
Range	0.1425	0.2432
Mean	0.0007	0.3346
Q1	-0.012	0.3128
Median	0.0005	0.3337
Q3	0.0135	0.3555
IQR	0.0255	0.0427
St. Dev.	0.0188	0.0311
67% CI ($\pm 0.431\sigma$) (Hawthorne)	[-0.0074, 0.0088]	[0.3212, 0.348]
95% CI ($\pm 1.96\sigma$)	[-0.0362, 0.0376]	[0.2737, 0.3955]
99% CI ($\pm 2.576\sigma$)	[-0.0478, 0.0492]	[0.2546, 0.4146]
Skewness	0.0299	0.1315
Kurtosis	-0.0106	-0.0795
D'Agostino-Pearson p of normality	0.4664	0.0

The graphs below show the distributions for each of these samples:-

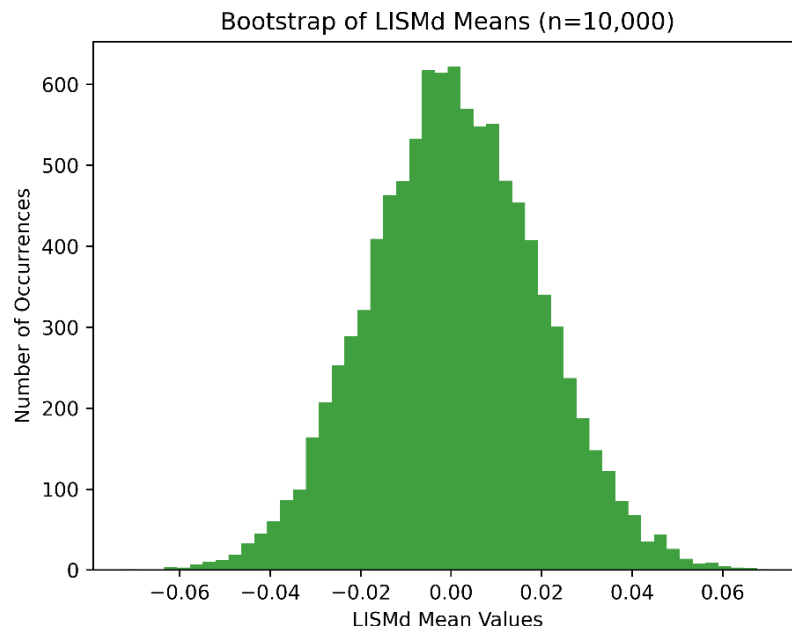


Figure 16 - The distribution of the LISMd_{Medians}

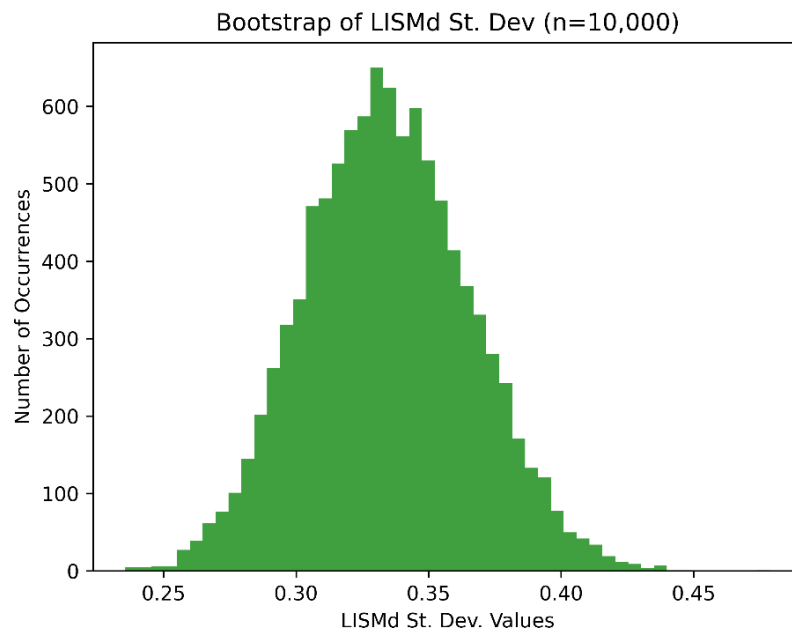


Figure 17 - The distribution of the LISMd_{Medians} standard deviations

It was expected that the mean would be 0.0. The actual value is 0.0007 (to 4 d.p.) is 0.0021σ from 0.0. It would therefore seem safe to assume a mean of 0.0 until further studies have been completed.

The mean of the standard deviations is 0.3346 which, given the size of its standard deviation (0.031), seems a reasonable approximation.

Again, as above, the 33%CI (0.431σ) for the $LISM_{MeanS}$ would be $[-0.1442, 0.1442]$ and the 95%CI would be $[-0.6558, 0.6558]$. Given that these are approximations, using ± 0.14 and ± 0.66 respectively would not alter the outcomes and make the results easier to understand. These values could therefore be used as a starting point to gauge the effectiveness of the experiment at moving the spread of the scores.

7.2.7 Overview of the three tests

The results from the three tests are summarised below with the figures in round brackets being the number of cases meeting the 5% criteria:-

Table 20 - Summary of the LIIM Results

	LILM	$LISM_{Mean}$	$LISM_{Median}$
n	318	318	318
Actual Mean	-0.0024	0.0011	0.0006
Actual St. Dev.	0.1036	0.3305	0.3368
Assumed Mean	0.0	0.0	0.0
Assumed St. Dev.	0.0988	0.3282	0.3346
Assumed 67% CI	$[-0.04, 0.04]$ (194 = 61%)	$[-0.14, 0.14]$ (170 = 53.5%)	$[-0.14, 0.14]$ (179 = 56.3%)
Assumed 95% CI	$[-0.2, 0.2]$ (16 = 5%)	$[-0.64, 0.64]$ (15 = 4.7%)	$[-0.66, 0.66]$ (15 = 4.7%)

Note: The figures in round brackets below the confidence intervals show the number of observations outside those intervals and the percentage they are of the 318 observations.

All the statistics stayed relatively close to the original values. The means stayed near 0 and the standard deviations stayed at about 0.1, 0.33 and 0.33 for the LILMs, $LISM_{MeanS}$ and $LISM_{MedianS}$ respectively.

The Hawthorne Effect values, where students know that an experiment is taking place and perform differently, are quite small. This would make detecting changes easier.

7.2.8 The case analyses

It is now possible to get a view of how the location and spread tests might interact. First, the interaction between the LILM and $LISM_{Mean}$.

A positive LILM is one that is over 0.431σ above the mean. The mean is 0.0 and the LILM standard deviation is 0.0988 therefore 0.431σ gives a LILM of 0.04. Anything above that value is being considered a positive LILM. A stable LILM would be one between -0.04 and 0.04 whilst a negative LILM would be one less than -0.04, i.e. limits which divide the LILMs in approximate thirds.

A similar classification can be made for the $LISM_{Mean}$ values. These can be divided into three similarly sized groups by going 0.431σ evenly sized groups ± 0.14 . In chapter 6 the interaction matrix was defined as:-

Table 21 - Case Matrix from Chapter 6

	Positive LILM	Stable LILM	Negative LILM
Positive $LISM_{Mean}$	Case 1	Case 2	Case 3
Stable $LISM_{Mean}$	Case 4	Case 5	Case 6
Negative $LISM_{Mean}$	Case 7	Case 8	Case 9

Using these limits the count in each case is:-

Table 22 - Case Matrix from using LILM (+/- 0.04) with $LISM_{Mean}$ (+/- 0.14)

	Positive LILM	Stable LILM	Negative LILM
Positive $LISM_{Mean}$	28	28	28

Stable LISM_{Mean}	40	65	43
Negative LISM_{Mean}	30	31	25

The equivalent table using the LISM_{Median} values is:-

Table 23 - Case Matrix from using LILM (+/- 0.04) with LISM_{Median} (+/- 0.14)

	Positive LILM	Stable LILM	Negative LILM
Positive LISM_{Median}	28	30	29
Stable LISM_{Median}	39	61	39
Negative LISM_{Median}	31	33	28

Both tables have similar numbers in each cell.

As one would expect, the most populous cell is case 5 – the one in which there is no appreciable movement in either the location or spread of the scores. However, this only accounts for just over 19% of the runs. The “best case scenario”, case 7, where the location improves and the spread goes down, accounts for approximately 9% of the runs. The “worst case scenario”, case 3, where the average grade goes down and the spread increases, accounted for around 9% of the runs. The remaining cases, over 65% of the run population, had scores which moved either the spread or the location of the scores on that run.

7.2.9 Summary

It seems that for this data set there is little difference between using LISM_{Mean} and LISM_{Median} values using any of the measures discussed in this chapter. That may change when more data becomes available. For this reason, the LISM_{Median} is being kept until the situation becomes clearer. Were these results to be replicated then there would be little to be gained by using the LISM_{Median} technique when the LISM_{Mean} is slightly more accurate and has a distribution that

is more like the normal distribution. The tests could then be reduced just LILM, to assess movements in location, and the $LISM_{Mean}$ to assess the effect of any learning intervention on the spread of scores. The $LISM_{Median}$ is therefore being dropped from the follow discussions.

7.3 Alternative uses

To examine possible ramifications of the tests, it is being assumed that the general thrust of the results would be replicated by other studies, even if the values of important and significant LILMs and $LISM_{Mean}$ values are amended later.

The discussion so far has looked at ways that researchers can gauge the effectiveness of an experiment which can be analysed using its positional score, LILM and $LISM_{Mean}$. However, these tests could be used more widely to monitor the performance of modules on a year-by-year basis using the cases discussed above as part of the annual institution monitoring cycle. They could be used to assess changes in staffing, rooming, times of tuition, etc. all of which would provide a greater insight into how students achieve best.

7.4 Applying the Results

Having established that these tests show some promise, this section takes a retrospective look at the module on which the experiment was run. It also examines other interesting LILMs, those whose values put themselves outside the 95% confidence interval, to see if anything can be learned from the annual monitoring reports where they exist.

7.4.1 Examining the effectiveness of the experimental site

The module used in the experiment was module 5~04 in the 2018/19 academic year. The results for that year and the one before are summarised in the table below:-

Table 24 - Experiment (Module 5~04, 2018/19) – Descriptive Statistics

	2017/18	2018/19
n	108	100
Mean	55.7407	59.64
Standard Deviation	17.8804	20.6459
Coefficient of Skewness	0.0488	-0.0502
Kurtosis	-0.6124	-1.1203
p of Normality	27.08%	0.00%
Positional Score (LILM)	0.9725	0.949
LILM	-0.1906	-0.0235
Positional Score _{Mean}	1.1962	1.5463
LISM _{Mean}	-0.286	0.3501

The following points are of interest:-

- Both runs had approximately the same number of active students.
- Both the mean and standard deviation went up in the year of the experiment.
- The kurtosis shows that both are platykurtic, i.e. showing thinner tails than would be expected if this were a normal distribution.
- Whilst the run scores for 2017/18 appear normally distributed, those for 2018/19 do not.
- Given that the mean score has increased, a LILM of over zero would have been expected. The LILM of -0.0235 can only happen if the mean scores for the other modules in that set have increased even more.
- The increase in spread shown by the LISM_{Mean} is less of a surprise given that the standard deviation of the scores has increased between the two years.

CHAPTER 7

The results place the module in the case 2 category, i.e. no appreciable change in location, but the spread increased. The spread increased in between the Hawthorne Effect and the 95% confidence interval, enough to be noticeable but not enough to be significant. A possible reason for this is that a proportion of the student were surface learners and felt that extra materials were an imposition.

These figures send an interesting message. It is not clear that the experimental site delivered any appreciable benefit to the students. In fact, a higher increase in mean grade would have been expected because this year group performed better in other modules.

Further insight can be gained by looking at the site usage statistics taken from the website. Of the 100 students, 96 enrolled with the site. Of these, 24 had fewer than four logins which means that they registered but did not use the site for study. 34 students used the site 12 times or more, which represents more than once per academic week, but only 7 averaged more than two visits per week.

If the 96 registered students had each completed (passed the quiz with 100%) all the tasks, then there would have been 3,552 completions. There were only 449 completions, 12.6% of the maximum possible. 46 students did not complete any tasks. At the other end of the scale, 4 students completed all the tasks.

There are four main reasons why this cohort were under-achieving relative to the other modules that they took:-

1. The terms of the ethical approval meant that the students had to be told in advance that this was an experiment and that participation was voluntary. The student feedback indicated that many saw this as extra work which would not benefit them directly.

Therefore, there was some reluctance to participate. Since this was their major source of extra-classroom materials, this had a detrimental effect on scores.

2. The experiment started with some items in a task being hidden, as explained in the methodology. Some were confused by this whilst others thought that the site was not working correctly even though this had been explained to them. This side of the experiment was stopped after two weeks for this reason and all tasks had all items after that. Whilst unwelcome, this did not appear to have a long term effect on the grades.
3. The site was not integrated into the teaching. The ethical approval conditions meant that using the site was optional. It was thought that making it part of the teaching, for example using the site instead of PowerPoint slides, might infringe that condition. This allowed students to feel that the site was an extra – something that they did not need in their crowded schedule of academic work. This was stated explicitly by some in the student feedback. This led to students finding the subject more difficult than necessary because they were not exposed to these materials and therefore the grades suffered. This was a unfortunate since those who did use the site gave positive feedback. Therefore, this appears to have had a detrimental effect on outcomes. However, those that did use the site appeared to have benefitted. The combination of these two effects increased the spread of results.
4. Finally, the tasks followed the scheme of work rather than the marking scheme. In hindsight, this was a mistake as there was no obvious link between the materials and their grades. Whilst this is unlikely to have effected the grades directly, a further opportunity for increasing them was missed.

When using the site in future, the site materials would be integrated more into the teaching. The progress of the students could be discussed during tutorial sessions with each student to

encouraging greater participation. Finally, the sequence of the tasks would be tied to the marking scheme so that students could see how their effort related to their grade directly.

7.4.2 Examining other anomalies discovered by the LIIMs

The first example used below shows how these techniques can be used to analyse a specific learning intervention. As mentioned in the methodology, one compulsory module, 4~01, was split into two for the 2017/18. It was felt that results might improve if students who underperformed in the first semester could catch up if the module was were split. The statistical data for the last three academic years were:-

Table 25 - Module 4~01 (2017/18) – Descriptive Statistics

	2016/17	2017/18
n	265	201
Mean	61.1585	63.7761
Standard Deviation	19.2373	15.1369
Coef. of Skewness	-0.4196	-0.1706
Kurtosis	-0.1795	-0.9178
p of Normality	2.01%	0.00%
Pos. Score (LILM)	1.0601	1.0076
LILM	-0.0193	-0.0525
Pos. Score _{Mean}	1.0138	0.8887
LISM _{Mean}	-0.0380	-0.1251

In the year the experiment started, 2017/18, the LILM and LISM_{Mean} scores push 4~01 just into case 8 by 0.0051. The slightly decreased spread seems to have been caused by those at the upper end improving their grades in return for no noticeable effect on those at the lower end.

Again, the mean score increased, but the LILM went down, implying that the other modules went up more. As well as the mean score going up, the standard deviation of the scores went down. It seems that students would be taken out of the fail/resit group and this is backed up by the LISM_{Mean} also going down. Whilst not bringing gains that were significant at the 5% level, it does appear that the change brought advantages to the students.

These techniques can also be used as research into those issues which affect students significantly.

7.4.3 The 16 other significant results

Appendix 21 shows the LILM calculations with those that fall outside the 95% CI highlighted when using the normal distribution. Given that there were 318 LILMs, one would expect that 5% of these (16) would be significant. It was argued earlier that the 95% confidence interval

limits should be relaxed to ± 0.2 . If these limits are used, then 16 LILMs were found to be significant. These are the 16 being examined here.

The module monitoring reports, where available, were used to see which events might have caused these LILM outliers. The reasons are shown in the table below were taken from the module monitoring reports (CMR) or by seeing whether there were staffing changes where this was shown.

7.4.4 The reasons behind the LILMs significant at the 5% level

Table 26 - Runs whose LILM was significant at the 5% level.

Row	Module	Year	Reason Given (if any)	t-Test p
1	4.05	2013/14	(LILM: -0.3208) Neither CMR nor staffing shown.	0.0%
2	4.05	2014/15	(LILM: 0.2384) A new format for the end of module examination was trialled.	0.0%
3	4.05	2017/18	(LILM: 0.4198) The lecturer was new the year before and made significant improvements in their second year.	0.0%
4	5.09	2017/18	(LILM: 0.2067) The lecturer was now in their second year and added extra materials as well as recording all lectures and making them available.	10.33%
5	6.02	2014/15	(LILM: -0.2532) The lecturer who took over this module used the same materials as the predecessor which led to some “interesting” situations.	15.3%
6	6.03	2012/13	(LILM: 0.2417) The lecturer introduced flipped classroom learning to this module in this academic year.	0.03%
7	6.04	2012/13	(LILM: -0.2035) Neither CMR nor staffing shown.	55.27%
8	6.05	2018/19	(LILM: -0.2352) A new member of staff took over and was finding their feet. Changes are planned for the next academic year.	20.87%
9	6.09	2014/15	(LILM: 0.2345) The same lecturer used the same materials as the year before. Whilst the lecturer was disappointed in the overall result (a drop of 20% in the mean score), it seems that other modules fared worse.	2.44%
10	6.12	2012/13	(LILM: -0.5197) No CMR and no change in lecturer.	100.0%
11	6.13	13/14	(LILM: -0.3781) No CMR, but the lecturer was new to the module.	0.0%
12	6.13	14/15	(LILM: 0.3402) The CMR implies that the module ran in much the same way as the year before, but the materials were updated.	0.26%
13	7.03	2013/14	(LILM: 0.2035) The lecturer changed.	21.43%
14	7.03	2016/17	(LILM: 0.2007) The lecturer took over who introduced a more transparent marking scheme and made audio recordings of the lectures available.	1.09%
15	7.03	2017/18	(LILM: -0.3121) A temporary lecturer took over as cover whilst their predecessor took a year out. The same materials were used.	0.02%
16	7.14	2018/19	(LILM: -0.2032) The same lecturer taught in both years and used similar materials. The feedback was equally positive, but the motivation of the students appeared to be less according to the CMR.	0.0%

There are six general categories indicated in the table above:-

- (1) No information (3) – rows 1, 7 and 10
- (2) No changes made (2) – rows 9 and 16
- (3) Lecturer change – positive (2) – rows 3 and 13
- (4) Lecturer change – negative (4) – rows 5, 8, 11 and 15
- (5) Learning intervention (4) – rows 4, 6, 12 and 14
- (6) Assessment intervention (1) – row 2

7.4.5 Conclusions

It seems that changes in the teaching staff are the major cause of differences in results between years, accounting for over one third (6 out of 16) of the significant variations. Some were positive, i.e. where a new lecturer had bedded in or an existing lecturer was replaced with one who appeared more effective.

The lecturer changes that had a negative effect were either from those who were new to teaching or where unexpected changes in staffing occurred.

Nearly one third (5 out of 16) of the significant changes were down to the assessment and materials offered to the students. These were either new materials or assessment methods being introduced, or no changes being made, which showed the positive effect of good materials on a less positive year group.

Finally, there were three significant results for which no information could be found. These lacked both the module monitoring reports as well as any indication of who taught them.

7.4.6 Summary

It seems that these new tests can be used either to assess the effect of a learning intervention, as happened with the experimental website or the changed format of 4~01, or as a retrospective tool to find factors which have a significant impact on student achievement. The 16 significant cases analysed showed how important the standard of teaching and the quality of materials available to the students has become. They also show the importance of experimentation in learning and how these can bring about positive student experiences.

7.5 Comparison to the Five Standard Tests

The five standard tests were compared to three new tests developed here for the number of cases where they indicated changes between years that were significant at the 5% level. There were two types of standard test used, parametric tests and non-parametric tests. The parametric tests, the Student t-test and the Fisher F-test are used in educational research even though, as has been shown in chapter 4, they may not be applicable. Student grades, as shown in appendix 9, are frequently not normally distributed or are used on groups too small to make a valid analysis.

The non-parametric tests used can be compared to the three tests created here. The LILM test is the equivalent of the Mann-Whitney U-tests. The $LISM_{Mean}$ test is the equivalent of the Levene test and the $LISM_{Median}$ is the equivalent of the Brown-Forsythe test. The quantity of year-on-year changes shown as significant at the 5% level is shown for each test below, but 16 would be expected (5% of 318):-

Table 27 - A comparison of statistical tests at the 5% level

Test	Number returned as significant at the 5% level
Student t-test	92
Fisher F-test	51
Mann-Whitney U-test	111
Levene test	75
Brown-Forsythe test	62
LILM test	12 (norm. dist.) or 16 (approximation)
LISM _{Mean} test	18 (norm. dist.) or 18 (approximation)
LISM _{Median} test	20 (norm. dist.) or 20 (approximation)

The table above demonstrates that the LIIM tests, LILM, LISM_{Mean} and LISM_{Median}, give results which are much more in line with expectations than any of the five standard statistical tests often used in educational research for the data in this study.

7.6 Conclusions

These new tests, based on like-with-like comparisons, have been shown to give better results in the educational setting of the institution than the five standard tests. The parametric tests seem not to work because the underlying data is either not normally distributed or the sample size is too small to allow valid conclusions to be drawn. The non-parametric tests appear not to work because there is an element of normality about the data – a Catch-22 situation.

It should be noted that the discussion above does not imply any criticism of the standard tests. They have been validated over decades and work if used correctly. The problem comes when these tests are applied in settings for which they were not designed. It seems that the analysis of grades may be just such a setting.

Chapter 8 Conclusions and Areas for Future Research

8.1 Introduction

The previous chapter examined the effect of using the LIIM tests, LILM, LISM_{Mean} and LISM_{Median}. This chapter seeks to draw the threads together by outlining the main contributions by giving a summary of the study, identifying the limitations of these new tests and suggesting areas for future research. The research question asked in this thesis was:-

Can the differences between modules, and their year groups, be used to assess whether a change in pedagogy affected learning outcomes?

8.2 Summary of the Study

The motivation for this research was the need to develop a robust statistical test to assess the impact of a new, educational theory-based, TEL site. The site was developed by searching the academic literature for examples of perceived good educational practice and then reducing these, where appropriate, to techniques which could be used to create a more useful website.

Two measures were needed, one to see whether a change in the average grade had taken place and one to assess any changes in the spread of grades when the new TEL site was used. This would usually be achieved by applying the Student t-test and Fisher F-test respectively. However, as was pointed out by Raymond Nickerson (Nickerson 2000), and Arianne Herrera-Bennett (Herrera-Bennett et al. 2020) and others, standard null hypothesis significance testing is unlikely to be the answer.

Doubts about the appropriate use of these standard statistical techniques when dealing with student grades were raised further when these tests were used to examine the year-on-year changes in module grades. It was found that a higher than expected number of these comparisons were significant at the 5% level. For the sample of results used, if the data had been random then 16 of these comparisons would have been expected to be classed as significant at the 5% level. In fact, for the five tests used (Student t-test, Fisher F-test, Mann-Whitney U-test, Levene test and the Brown-Forsythe test) the number of significant results ranged between 51 and 111 – significantly above the 16 expected. Using the binomial theorem, the 99% confidence intervals would have been between 7 and 25. Using the binomial distribution, the probability of getting even the lowest of these by chance is 6.8612×10^{-14} . These tests are therefore being used in this study for purposes for which they were not designed. It is not the tests that are at fault, but rather the way they were used in this educational setting. It is therefore possible for others to assume that these are safe techniques in this area when there may be problems.

A further interesting observation was that the grades in few runs appeared to normally distributed (as shown in appendix 9). This meant that non-parametric tests, ones which do not assume an underlying normal distribution have to be used. Changes in the location of the group are often measured using the Mann-Whitney U-test, whilst changes in spread are measured by using the Levene or Brown-Forsythe tests. Any tests developed would therefore have to shadow these.

The theoretical basis for the proposed new tests was to compare like-with-like. Rather than look at the year-on-year movements in grades, the new techniques compare the grades within a year to find the relative position of the modules. Each module was compared in turn to each of the other modules taken by that year group. Only grades from students who took both

modules were used in the analysis (the like-with-like principle) and then an average position for each module was calculated using these inter-module comparisons.

If a learning intervention took place, i.e. any activity which might affect scores such as changed teaching methods, times of classes, rooming or lecturer changes, etc., then the relative position of that module might be expected to change. If, say, a flipped classroom approach was being trialled then the hope would be that the scores in the trial year would go up. A straight comparison of the mean grades might be misleading. The year on which the experiment was tried might have been a particularly good one and so the grades would have gone up anyway. A within year comparison (comparing like-with-like again) might show that this module has improved its scores relative to the other modules. This inter-year comparison would only happen after the within-year (intra-year) analyses have taken place.

Three types of analysis were proposed; the average scores (the learning intervention location metric, LILM), the squared deviation from the mean (the learning intervention spread metric, mean – LISM_{Mean}) and the squared deviation from the median (the learning intervention spread metric, median – LISM_{Median}). These shadowed the Mann-Whitney U, Levene and the Brown-Forsythe tests respectively.

These three tests were carried out on the 48 modules over the eight academic years from 2011/12 to 2018/19 although in 20 cases there were insufficient students (less than 20) to make clear statistical comparisons. This left 318 valid comparisons which gave a sufficiently large sample for a distribution to become apparent for this data set. Whilst the distributions each looked normal when graphed it was shown that they were not. Whilst there was some skewness, the kurtosis was more of a problem.

Three methods were compared to find a way to use the distribution: using the normal distribution, using a dose-response curve and shifting the limits outwards until 16 instances

were left outside. The first and last options produced similar results whilst the dose-response curves gave far narrower limits which left a greater number outside the 95% confidence intervals than there should have been (46, 40 and 46 respectively as shown in table 16).

The 95% confidence intervals for the three tests, LILM, LISM_{Mean} and LISM_{Median} were estimated as:-

$$\text{LILM 95\% CI} = [-0.2, 0.2]$$

$$\text{LISM}_{\text{Mean}} \text{ 95\% CI} = [-0.58, 0.58]$$

$$\text{LISM}_{\text{Median}} \text{ 95\% CI} = [-0.55, 0.55]$$

for the data set used. Data from other institutions and levels of education may change these.

It was found that the LISM_{Mean} and LISM_{Median} gave very similar results and that there was little advantage to using one over the other therefore LISM_{Median} was dropped from the final analyses because it did not perform quite as well as the LISM_{Mean}.

There were then three types of analysis performed using these test results.

1. First, a 3x3 matrix was created by splitting the LILM and the LISM_{Mean} each into three similarly sized groups by taking 0.4 standard deviations either side of the mean and the number of modules in each cell counted. These nine cells classified the runs of modules and each classification had its own characteristics.
2. The second analysis looked at two special cases, the TEL experiment and a module which had been split in two. The analysis of the TEL experiment showed that the mean score had gone up from 55.7 to 59.6, an increase of 4%, but the standard deviation of the grades went up from 17.9 to 20.6. This would normally be taken as a success, after all, the average went up significantly, but the LILM dropped by 0.0235 and the

LISM_{Mean} of 0.3501 showed that the spread had increased. Some possible reasons for this were explained in chapter 7.

The second example of an experiment occurred where a module had been split in two with the expectation that smaller would equal more manageable. The mean score went up after the change (61.2 to 63.8) and the standard deviation went down (19.2 to 15.1). However, the LILM dropped by 0.0525, indicating that on average the change had not worked. On a more positive note, the LISM_{Mean} also decreased, showing that fewer students would have scored poorly.

3. The final analysis looked at the cases where the LISM lay outside the 95% confidence interval to see what conclusions could be drawn from them. The module monitoring reports for 16 runs showed that about one-third were explained by changes in staff, another third were down to changes in the assessment strategy or teaching materials.

These tests were used for three purposes; to classify results, to analyse the effects of specific learning interventions and to investigate anomalous cases to see what can be learned.

8.3 How the Research Aims and Objectives were Met

In order to answer the research question, this study sought:-

- 1) To develop a TEL which used educational theory in its design as an example of a learning intervention.
- 2) To assess the appropriateness of standard statistical techniques, such as the t- and F-test.
- 3) To offer and test an alternative statistical technique if the standard measures were

shown to be inappropriate.

The TEL website was created and ran in 2018/19 using a second-year degree level module on database design and management. The results from all the tests used show that the experiment made little difference. It fell within the central cell of the classification tables (see section 8.4 below). The average scores went down a little and their spread increased, a poor combination. This was unexpected, but the experiment suffered from both logistical and institutional handicaps which would not be the case if used again. Future researchers might find avoiding the pitfalls listed towards the bottom of section 7.4.2 helpful.

The five standard statistical tests showed poor levels of applicability to this educational setting since each failed to approach identifying 5% of cases when the 5% significance level was used.

Therefore, three new non-parametric tests were created which were built upon like-with-like comparisons. Together these three tests, called the learning intervention impact metrics (LIIM), each performed better than their more established counterparts.

Overall, the objectives were met and the research question was answered – new statistical tests could be developed which gave a more accurate representation of the null hypothesis significance tests in an educational setting. The TEL website was created from educational theory and the outcomes tested using conventional statistical tests as well as those newly developed. It was shown that these new tests could be used for purposes beyond their immediate purpose.

8.4 Main Contributions of the Research

The contributions of the research are:-

1. Empirical proof that the standard tests are inappropriate when comparing grades in successive years. There appear to be other factors such as the non-normality of grades, differing cohort characteristics and changes in the learning environment which can confound their use. The types of tests taught as standard tools in university when analysing data may not always be applied appropriately.
2. The use of the like-with-like principle was used to develop three new statistical tests, the learning intervention location metric (LILM), the learning intervention spread metric using the mean (LISM_{Mean}) and the learning intervention spread metric using the median (LISM_{Median}). Together these three make the learning intervention impact metrics (LIIM). These tests, when used in an educational setting, were shown to help accuracy. The three standard non-parametric tests (Mann-Whitney U-test, the Levene test and the Brown-Forsythe test) returned an average of 82 cases when 16 were expected. The new tests, which shadowed these standard tests, returned an average of 18 cases.
3. A system which classifies changes in module results was created using the newly developed tests which spit the results into nine cases based on their LILM and LISM_{Mean}. These cases can be used to inform a monitoring process. This in turn can be used as a research tool itself by highlighting educational activities which affect student achievement.
4. Providing students with optional extra materials did not work in this experiment. A solution would be to integrate the materials into the teaching so that they are not seen as something extra.

5. The literature review provided an overview of educational theories and organised them into categories for easy reference.

8.5 Limitations of the Study

Although the aims and objective of the study were fulfilled, there were limitations.

- The effectiveness of the TEL was hampered by having to make it clear to the students that the use of the TEL site was optional. This gave the impression that the students were being asked to complete a significant amount of extra work. The experiment also suffered by having items on the webpages turned off in some cases. This gave the impression that the site did not always work correctly – even though the students had been warned about this. Finally, the tasks used by the website were laid out in a scheme of work order. Afterwards it became apparent that presenting them in the order determined by the marking scheme would have made it easier for the students to see the relationship between the tasks and their grades.
- The three tests created are based upon within-year comparisons between modules. If there are, say, four modules within a year and one has an intervention which brings significant benefits then the other three modules will show a decline by comparison. This intradependency might show the other modules in a negative light if this not taken into account.
- Multiple changes within an academic year would not show the same level of significance as they would if they happened individually because the intradependency. Both this and the previous point might be overcome by normalising the grades for each module individually in turn, but that has not been part of this study.

8.6 Future Research

There are some areas for future research:-

- Increasing the quantity of data would help to establish the probability distribution more accurately. It is not known whether the distribution is different at other institutions or at different levels. This would need to be explored by others with access to grades data at those institutions.
- It would be interesting to take the lessons learned from the implementation of the TEL website and run it again. This has the ability to generate a significant volume of data on how students approach their learning. For example, which times of the day do they prefer, when are they at their most productive? Do they prefer to work in bursts, up-front of the teaching or as revision after the teaching has happened? Do they only dip in when having problems with coursework or revising for an exam?
- This TEL site was new to the students and ran on one module in one term. It might be instructive to see what would happen if it was used more widely and became a part of the learning system.
- Other researchers may want to examine the applicability of other statistical tests not included here.

8.7 Final Reflections

It has been said before, but bears repeating, that the techniques outlined here do not imply any criticism of the five standard statistical tests used as examples here. They are established and

are likely to remain that way. However, as this study shows, they may not be applicable in situation where a learning intervention is being judged.

The motivation for this study was to create a TEL which would allow a wider range of learning styles (whether temporary or permanent) to be used when studying. It has yet to be established whether the method proposed would benefit the students in the way hoped. However, the new statistical techniques, which were outlined before work began on the TEL, show promise and may eventually be of benefit to the educational research community.

References

- Ausubel, D.P., 1963. The psychology of meaningful verbal learning.
- Bacon, L. & MacKinnon, L., 2014. A flexible framework for metacognitive modelling and development. *Proceedings of ICEL*, 9, pp.7–14.
- Bergeron, P.-J. & Rivard, L., 2017. How to Engage in Pseudoscience With Real Data: A Criticism of John Hattie's Arguments in Visible Learning From the Perspective of a Statistician. *McGill Journal of Education/Revue des sciences de l'éducation de McGill*, 52(1), pp.237–246.
- Bergmann, J. & Sams, A., 2012. *Flip your classroom: Reach every student in every class every day*, International Society for Technology in Education. Available at: <https://books.google.co.uk/books?hl=en&lr=&id=72YqBgAAQBAJ&oi=fnd&pg=PT15&dq=Bergmann+and+Sams&ots=b0KUwTRvAq&sig=I7zukMou57bRoU76qj4aU7Nh5BU#v=onepage&q=Bergmann%20and%20Sams&f=false>.
- Binet, A. & Simon, T., 1916. *The development of intelligence in children: The Binet-Simon Scale*, Williams & Wilkins Company.
- Bloom, B.S., 1968. Learning for Mastery. Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. *Evaluation comment*, 1(2), p.n2.
- Bloom, B.S. & Krathwohl, D.R., 1956. Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.
- De Bono, E., 1989. *Six thinking hats*, Taylor & Francis.
- Bonwell, C.C. & Eison, J.A., 1991. *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports.*, ERIC.
- De Bra, P., Brusilovsky, P. & Houben, G.-J., 1999. Adaptive hypermedia: from systems to framework. *ACM Computing Surveys (CSUR)*, 31(4es), p.12.
- De Bra, P., Brusilovsky, P. & Houben, G.J., 1999. Adaptive hypermedia: from systems to framework. *ACM Computing Surveys (CSUR)*, 31(4es), p.12.
- Brown, M.B. & Forsythe, A.B., 1974. Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346), pp.364–367.
- Bruner, J.S., 1961. The act of discovery. *Harvard educational review*.
- Brusilovsky, P., 2001. Adaptive hypermedia. *User modeling and user-adapted interaction*, 11, pp.87–110.

References

- Brusilovsky, P., 1998. Methods and techniques of adaptive hypermedia. In *Adaptive hypertext and hypermedia*. Springer, pp. 1–43.
- Brusilovsky, P., Karagiannidis, C. & Sampson, D., 2004. Layered evaluation of adaptive learning systems. *International Journal of Continuing Engineering Education and Life Long Learning*, 14(4-5), pp.402–421.
- Carroll, J., 1963. A model of school learning. *The Teachers College Record*, 64(8), pp.723–723.
- Cemal Nat, M. et al., 2011. Impact of metacognitive awareness on learning in technology enhanced learning environments.
- Cemal Nat, M. & others, 2012. *Design, development and evaluation of technology enhanced learning environments: learning styles as an evaluation tool for metacognitive skills*.
- Chabris, C. & Simons, D., 2010. *The invisible gorilla: And other ways our intuitions deceive us*, Harmony.
- Chou, S.-W. & Liu, C.-H., 2005. Learning effectiveness in a Web-based virtual learning environment: a learner control perspective. *Journal of computer assisted learning*, 21(1), pp.65–76.
- Clarke, J.A., 1993. Cognitive style and Computer-Assisted Learning: problems and a possible solution. *Alt-j*, 1(1), pp.47–59.
- Cleveland-Innes, M. et al., 2019. Report of the Massive Open Online Course on Introduction to Technology-Enabled Learning (TEL MOOC 3).
- Cohen, J., 2016. The earth is round ($p < .05$). In *What if there were no significance tests?* Routledge, pp. 69–82.
- Connolly, T.M. et al., 2012. A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2), pp.661–686.
- Corcoran, T., 1927. The Winnetka school plan. *The Irish Monthly*, 55(644), pp.63–67.
- Craig M. Bennett, A.A.B.M.B.M. & Wolford, G.L., 2010. Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction. *Journal of Serendipitous and Unexpected Results*, 1 No. 1, pp.1–5. Available at: <https://www.improbable.com/ig-about/winners/#ig2012> [Accessed April 3, 2020].
- Cuevas, J., 2015. Is learning styles-based instruction effective? A comprehensive analysis of recent research on learning styles. *Theory and Research in Education*, 13(3), pp.308–333.
- D'Agostino, R. & Pearson, E.S., 1973. Tests for departure from normality. Empirical results for the distributions of b^2 and $\sqrt{b}$. *Biometrika*, 60(3), pp.613–622.

References

- Deterding, S. et al., 2011. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*. pp. 9–15.
- Deterding, S., 2015. Sixty-two Reasons Why “Gamification” Is Played Out. Available at: <http://www.slideshare.net/dings/pawned-gamification-and-its-discontents>.
- Dewey, J., 2007. *Experience and education*, Simon and Schuster. Available at: http://books.google.co.uk/books?id=UE2EusaU53IC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false.
- Dirnagl, U., 2019. The p value wars (again). Available at: <https://link.springer.com/article/10.1007/s00259-019-04467-5>.
- Downes, S., 2012. Connectivism and Connective Knowledge: essays on meaning and learning networks. *Stephen Downes Web*.
- Downes, S., 2008. Places to go: Connectivism \& connective knowledge. *Innovate Blog*.
- Dunn, R. et al., 1995. A meta-analytic validation of the Dunn and Dunn model of learning-style preferences. *The Journal of Educational Research*, 88(6), pp.353–362.
- Dweck, C., 2017. *Mindset: changing the way you think to fulfil your potential*, Hachette UK.
- Dweck, C.S., 2010. Even geniuses work hard. *Educational Leadership*, 68(1), pp.16–20.
- Efron, B., 1992. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics*. Springer, pp. 569–593.
- Entwistle, N., 2000. Promoting deep learning through teaching and assessment: conceptual frameworks and educational contexts. In *TLRP conference, Leicester*. Available at: <http://www.tlrp.org/pub/acadpub/Entwistle2000.pdf>.
- Entwistle, N., Hanley, M. & Hounsell, D., 1979. Identifying distinctive approaches to studying. *Higher education*, 8(4), pp.365–380.
- Felder, R.M. & Silverman, L.K., 1988. Learning and teaching styles in engineering education. *Engineering education*, 78(7), pp.674–681.
- Festinger, L., 1957. A theory of cognitive dissonance.
- Feuerstein, R. et al., 1985. Learning to learn: Mediated learning experiences and instrumental enrichment. *Special Services in the Schools*, 3(1-2), pp.49–82.
- Flavell, J.H., 1979. Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American psychologist*, 34(10), p.906.
- Fleming, N. & Baume, D., 2006. Learning Styles Again: VARKing up the right tree! *Educational Developments*, 7(4), p.4.

References

- Fleming, N.D., 1995. I'm different; not dumb. Modes of presentation (VARK) in the tertiary classroom. In *Research and development in higher education, Proceedings of the 1995 Annual Conference of the Higher Education and Research Development Society of Australasia (HERDSA)*, HERDSA. pp. 308–313.
- Fleming, N.D. & Mills, C., 1992. Not another inventory, rather a catalyst for reflection. *To improve the academy*, 11(1), pp.137–155.
- Flynn, J.R., 1987. Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological bulletin*, 101(2), p.171.
- Fröbel, F., 1885. *The education of man*, A. Lovell & Company. Available at: <http://ridgeway-ohara-muller-giffard.wikispaces.com/file/view/The+Education+of+Man+By+Friedrich+Froebel.pdf>.
- Gagne, R.M., 1988. Mastery learning and instructional design. *Performance Improvement Quarterly*, 1(1), pp.7–18.
- Gagne, R.M., 1968. Presidential address of division 15 learning hierarchies. *Educational psychologist*, 6(1), pp.1–9.
- Gardner, H., 1995. Reflections on multiple intelligences: Myths and messages. *Phi Delta Kappan*, 77, pp.200–200.
- Gee, J.P., 2003. What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)*, 1(1), pp.20–20.
- Goldenberg, A. & Gross, J.J., 2020. Digital Emotion Contagion. *Trends in Cognitive Sciences*.
- Gregorc, A.F., 1982. *An adult's guide to style*, Gregorc Associates Columbia, Conn.
- Harrison, L.A. et al., 2006. The consequences of stereotype threat on the academic performance of White and non-White lower income college students. *Social Psychology of Education*, 9(3), pp.341–357.
- Hattie, J., 2009. *Visible learning*, London: Routledge. Available at: http://samples.sainsburysebooks.co.uk/9781134024124_sample_526335.pdf.
- Herrera-Bennett, A.C. et al., 2020. Improving statistical inferences: Can a MOOC reduce statistical misconceptions about p-values, confidence intervals, and Bayes factors?
- Herrmann, N., 1991. The creative brain*. *The Journal of Creative Behavior*, 25(4), pp.275–295.
- Hill, N. & Stone, W.C., 1991. *Success through a positive mental attitude*, Simon and Schuster.
- Holt, J., 1966. How children fail. *Literacy Research and Instruction*, 6(1), pp.4–7.

References

- Honey, P. & Mumford, A., 1989. *Learning styles questionnaire*, Organization Design and Development, Incorporated.
- Illich, I., 1971. *Deschooling society*. New York, 56.
- Johansson, F., 2004. *The Medici effect: Breakthrough insights at the intersection of ideas, concepts, and cultures*, Harvard Business Press.
- Jovanovic, S., 2016. John Dewey's Experience and Education. *Partnerships: A Journal of Service-Learning and Civic Engagement*, 7(1), pp.55–57.
- Kazimoglu, C., 2013. *Empirical evidence that proves a serious game is an educationally effective tool for learning computer programming constructs at the computational thinking level*.
- Keller, F.S., 1968. Good-bye, teacher... *Journal of applied behavior analysis*, 1(1), p.79.
- Khan, S., 2015. Salman Kahn: Let's teach for mastery—not test scores [Video file].
- Kolb, D.A., 1984. *Experiential learning: experience as the source of learning and development*, Englewood Cliffs, NJ: Prentice Hall. Available at: <http://academic.regis.edu/ed205/kolb.pdf>.
- Kruger, J. & Dunning, D., 1999. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, 77(6), p.1121.
- Land, R., Meyer, J.H. & Smith, J., 2008. *Threshold concepts within the disciplines*, Sense Publishers.
- Landsberger, H.A., 1958. Hawthorne Revisited: Management and the Worker, Its Critics, and Developments in Human Relations in Industry.
- Lave, J. & Wenger, E., 1991. *Situated learning: Legitimate peripheral participation*, Cambridge university press.
- Levene, H., 1960. Robust tests for equality of variances: In Ingram Olkin, Harold Hotelling, et alia.
- Lublin, J., 2003. Deep, surface and strategic approaches to learning. *Centre for Teaching and Learning. UCD Dublin, nd*.
- MacKinnon, L. & Bacon, L., 2015. The move to student-centric learning: Progress and pitfalls. In *International Conference on e-Learning*. p. 188.
- MacKinnon, L. & Bacon, L., 2016. Virtual scaffolding-constructivism in online learning.
- Makel, M.C. et al., 2019. Questionable and Open Research Practices in Education Research.

References

- Mann, H.B. & Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pp.50–60.
- Marton, F. & Säljö, R., 1976. On qualitative differences in learning: I—Outcome and process. *British journal of educational psychology*, 46(1), pp.4–11. Available at: <https://scinapse.io/papers/1975912968> [Accessed December 31, 2019].
- Maslow, A.H., 1943. A theory of human motivation. *Psychological review*, 50(4), p.370.
- Mayo, E., 2009. Hawthorne and the Western Electric Company.
- McGregor, D., 1960. Theory X and theory Y. *Organization theory*, pp.358–374.
- Meyer, J. & Land, R., 2003. *Threshold concepts and troublesome knowledge: Linkages to ways of thinking and practising within the disciplines*, University of Edinburgh Edinburgh.
- Michalos, A.C., 1985. Multiple discrepancies theory (MDT). *Social Indicators Research*, 16(4), pp.347–413.
- Montessori, M., 1913. *Pedagogical anthropology*, Frederick A. Stokes Company. Available at: https://ia902507.us.archive.org/8/items/pedagogicalanthr00montiala/pedagogicalanthr00montiala_bw.pdf.
- Morrison, H.C., 1926. The practice of teaching in the secondary school.
- Myers, I.B., McCaulley, M.H. & Most, R., 1985. *Manual, a guide to the development and use of the Myers-Briggs type indicator*, Consulting Psychologists Press.
- Nickerson, R.S., 2000. Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological methods*, 5(2), p.241.
- Pashler, H. et al., 2008. Learning styles: Concepts and evidence. *Psychological science in the public interest*, 9(3), pp.105–119.
- Pask, G., Scott, B. & Kallikourdis, D., 1975. *Conversation, Cognition and Learning: A Cybernetic Theory and Methodology... with Appendices by BCE Scott and D.*, Elsevier.
- Perkins, D., 1999. The many faces of constructivism. *Educational leadership*, 57(3), pp.6–11.
- Pestalozzi, J.H., 1894. How Gertrude Teaches her Children translated by Lucy, E. Holland and Frances C. Turner. Edited with an introduction by Ebenezer Cooke. London: Swan Sonnenschein. Available at: <http://studentzone.roehampton.ac.uk/library/digital-collection/froebel-archive/gertrude/Gertrude%20complete.pdf>.
- Piaget, J., Cook, M. & Norton, W., 1952. *The origins of intelligence in children*, International Universities Press New York.
- Postman, N., 2009. *Teaching as a subversive activity*, Delta. Available at: <http://blogs.oregonstate.edu/smedcohort/files/2009/07/Teaching-as-a-Subversive-Activity->

References

- Postman.pdf.
- Riener, C. & Willingham, D., 2010. The myth of learning styles. *Change: The magazine of higher learning*, 42(5), pp.32–35. Available at: http://sites.psu.edu/wp-content/uploads/sites/10091/2012/03/Reiner-and-Willingham_2010.pdf.
- Robinson, K., 2008. Changing Education Paradigms. *RSA*.
- Seligman, M.E. et al., 2009. Positive education: Positive psychology and classroom interventions. *Oxford review of education*, 35(3), pp.293–311.
- Shen, L. & Hsee, C.K., 2017. Numerical Nudging: Using an Accelerating Score to Enhance Performance. *Psychological Science*, p.0956797617700497.
- Shrock, S.A., 1995. A brief history of instructional development. *Instructional technology: Past, present, and future*, 2, pp.11–19.
- Siemens, G., 2004. Connectivism. *A Learning Theory for the Digital Age*. In *eLearnSpace*.
- Siemens, G., 2005. Connectivism: A learning theory for the digital age. *International journal of instructional technology and distance learning*, 2(1), pp.3–10.
- Snedecor, G.W., 1934. Calculation and interpretation of analysis of variance and covariance.
- Sparrowhawk, A., 2007. Digital resources to support basic skills education for 14-19 year-olds. *The landscape, and opportunities for development, futurelab, Sparrowhawk & Heald*.
- Spearman, C., 1904. “General Intelligence”, objectively determined and measured. *The American Journal of Psychology*, 15(2), pp.201–292. Available at: <http://www.jstor.org/stable/pdf/1412107.pdf>.
- Steiner, R., 1989. The education of the child in the light of spiritual science. *Rhythms of learning*. Great Barrington, MA: Anthroposophical Press. (Original work published in 1907).
- Stone, C.A., 1993. What is missing in the metaphor of scaffolding. *Contexts for learning: Sociocultural dynamics in children’s development*, pp.169–183.
- Stonehouse, P., Allison, P. & Carr, D., 2011. ARISTOTLE, PLATO, AND SOCRATES: ANCIENT GREEK PERSPECTIVES ON EXPERIENTIAL LEARNING*. In *Sourcebook of experiential education: Key thinkers and their contributions*. pp. 18–25. Available at: http://www.researchgate.net/profile/Pete_Allison/publication/236896596_ARISTOTLE_PLATO_AND_SOCRATES_ANCIENT_GREEK_PERSPECTIVES_ON_EXPERIENTIAL_LEARNING*/links/0046351a0dd0115974000000.pdf.
- Student, 1908. The probable error of a mean. *Biometrika*, pp.1–25.
- Tate, W.K., 1914. Some Suggestive Features of the Swiss School System. Bulletin, 1913, No. 56. Whole Number 567. *United States Bureau of Education, Department of the Interior*,

- p.15.
- Terhart, E., 2011. Has John Hattie really found the holy grail of research on teaching? An extended review of Visible Learning. *Journal of curriculum studies*, 43(3), pp.425–438.
- Topping, K.J., Samuels, J. & Paul, T., 2007. Does practice make perfect? Independent reading quantity, quality and student achievement. *Learning and instruction*, 17(3), pp.253–264.
- Tucker, B., 2012. The flipped classroom. *Education Next*, 12(1), pp.82–83. Available at: http://wardwcom.webstarts.com/uploads/the_flipped_classroom_article.pdf.
- Tukey, J.W., 1949. Comparing individual means in the analysis of variance. *Biometrics*, 5(2), pp.99–114.
- Vygotsky, L., 1978. *Mind in society: The development of higher psychological processes* M. Cole, V. John-Steiner, S. Scribner & E. Souberman, ed., Harvard University Press.
- Wetzel Wolfgang, J.M.-D.N.P., 1967. *Statistische Tabellen*, Verlag Walter de Gruyter and Co, Berlin.
- Wickham, H. & Stryjewski, L., 2011. 40 years of boxplots. *Am. Statistician*.
- Wiggins, B.E., 2016. An overview and study on the use of games, simulations, and gamification in higher education. *International Journal of Game-Based Learning (IJGBL)*, 6(1), pp.18–29.
- Willingham, D., 2009. Why don't students like school? *American Educator*, Spring, pp.4–13.
- Wood, D., Bruner, J.S. & Ross, G., 1976. The role of tutoring in problem solving. *Journal of child psychology and psychiatry*, 17(2), pp.89–100.
- Zumbo, B.D. & Michalos, A.C., 2014. Michalos-Zumbo Well-Being Index. *Encyclopedia of Quality of Life and Well-Being Research*, pp.4032–4036.

The raw and clean grade data sets overviews

	Raw Data Set	Clean Data Set
n	30975	28610
Minimum	0	1
Maximum	100	100
Mean	57.620694	62.383817

Quartile 1	49	52
Median	62	64
Quartile 3	73	74
Std. Deviation	23.145536	16.818125
Coef. of Skew.	-1.062063	-0.54497
Kurtosis	0.798298	0.490802
p of normality	0%	0%

The LIIM statistical overviews

Learning Intervention Location Metric (LILM)

	Pos. Score	LILM
n	366	318
Minimum	0.649194	-0.519756
Maximum	1.575931	0.419805
Mean	1.001123	-0.002462
Quartile 1	0.942109	-0.054531
Median	0.990875	-0.002788
Quartile 3	1.050946	0.052369
Std. Deviation	0.097168	0.104327
Coef. of Skew.	0.856251	-0.305512
Kurtosis	3.813218	2.895432
p of normality	7.23713×10^{-17}	1.38022×10^{-7}

Learning Intervention Spread Metric – Mean (LISM_{Mean})

	Pos. Score	LISM _{Mean}
n	366	318
Minimum	0.3288724117583147	-1.470057
Maximum	2.969655576359739	1.434667
Mean	1.0555214369370154	0.000288
Quartile 1	0.882149234610999	-0.15311
Median	1.0341093534792956	-0.008614
Quartile 3	1.1968647958900964	0.145904
Std. Deviation	0.3187923015509858	0.292823
Coef. of Skew.	1.40149937818948	-0.234778
Kurtosis	5.827752557395204	4.666853
p of normality	1.782468×10^{-28}	4.03993×10^{-7}

Learning Intervention Spread Metric – Median (LISM_{Median})

	Pos. Score	LISM _{Median}
n	366	318
Minimum	0.32733854965426445	-1.319518
Maximum	2.9174810893016514	0.727551
Mean	1.0580049768516382	-0.000308
Quartile 1	0.8767857022885306	-0.155835
Median	1.0377572242210014	-0.006379
Quartile 3	1.2032349016825836	0.151188
Std. Deviation	0.32361560044699617	0.276171
Coef. of Skew.	1.4038667124939477	-0.363552
Kurtosis	5.505407522722788	2.350132
p of normality	4.927533×10^{-28}	7.447609×10^{-7}

The number of students

	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	294	223	262	293	253	265	201	216
4~02	227	230	256	282				
4~03	281	240	253	294	253	263	221	
4~04	141	104	132	164	145	158	133	
4~05	130	109	105	117	74	76	57	
4~06	159	134	176	207	203	172	162	235
4~07	127	124	164	193	190	166	156	247
5~01	53	75	55	68	64	56	74	61
5~02	133	75	117	141	136	111	120	98
5~03	66	86	78	99	112	94	102	96
5~04	55	194	139	156	145	100	108	100
5~05	43	24	17	35	53	39	53	27
5~06	39	24	17	34	137	37	51	26
5~07	95	104	91	132	155	135	150	119
5~08	43	56	56	87	80	85	78	71
5~09	76	43	27	39	23	15	11	
5~10	25	122	77	71	71	58	73	47
5~11	87	264	198	257	279	212	241	193
5~12	26	59	52	52	48	44	49	49
5~13	50	258	200	249	268	197	232	184
5~14	33	36	30	50	70	61	69	41
6~01	17	22	24	36	17	21		21
6~02	5	5	12	11	15	17	3	16
6~03	59	26	22	21	25	17	23	19
6~04	67	54	28	23	31	14	12	11
6~05	56	108	104	76	80	63	77	8
6~06	54	55	34	26	33	25	14	40
6~07	64	70	82	53	54	46	31	36
6~08	43	23	22	18	20	25	9	17
6~09	46	23	16	10	9	8	8	7
6~10	69	144	150	105	115	95	76	96
6~11	21	91	93	57	65	67	43	28
6~12	96	192	107	92	111	83	96	69
6~13	23	21	112	11	14	20	26	17
7~01	108	94	61	80	73	64		
7~02	90	54	56	63	37	45	34	53
7~03	47	29	44	30	37	18	22	19
7~04	44	44	45	43	24	21	26	31
7~05	79	60	72	69	65	45	38	58
7~06	31	31	23	37	24	25	12	26
7~07	47	39	32	37	15	26	10	19
7~08	12	53	53	43	29	15	17	15
7~09	19	54	64	81	54	48	49	47
7~10	77	14	34	43	34	48	33	42
7~11	41	60	54	52	44	23	36	40
7~12	15	15	31	19	15	13	22	20
7~13	22	13	12	23	32	13	15	11
7~14	61	69	64	61	34	36	37	38

The mean score in each run

	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	63.3197	60.6054	68.3168	71.4915	63.1502	61.1585	63.7761	63.0139
4~02	58.9736	59.6348	60.5742	57.7908				
4~03	67.2100	70.9125	69.4822	63.4184	55.2609	57.7034	53.1765	
4~04	57.4397	62.9423	56.6667	57.4451	55.2897	54.2532	56.0752	
4~05	70.9615	78.2752	59.4381	74.0855	59.1757	55.6974	76.9474	
4~06	61.8994	64.5821	64.7159	74.4444	66.2365	68.8605	72.2407	66.2809
4~07	60.4646	63.2339	64.1890	61.4767	58.6053	60.3253	60.8590	62.1579
5~01	66.5660	59.7200	59.9455	58.0441	57.5156	55.3571	51.0000	58.5246
5~02	63.6391	59.7200	61.0513	56.2837	58.8088	53.7117	50.4250	61.4592
5~03	52.8182	57.4070	55.6154	58.0000	61.0714	54.8191	55.2843	61.4688
5~04	52.9636	56.2216	66.3237	70.1474	60.8414	67.9900	55.7407	59.6400
5~05	64.4884	65.5417	62.7647	64.3714	58.0377	50.3590	56.4151	61.2593
5~06	66.9487	68.1667	67.2353	68.4118	57.8321	61.7838	67.8431	67.0385
5~07	60.3579	62.9615	65.1758	58.9318	60.0387	55.4148	60.7000	68.3277
5~08	57.8605	56.0179	62.2500	56.5632	57.3750	55.0471	52.9615	59.7746
5~09	61.5395	61.5814	61.9259	59.9487	58.2174	45.8000	58.4545	
5~10	73.1600	68.6148	66.7273	74.4789	63.9718	62.9310	63.2055	64.3191
5~11	62.2989	64.6212	63.1818	56.3580	59.0466	59.6698	58.3029	65.3782
5~12	59.5385	54.0678	54.9808	58.9231	60.5625	56.3636	55.6735	60.8163
5~13	62.5400	65.0310	62.6400	67.3574	65.0858	63.0660	60.7931	66.0380
5~14	63.5758	65.8611	79.7667	72.9000	66.4857	71.6721	66.3188	63.8049
6~01	72.4706	63.1818	66.3333	73.1667	67.4706	64.0000		69.1905
6~02	66.4000	66.0000	69.2500	54.4545	72.7333	60.1765		59.6875
6~03	46.2542	62.5000	59.4545	58.6190	60.1200	63.8824	58.4783	67.5789
6~04	64.2985	62.2593	60.6071	53.2174	53.5806	53.6429	51.1667	70.9091
6~05	55.3571	59.9167	65.4808	59.2763	65.3750	71.5556	68.4416	63.1250
6~06	64.1296	64.0182	67.1176	70.5000	65.2121	65.2800	72.0714	65.2500
6~07	60.3906	63.3286	62.5122	60.9811	62.4815	59.0435	57.3226	62.3056
6~08	58.8372	66.5217	63.0000	63.5000	69.1500	71.0000	68.3333	71.7059
6~09	56.6957	66.5217	61.6875	73.2000	51.1111	66.0000	60.7500	55.8571
6~10	62.1739	61.2361	62.9400	66.0857	66.5913	62.4632	59.2237	63.9375
6~11	63.1429	64.2418	63.7957	62.4912	58.6154	58.5224	57.3721	54.4286
6~12	66.1250	66.1250	65.0000	64.7283	55.8829	58.6747	51.8854	54.5652
6~13	61.3478	69.4286	50.5000	66.1818	60.2143	61.0000	54.0769	60.2941
7~01	58.7407	63.1596	62.6393	62.2250	66.0411	67.8594		
7~02	57.4000	60.1296	66.6071	63.3651	61.6216	65.5778	61.8824	60.2264
7~03	65.5745	58.4828	62.4091	64.7667	68.4865	75.9444	62.5909	68.6316
7~04	63.5909	72.8182	63.3778	63.3023	68.2500	69.5238	77.3846	70.6452
7~05	57.1646	59.0167	62.9722	60.5942	65.0000	60.1556	59.3158	58.7759
7~06	63.9355	64.5484	63.3043	64.4865	65.9583	64.7200	70.7500	65.1154
7~07	51.2128	50.0256	60.4688	59.5946	58.0000	61.8846	51.2000	63.8947
7~08	66.3333	63.5283	63.6415	62.0465	62.5862	65.9333	65.2941	66.8667
7~09	61.3684	64.9259	67.8906	67.1728	74.0370	72.4792	70.3061	68.2979
7~10	61.7013	63.5000	66.0882	65.0698	65.9118	66.7500	67.0606	65.0952
7~11	63.8293	59.7167	62.8704	62.8077	63.2045	64.0435	63.6389	62.3750
7~12	64.0000	73.3333	63.8387	65.8947	72.1333	66.3846	67.5000	66.9000
7~13	65.4545	68.5385	61.5000	62.1304	67.2188	73.6154	71.1333	64.0909
7~14	64.0492	67.4058	67.7812	67.1803	68.8235	66.6667	74.4054	56.8684

The median score for each run

	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	64	60	70	74	66	62	66	64
4~02	62	64	64	60				
4~03	72	78	74	65.5	56	60	54	
4~04	61	61.5	59	57	58	55.5	55	
4~05	73	86	59	79	58	56	77	
4~06	61	65.5	66	78	67	72	75	69
4~07	61	64.5	64.5	61	59	60	56.5	63
5~01	73	63	62	59	60.5	59.5	48	60
5~02	68	63	61	56	60	55	49	58
5~03	50	60	56	57	62.5	55	55	63
5~04	54	56	68	74	58	72	55.5	61.5
5~05	66	67	60	63	61	49	56	62
5~06	70	69.5	65	69.5	59	58	68	68
5~07	61	64	66	57	60	53	59.5	69
5~08	61	59.5	66	59	59	55	52	56
5~09	62	58	63	63	56	45	61	
5~10	72	70	67	76	67	65	66	68
5~11	64	68	65.5	59	62	62	58	66
5~12	61	55	57.5	60	60	55	51	61
5~13	66	66	66	68	66	64	61	67
5~14	66	66.5	81.5	80	70	74	68	62
6~01	70	65.5	71	74	67	65		68
6~02	62	63	75.5	46	73	65		60.5
6~03	45	62	60	54	66	70	60	67
6~04	65	65	63	50	54	50.5	51	77
6~05	65	65	66	60	70	75	71	63
6~06	70	66	70	70	68	68	75	65
6~07	63.5	65	64	60	64	62.5	60	65
6~08	60	70	64.5	64.5	70	70	70	74
6~09	54.5	70	59	79	46	63.5	61	60
6~10	65	63	65	67	70	63	60.5	65
6~11	62	66	65	63	62	60	57	56
6~12	70	70	66	66	60	60	53.5	54
6~13	57	73	51	63	65.5	61.5	49	60
7~01	56	64	64	66	68	67.5		
7~02	58	60	66	64	60	64	66.5	60
7~03	65	57	62	66	68	75	63.5	70
7~04	61	75	62	62	70	70	85.5	71
7~05	56	60	63	62	65	59	63	60
7~06	64	64	63	64	64.5	66	69.5	65
7~07	54	50	61	62	60	64.5	53.5	63
7~08	69.5	64	64	63	61	65	66	68
7~09	65	65.5	68	68	73.5	72	70	72
7~10	61	62.5	65	68	67	69.5	69	63.5
7~11	63	59	63	61	64	66	63.5	64
7~12	65	72	60	67	70	64	66.5	65.5
7~13	67	66	65.5	60	66	81	70	67
7~14	64	68	68.5	66	68	67	77	57.5

The standard deviation of the scores for each run

	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	21.0930	22.3066	18.0962	16.4840	19.6772	19.2373	15.1369	15.2896
4~02	19.3635	20.6312	15.6712	11.6079				
4~03	19.2894	19.8177	17.9873	17.0313	18.6069	17.0691	15.2423	
4~04	21.5802	21.4444	19.7992	17.2033	17.2510	17.7939	21.2929	
4~05	15.9191	20.8375	16.4871	17.7975	18.5586	17.5795	12.8277	
4~06	18.6975	18.6174	17.1057	16.0529	16.2454	18.5577	18.0891	19.1067
4~07	21.9714	21.7305	20.0330	17.9174	20.6338	19.0404	20.6746	18.5554
5~01	20.7471	19.9618	15.9518	19.9383	15.1228	17.6331	16.1041	15.9192
5~02	19.3016	19.9618	16.4101	13.6457	14.7846	19.5650	22.0895	17.3266
5~03	17.2019	20.5285	18.9594	20.6511	18.7216	20.1580	16.1756	16.5739
5~04	19.5154	17.1687	18.0800	17.6005	17.7227	19.3294	17.8804	20.6459
5~05	17.2908	12.2154	13.4235	16.7281	20.1694	14.4941	14.6069	17.1903
5~06	17.1095	13.9710	17.6193	18.3090	17.7012	15.4706	13.5667	10.1212
5~07	19.1066	15.4454	15.1523	15.1321	16.0251	17.2187	16.1190	16.4724
5~08	23.1502	18.5791	17.1626	20.5294	16.9253	17.4287	17.5503	20.5997
5~09	15.2451	16.4855	15.6154	19.7470	16.4399	20.9939	17.0667	
5~10	7.8935	12.4664	15.1168	12.1524	10.6368	12.8416	14.2809	10.1234
5~11	16.6413	12.1527	12.5955	12.2342	13.9054	11.7205	12.7271	13.5115
5~12	13.3244	14.4029	14.2037	14.4261	17.0879	14.0002	14.9267	13.8574
5~13	14.6095	11.2228	12.7803	9.5860	9.1125	8.7348	9.3049	8.4066
5~14	13.5440	18.7467	12.9553	19.9195	17.5967	14.2907	16.9649	19.7778
6~01	9.7026	14.9623	17.6307	9.2567	12.4354	16.6553		12.3516
6~02	13.6858	15.6205	23.1089	24.7522	16.0733	21.0987		15.3611
6~03	17.1782	18.8409	17.4349	17.3219	22.9770	23.6296	22.7594	16.9191
6~04	18.1768	19.1963	15.5429	15.0543	14.0375	25.2331	17.4399	15.4561
6~05	25.3028	15.3459	17.4795	16.6089	15.5241	17.4215	13.2163	11.1283
6~06	19.2433	13.5749	13.2087	13.5713	15.5156	15.0126	15.0306	10.8220
6~07	12.9610	14.8994	11.5878	14.3641	14.7453	15.0066	14.2323	15.3731
6~08	12.9024	8.4038	11.5882	10.6121	7.0058	10.8128	14.2916	14.5635
6~09	14.0426	8.4038	13.2198	11.0030	22.5746	15.9553	13.3924	19.0651
6~10	15.8299	13.3306	11.5458	11.3113	14.7558	12.4830	14.1418	13.6850
6~11	15.1634	13.1878	13.2514	13.8140	15.5751	14.4005	13.5910	14.9627
6~12	13.8642	13.8278	9.8441	12.3218	16.6383	13.5648	15.3279	14.1333
6~13	10.4386	15.1511	16.4350	12.7265	19.0714	12.4689	13.9454	11.1902
7~01	15.0980	14.4060	12.7711	16.7226	13.7684	12.1405		
7~02	14.0415	12.4173	12.1800	12.6469	9.3016	11.1895	18.3167	13.9279
7~03	9.5616	13.7077	12.2119	7.6234	10.7719	9.3775	4.2612	4.9127
7~04	19.6898	13.1012	14.3405	16.1283	20.4668	13.7354	23.0739	12.8337
7~05	12.7515	12.3460	10.3079	12.0090	10.8700	10.8940	11.8734	11.3262
7~06	6.4753	6.8646	6.2336	5.8387	7.8268	6.1205	6.3407	13.5833
7~07	17.2462	16.3634	13.2006	12.5841	9.9067	14.1713	17.1581	10.4716
7~08	14.5935	12.7876	13.5065	13.4836	13.5897	12.1917	9.3992	17.2497
7~09	11.6247	12.3851	6.6192	8.7504	7.1874	9.7566	12.2546	13.5003
7~10	10.6512	10.6825	7.3086	14.9609	15.4633	16.2186	22.1500	14.7562
7~11	17.3866	14.3469	13.8333	11.8967	11.3252	15.2240	13.4039	16.4561
7~12	18.0436	11.9383	14.1730	20.2619	11.9634	15.1412	14.3651	10.5076
7~13	9.2671	11.5081	13.7477	12.1629	13.6966	15.7984	12.5634	19.5471
7~14	10.6041	14.9974	10.4220	13.1979	16.1629	10.0371	14.3903	15.8867

The coefficient of skewness of the scores for each run

	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	-0.4402	-0.2269	-0.7430	-0.6468	-0.5945	-0.4196	-0.1706	-0.2364
4~02	-0.9167	-1.0892	-1.2029	-0.7083				
4~03	-1.0759	-1.3964	-1.5303	-1.0320	-0.5747	-0.6891	-0.6578	
4~04	-0.5940	-0.4719	-0.3174	-0.5933	-0.4759	-0.4221	-0.2839	
4~05	-0.4923	-1.1411	-0.0953	-0.5309	0.2715	0.2605	-0.4410	
4~06	-0.4210	-0.5235	-0.6002	-0.7283	-0.4884	-0.6380	-0.7133	-0.5361
4~07	-0.2847	-0.2210	-0.3956	0.0292	-0.0674	-0.2658	-0.0405	-0.3901
5~01	-1.2706	-0.4959	-0.0143	-0.7344	-0.9762	-0.8262	0.1454	-0.7974
5~02	-0.8824	-0.4959	-0.2891	0.1318	-0.2956	-0.4495	-0.4885	0.3653
5~03	-0.1894	-0.4260	-0.3105	-0.1084	-0.3693	-0.3777	0.0697	-0.4941
5~04	-0.2270	-0.1914	-0.2094	-0.5689	0.2389	-0.1746	0.0488	-0.0502
5~05	-0.3658	-0.3266		-0.0897	-0.6783	-0.2245	-0.2860	-1.1501
5~06	-0.3991	-0.3188		-0.1852	-0.2999	0.2202	-0.3282	-0.4061
5~07	-0.2559	-0.5629	-0.5560	0.0768	-0.2167	0.2674	0.0288	-0.5445
5~08	-0.3192	-0.6260	-0.4370	-0.2541	-0.4674	-0.0989	-0.0810	0.2815
5~09	0.0993	0.1843	-0.1048	-0.6010	0.1165			
5~10	-0.4041	-0.7356	-0.2683	-0.9017	-0.2530	-0.4208	-1.2605	-1.1891
5~11	-0.8860	-1.1317	-1.0974	-1.5883	-1.2587	-1.1801	-0.4446	-1.0330
5~12	-0.3564	0.1676	-0.1684	-1.0810	-0.0652	0.3448	0.3467	-0.2640
5~13	-0.8900	-0.7420	-1.3130	-1.3239	-1.6569	-1.2081	-0.5157	-0.3555
5~14	-0.1311	-0.3212	-0.9575	-1.3161	-0.7289	-0.8399	-0.3376	0.2835
6~01		-0.8762	-0.7518	-0.6946		-1.2065		0.1403
6~02								
6~03	-0.0379	-0.6008	-0.8394	-0.0790	-1.0671		-0.6130	
6~04	-0.1826	-0.4162	-1.0307	0.4027	-0.2290			
6~05	-0.6735	-0.7938	-1.0790	-0.3637	-0.7588	-1.9738	-1.0776	
6~06	-1.0447	-0.8758	-0.4413	-0.0987	-0.4992	-0.1925		-0.0212
6~07	-0.9205	-1.3007	-0.7823	-0.2747	-0.7179	-0.5336	-0.8323	-0.3443
6~08	0.4370	-0.6633	-1.0156		0.2794	-0.5250		
6~09	0.6798	-0.6633						
6~10	-0.6786	-0.8954	-0.3686	-0.5863	-1.6376	-0.8364	0.0399	-0.3951
6~11	-0.2031	-1.1641	-1.1653	-0.9243	-0.5609	-0.2077	0.1589	-0.3553
6~12	-1.1100	-1.1100	-0.8338	-0.9500	-0.4050	-0.2300	-0.2762	-0.0988
6~13	0.5941	-0.5783	-0.1124			-0.2900	0.7839	
7~01	-0.1611	0.0492	-0.2468	-0.9219	-0.0670	-0.1392		
7~02	-0.3474	-0.1543	-0.0577	-0.9017	0.8504	0.5083	-0.5694	-0.1995
7~03	-0.1671	-0.1544	-0.3430	-0.1498	-0.8953		0.0784	
7~04	-0.2432	-0.6568	0.3470	-0.2315	-0.4229	-0.9132	-2.0856	-0.4407
7~05	-0.2763	-1.1497	-1.2139	-0.7292	-0.1334	-0.6070	-0.6752	-0.4588
7~06	-1.5401	0.9398	0.1174	0.3029	0.5642	-0.6513		-1.4107
7~07	-1.0821	-0.7930	-0.3075	-0.3258		-0.8857		
7~08		-0.9307	-0.5658	-0.9607	-0.3491			
7~09		-1.5325	0.2459	-0.0348	-0.2211	-0.1447	-0.1301	-0.3615
7~10	-0.1574		-0.0701	-1.4983	-1.1986	-0.9556	-1.2497	-0.2078
7~11	-0.2529	-0.6359	-1.3111	0.4732	-0.3340	-0.6779	-0.8038	-0.9874
7~12			0.8576				0.3823	0.3272
7~13	-1.2139			-0.3902	0.0452			
7~14	-0.6273	-0.9479	-1.2820	-0.5974	-1.1247	-0.1649	-2.2956	-1.0349

The kurtosis of the scores for each run

	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	-0.3795	-0.3869	0.6623	0.2257	-0.1187	-0.1795	-0.9178	-0.5535
4~02	0.6472	0.8643	1.5702	0.1344				
4~03	0.9473	1.6620	2.5782	1.8365	0.2766	0.4689	1.1074	
4~04	-0.4796	-0.1091	-0.5174	0.5773	0.0325	0.3494	-0.5895	
4~05	-0.6664	0.5336	-0.2644	-0.8659	-0.9334	-0.6217	0.1111	
4~06	0.3226	-0.0706	-0.1411	-0.1133	0.4780	-0.4177	0.3345	-0.1801
4~07	-0.5111	-0.6793	-0.4084	-0.9213	-0.6500	0.0329	-0.7139	0.0514
5~01	1.4440	0.0074	-1.0235	0.5619	0.8715	0.1194	-0.1411	0.9543
5~02	0.6055	0.0074	0.2847	-0.7676	0.8946	-0.1089	-0.3105	-1.0876
5~03	-0.3944	-0.3659	-0.2551	-0.7616	-0.3061	-0.1979	0.3134	0.4373
5~04	-0.7778	-0.5264	-0.6480	-0.4816	-0.7318	-1.1643	-0.6124	-1.1203
5~05	-0.4888	0.4306		-1.1386	0.3809	0.5422	-0.3066	2.7364
5~06	-0.4151	-0.9901		-1.0838	-0.7930	-1.1969	-1.0057	0.3516
5~07	-0.6731	1.6966	0.9414	-0.0518	0.4624	0.3507	-0.3015	0.7214
5~08	-0.9106	0.0832	-0.5896	-0.8037	0.4727	-0.1057	0.2649	-0.8608
5~09	-0.4861	-0.6353	-1.0025	-0.4095	-0.8697			
5~10	-0.3864	0.1923	-0.3420	0.9495	0.3803	-0.7284	3.2059	0.5457
5~11	0.8053	1.7975	2.0950	3.7227	1.9274	2.3726	0.5550	1.8856
5~12	-0.1080	-0.7639	-0.4057	2.7993	-1.3684	-1.0014	-0.7462	-0.8568
5~13	-0.1821	2.4454	2.0749	5.2474	8.4179	4.3381	1.5243	-0.0306
5~14	-0.2641	-0.6993	-0.0846	1.2806	0.0935	0.0645	-0.5916	-0.9139
6~01		0.1809	0.0593	0.6720		1.7146		-1.1476
6~02								
6~03	-0.0328	-0.1682	0.2289	-0.6731	0.6224		-0.3326	
6~04	-1.1210	-0.3460	1.1407	-0.6938	-0.7607			
6~05	-0.5761	1.4886	2.1011	-0.0572	-0.6138	4.0859	0.9885	
6~06	0.7805	0.6683	-0.7532	-0.8171	-0.6132	-1.0590		-0.3177
6~07	0.7992	2.0975	0.2324	-0.8132	0.5503	-0.3310	0.6070	-0.8194
6~08	0.1077	0.1481	0.9047		-0.1828	1.0654		
6~09	-0.0048	0.1481						
6~10	-0.0265	0.8380	-0.3427	-0.1066	4.1622	3.2613	-1.0242	0.6111
6~11	-1.0004	1.6390	1.5618	1.0279	-0.2452	-0.7076	-0.6718	0.1683
6~12	1.3665	1.3665	0.2441	1.2024	-0.0465	-0.7337	-0.3551	0.2229
6~13	-0.5070	-0.9220	0.0381			-0.1087	-0.6115	
7~01	0.3154	-0.3601	-0.1086	0.6366	-0.7550	-0.2542		
7~02	1.3237	0.2787	-0.5675	2.0858	0.6717	-0.2287	-0.4689	1.2637
7~03	-0.9754	-0.3996	0.0668	-0.5227	1.6834		-1.2042	
7~04	0.5095	0.0623	-0.6276	-0.0461	-0.2303	0.8821	3.4347	-0.2783
7~05	-0.1302	2.6234	3.5454	1.2612	-0.1576	1.4813	0.7711	1.0797
7~06	5.3426	0.1980	-0.1264	1.1376	-0.1779	-0.3521		4.2539
7~07	2.1688	0.4455	0.4124	-0.3139		0.2852		
7~08		1.8721	2.1570	1.1778	2.0026			
7~09		4.9547	0.0657	-0.3758	1.3565	-0.9605	-0.8120	-0.8700
7~10	0.4421		-1.0998	4.1148	3.0199	1.6646	1.0428	-0.3727
7~11	0.3099	1.5958	2.9026	1.2201	1.4021	1.2513	1.2284	1.4021
7~12			-0.4730				-0.9089	-0.1170
7~13	2.6128			0.0313	-0.7348			
7~14	1.3724	1.8783	6.2943	1.7636	2.5763	-0.9384	7.2769	1.7699

p of normality (D'Agostino-Pearson) for each run

	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	0.34%	16.51%	0.00%	0.01%	0.10%	2.01%	0.00%	3.75%
4~02	0.00%	0.00%	0.00%	0.00%				
4~03	0.00%	0.00%	0.00%	0.00%	0.09%	0.00%	0.00%	
4~04	0.80%	13.45%	11.70%	0.32%	6.04%	5.14%	8.69%	
4~05	0.82%	0.00%	85.83%	0.05%	1.80%	26.88%	29.31%	
4~06	5.44%	4.77%	0.68%	0.03%	0.70%	0.22%	0.11%	0.45%
4~07	16.43%	6.57%	6.07%	0.00%	3.83%	33.65%	3.17%	4.16%
5~01	0.03%	17.84%	3.50%	2.14%	0.26%	3.38%	85.70%	0.99%
5~02	0.02%	17.84%	27.61%	1.63%	5.32%	14.36%	7.56%	0.00%
5~03	70.50%	21.14%	48.34%	8.03%	22.71%	29.61%	61.50%	6.88%
5~04	23.88%	9.93%	6.76%	0.66%	1.58%	0.00%	27.08%	0.00%
5~05	49.58%	42.84%		6.54%	6.97%	41.76%	63.90%	0.15%
5~06	50.80%	33.09%		10.92%	0.49%	2.11%	3.54%	38.88%
5~07	12.75%	0.20%	1.80%	92.01%	22.31%	24.27%	80.09%	1.41%
5~08	14.81%	11.91%	23.63%	5.02%	10.76%	91.73%	67.39%	5.21%
5~09	65.81%	60.18%	33.88%	24.31%	62.73%			
5~10	62.56%	0.48%	54.02%	0.26%	39.17%	14.17%	0.00%	0.33%
5~11	0.16%	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%	0.00%
5~12	63.26%	25.76%	79.83%	0.02%	0.00%	6.42%	24.33%	17.48%
5~13	3.59%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	13.66%
5~14	93.52%	45.98%	6.99%	0.04%	3.87%	2.58%	26.12%	17.87%
6~01		12.02%	19.05%	8.14%		0.65%		26.97%
6~02								
6~03	93.68%	34.24%	13.12%	89.54%	3.05%		37.54%	
6~04	0.11%	39.03%	1.56%	55.94%	55.18%			
6~05	7.14%	0.03%	0.00%	38.15%	1.04%	0.00%	0.04%	
6~06	0.36%	1.26%	31.16%	64.61%	35.26%	27.81%		99.47%
6~07	0.45%	0.00%	1.36%	18.53%	4.25%	27.84%	6.01%	30.72%
6~08	36.43%	24.48%	3.31%		76.74%	12.95%		
6~09	12.85%	24.48%						
6~10	6.31%	0.00%	12.58%	4.98%	0.00%	0.00%	0.49%	9.92%
6~11	45.35%	0.00%	0.00%	0.46%	15.83%	25.50%	57.27%	51.06%
6~12	0.00%	0.00%	0.27%	0.03%	19.87%	12.46%	43.11%	70.18%
6~13	39.28%	27.46%	81.14%			72.57%	17.28%	
7~01	49.16%	82.61%	68.19%	0.23%	20.61%	87.68%		
7~02	3.35%	61.68%	68.21%	0.06%	3.78%	32.09%	30.90%	13.36%
7~03	9.76%	91.94%	50.80%	87.19%	0.78%		18.34%	
7~04	40.60%	14.57%	39.98%	73.28%	59.01%	5.29%	0.00%	52.02%
7~05	56.45%	0.01%	0.00%	0.63%	88.96%	2.75%	7.23%	6.64%
7~06	0.00%	5.77%	88.55%	16.50%	40.18%	31.51%		0.01%
7~07	0.06%	6.03%	43.12%	65.85%		8.69%		
7~08		0.15%	0.85%	0.73%	5.77%			
7~09		0.00%	60.49%	85.36%	10.80%	10.78%	29.45%	13.39%
7~10	45.36%		10.64%	0.00%	0.04%	0.26%	0.40%	80.77%
7~11	51.39%	1.01%	0.00%	5.87%	9.12%	6.98%	2.38%	0.57%
7~12			10.79%				42.07%	68.97%
7~13	0.21%			54.93%	68.08%			
7~14	1.48%	0.04%	0.00%	0.91%	0.10%	25.63%	0.00%	0.31%

Student t-test values by run

	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	1.4135	-4.2023	-2.1627	5.3895	1.1648	-1.5912	0.5111
4~02	-0.3532	-0.5685	2.3548				
4~03	-2.1564	0.8398	4.0453	5.3510	-1.5548	3.0508	
4~04	-1.9779	2.3303	-0.3617	1.0977	0.5139	-0.7952	
4~05	-3.0731	7.3161	-6.3384	5.5475	1.1787	-7.7124	
4~06	-1.2259	-0.0657	-5.7349	5.1457	-1.4599	-1.6842	3.1212
4~07	-1.0038	-0.3862	1.3500	1.4548	-0.8133	-0.2411	-0.6546
5~01	1.8803	-0.0691	0.5740	0.1708	0.7219	1.4662	-2.7159
5~02	1.3889	-0.5034	2.5481	-1.4779	2.3307	1.1933	-4.0339
5~03	-1.4637	0.5787	-0.7905	-1.1331	2.3052	-0.1788	-2.6568
5~04	-1.2044	-5.1786	-1.8388	4.5683	-2.9898	4.7478	-1.4589
5~05	-0.2636		0.0000	1.5399	2.0229	-1.9716	-1.3204
5~06	-0.2934		0.0000	3.0983	-1.2359	-1.9493	0.2666
5~07	-1.0611	-1.0076	3.0268	-0.5983	2.3674	-2.6759	-3.8176
5~08	0.4394	-1.8439	1.7211	-0.2774	0.8696	0.7606	-2.1789
5~09	-0.0140	-0.0868	0.4344	0.3540			
5~10	1.7498	0.9570	-3.4201	5.4820	0.5036	-0.1142	-0.4644
5~11	-1.4020	1.2403	5.8232	-2.3682	-0.5259	1.1834	-5.5993
5~12	1.6498	-0.3354	-1.4042	-0.5197	1.2823	0.2292	-1.7675
5~13	-1.3627	2.1278	-4.4670	2.7621	2.4035	2.5929	-5.9571
5~14	-0.5758	-3.4345	1.6842	1.8628	-1.8342	1.9318	0.7060
6~01	0.0000	-0.6506	-1.9603		0.0000		0.0000
6~02						0.0000	
6~03	-3.9002	0.5772	0.1576	-0.2462		0.0000	
6~04	0.5983	0.3931	1.7135	-0.0912			
6~05	-1.4338	-2.4654	2.4018	-2.3705	-2.2393	1.2020	
6~06	0.0350	-1.0573	-0.9713	1.3723	-0.0167		0.0000
6~07	-1.2127	0.3796	0.6816	-0.5330	1.1526	0.5037	-1.3688
6~08	-2.5743	1.1709		0.0000	-0.6614		
6~09	-3.0842						
6~10	0.4516	-1.1729	-2.1592	-0.2832	2.1615	1.5893	-2.2105
6~11	-0.3345	0.2288	0.5758	1.4452	0.0356	0.4178	0.8570
6~12	0.0000	0.7429	0.1728	4.2274	-1.2492	3.1159	-1.1441
6~13	-2.0756	4.8998			0.0000	1.7465	
7~01	-2.1195	0.2295	0.1610	-1.5324	-0.8147		
7~02	-1.1783	-2.7619	1.4202	0.7300	-1.7170	1.1083	0.4777
7~03	2.6553	-1.2802	-0.9389	-1.5943		0.0000	
7~04	-2.5880	3.2403	0.0232	-1.0918	-0.2414	-1.3751	1.3912
7~05	-0.8598	-2.0063	1.2634	-2.2221	2.2961	0.3358	0.2241
7~06	-0.3616	0.6844	-0.7431	-0.8401	0.6184		0.0000
7~07	0.3252	-2.9140	0.2813		0.0000		
7~08	0.0000	-0.0443	0.5758	-0.1661			
7~09	0.0000	-1.6559	0.5445	-4.7861	0.9247	0.9649	0.7637
7~10		0.0000	0.3638	-0.2416	-0.2350	-0.0729	0.4600
7~11	1.2972	-1.1919	0.0250	-0.1665	-0.2551	0.1072	0.3646
7~12		0.0000				0.0000	0.1532
7~13			0.0000	-1.4229			
7~14	-1.4555	-0.1664	0.2832	-0.5361	0.6748	-2.6581	5.0062

Student t-test probabilities by run

	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	15.81%	0.00%	3.10%	0.00%	24.46%	11.22%	60.95%
4~02	72.41%	56.99%	1.89%				
4~03	3.15%	40.14%	0.01%	0.00%	12.06%	0.24%	
4~04	4.91%	2.06%	71.78%	27.32%	60.77%	42.71%	
4~05	0.24%	0.00%	0.00%	0.00%	24.04%	0.00%	
4~06	22.12%	94.77%	0.00%	0.00%	14.52%	9.31%	0.19%
4~07	31.65%	69.96%	17.79%	14.65%	41.66%	80.96%	51.31%
5~01	6.24%	94.50%	56.71%	86.47%	47.18%	14.50%	0.75%
5~02	16.64%	61.52%	1.14%	14.06%	2.06%	23.40%	0.01%
5~03	14.54%	56.36%	43.03%	25.85%	2.22%	85.83%	0.85%
5~04	22.96%	0.00%	6.70%	0.00%	0.31%	0.00%	14.61%
5~05	79.29%		0.00%	12.73%	4.61%	5.17%	19.06%
5~06	77.02%		0.00%	0.23%	21.82%	5.45%	79.05%
5~07	28.99%	31.49%	0.28%	55.01%	1.86%	0.79%	0.02%
5~08	66.14%	6.79%	8.74%	78.18%	38.58%	44.80%	3.09%
5~09	98.89%	93.11%	66.55%	72.46%			
5~10	8.23%	33.97%	0.08%	0.00%	61.54%	90.93%	64.32%
5~11	16.18%	21.55%	0.00%	1.82%	59.92%	23.73%	0.00%
5~12	10.28%	73.80%	16.33%	60.44%	20.30%	81.92%	8.03%
5~13	17.40%	3.39%	0.00%	0.59%	1.66%	0.98%	0.00%
5~14	56.67%	0.10%	9.61%	6.50%	6.89%	5.56%	48.17%
6~01	0.00%	51.87%	5.48%		0.00%		0.00%
6~02						0.00%	
6~03	0.02%	56.66%	87.56%	80.67%		0.00%	
6~04	55.08%	69.53%	9.29%	92.77%			
6~05	15.35%	1.45%	1.73%	1.90%	2.67%	23.14%	
6~06	97.22%	29.33%	33.54%	17.54%	98.67%		0.00%
6~07	22.74%	70.48%	49.67%	59.51%	25.19%	61.59%	17.58%
6~08	1.24%	24.81%		0.00%	51.19%		
6~09	0.30%						
6~10	65.20%	24.18%	3.18%	77.73%	3.18%	11.39%	2.84%
6~11	73.86%	81.93%	56.56%	15.10%	97.16%	67.69%	39.44%
6~12	100.00%	45.81%	86.30%	0.00%	21.31%	0.21%	25.43%
6~13	4.41%	0.00%			0.00%	8.77%	
7~01	3.53%	81.88%	87.24%	12.75%	41.67%		
7~02	24.07%	0.68%	15.82%	46.72%	8.98%	27.12%	63.41%
7~03	0.97%	20.46%	35.09%	11.57%		0.00%	
7~04	1.13%	0.17%	98.15%	27.89%	81.04%	17.59%	16.98%
7~05	39.14%	4.69%	20.86%	2.80%	2.36%	73.79%	82.32%
7~06	71.89%	49.68%	46.04%	40.42%	53.93%		0.00%
7~07	74.58%	0.48%	77.94%		0.00%		
7~08	0.00%	96.47%	56.61%	86.86%			
7~09	0.00%	10.04%	58.70%	0.00%	35.73%	33.70%	44.69%
7~10		0.00%	71.71%	80.97%	81.48%	94.21%	64.69%
7~11	19.76%	23.58%	98.01%	86.82%	79.94%	91.50%	71.65%
7~12		0.00%				0.00%	87.90%
7~13			0.00%	16.06%			
7~14	14.80%	86.81%	77.75%	59.31%	50.21%	0.97%	0.00%

Number of blank cells = 60 (either did not run or had less than 4 participating grades)

Number of normal cells = 184 (within 95% CI)

Number of abnormal cells = 92 (outside 95% CI)

Total number of cells = 336

Fisher F-test values by run

	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	0.8942	1.5195	1.2052	0.7018	1.0463	1.6152	0.9801
4~02	0.8809	1.7332	1.8226				
4~03	0.9474	1.2139	1.1154	0.8378	1.1883	1.2541	
4~04	1.0127	1.1731	1.3246	0.9945	0.9399	0.6983	
4~05	0.5836	1.5974	0.8582	0.9197	1.1145	1.8781	
4~06	1.0086	1.1846	1.1355	0.9764	0.7663	1.0525	0.8963
4~07	1.0223	1.1767	1.2501	0.7540	1.1744	0.8482	1.2415
5~01	1.0802	1.5660	0.6401	1.7383	0.7355	1.1989	1.0234
5~02	0.9349	1.4797	1.4462	0.8519	0.5710	0.7845	1.6253
5~03	0.7022	1.1724	0.8429	1.2167	0.8626	1.5530	0.9525
5~04	1.2921	0.9017	1.0552	0.9863	0.8407	1.1686	0.7500
5~05	2.0036		0.0000	0.6879	1.9365	0.9846	0.7220
5~06	1.4998		0.0000	1.0699	1.3091	1.3004	1.7967
5~07	1.5303	1.0391	1.0027	0.8917	0.8662	1.1411	0.9575
5~08	1.5526	1.1719	0.6989	1.4712	0.9431	0.9862	0.7258
5~09	0.8552	1.1146	0.6253	1.4428			
5~10	0.4009	0.6801	1.5474	1.3053	0.6861	0.8086	1.9900
5~11	1.8751	0.9309	1.0599	0.7741	1.4076	0.8481	0.8873
5~12	0.8558	1.0282	0.9694	0.7127	1.4897	0.8797	1.1603
5~13	1.6946	0.7711	1.7775	1.1066	1.0884	0.8812	1.2251
5~14	0.5220	2.0939	0.4230	1.2814	1.5162	0.7096	0.7358
6~01	0.0000	0.7202	3.6277		0.0000		0.0000
6~02						0.0000	
6~03	0.8313	1.1678	1.0131	0.5683		0.0000	
6~04	0.8966	1.5254	1.0660	1.1501			
6~05	2.7186	0.7708	1.1076	1.1446	0.7940	1.7376	
6~06	2.0095	1.0562	0.9473	0.7651	1.0681		0.0000
6~07	0.7567	1.6532	0.6508	0.9490	0.9655	1.1118	0.8571
6~08	2.3572	0.5259		0.0000	0.4198		
6~09	2.7921						
6~10	1.4101	1.3331	1.0419	0.5876	1.3973	0.7792	1.0679
6~11	1.3220	0.9904	0.9202	0.7866	1.1698	1.1227	0.8251
6~12	1.0053	1.9732	0.6383	0.5484	1.5045	0.7832	1.1762
6~13	0.4747	0.8499			0.0000	0.7995	
7~01	1.0984	1.2724	0.5832	1.4752	1.2861		
7~02	1.2787	1.0394	0.9275	1.8487	0.6910	0.3732	1.7295
7~03	0.4866	1.2600	2.5661	0.5009		0.0000	
7~04	2.2587	0.8346	0.7906	0.6210	2.2203	0.3544	3.2325
7~05	1.0668	1.4345	0.7368	1.2205	0.9956	0.8418	1.0990
7~06	0.8898	1.2127	1.1399	0.5565	1.6353		0.0000
7~07	1.1108	1.5366	1.1004		0.0000		
7~08	0.0000	0.8964	1.0034	0.9844			
7~09	0.0000	3.5010	0.5722	1.4822	0.5427	0.6339	0.8240
7~10		0.0000	0.2386	0.9361	0.9090	0.5361	2.2532
7~11	1.4686	1.0756	1.3521	1.1035	0.5534	1.2900	0.6635
7~12		0.0000				0.0000	1.8690
7~13			0.0000	0.7886			
7~14	0.4999	2.0708	0.6236	0.6668	2.5931	0.4865	0.8205

Fisher F-test probabilities by run

	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	18.51%	99.94%	93.96%	0.18%	64.20%	99.98%	44.33%
4~02	16.99%	100.00%	100.00%				
4~03	33.10%	93.54%	81.65%	7.21%	91.65%	95.88%	
4~04	52.36%	80.65%	95.57%	48.50%	35.32%	1.55%	
4~05	0.17%	99.16%	21.36%	34.00%	67.93%	99.28%	
4~06	51.87%	85.32%	81.00%	43.23%	3.46%	62.83%	22.87%
4~07	54.86%	83.46%	93.15%	2.60%	85.53%	14.88%	93.46%
5~01	62.40%	95.77%	4.57%	98.60%	11.91%	76.73%	53.40%
5~02	36.43%	97.09%	98.15%	17.38%	0.10%	9.88%	99.32%
5~03	6.86%	76.04%	21.78%	84.22%	22.69%	98.46%	40.43%
5~04	89.30%	25.28%	62.84%	46.56%	17.01%	78.56%	7.25%
5~05	96.14%		0.00%	12.46%	98.25%	48.59%	15.73%
5~06	84.71%		0.00%	61.89%	82.50%	80.69%	94.28%
5~07	98.24%	57.25%	51.03%	24.97%	19.40%	78.41%	39.92%
5~08	93.72%	72.08%	7.76%	95.85%	39.69%	47.39%	8.51%
5~09	27.39%	60.87%	10.61%	81.80%			
5~10	0.58%	2.90%	96.72%	86.63%	6.67%	20.32%	99.31%
5~11	99.99%	29.33%	67.01%	1.87%	99.55%	10.98%	18.96%
5~12	34.19%	53.84%	45.60%	11.86%	90.58%	33.61%	69.57%
5~13	99.51%	2.51%	100.00%	79.20%	73.48%	18.04%	92.46%
5~14	3.32%	97.77%	0.76%	83.02%	94.96%	8.84%	13.13%
6~01	0.00%	22.66%	99.97%		0.00%		0.00%
6~02						0.00%	
6~03	27.63%	63.81%	51.02%	10.16%		0.00%	
6~04	33.47%	88.23%	55.63%	64.45%			
6~05	100.00%	9.16%	67.80%	72.33%	16.58%	98.90%	
6~06	99.41%	55.90%	43.62%	24.74%	56.06%		0.00%
6~07	13.16%	98.51%	4.08%	42.54%	44.84%	61.49%	33.55%
6~08	98.31%	7.12%		0.00%	2.88%		
6~09	99.43%						
6~10	95.55%	95.85%	58.54%	0.31%	95.30%	12.54%	62.12%
6~11	81.37%	48.21%	35.71%	18.03%	73.59%	65.17%	28.23%
6~12	51.97%	99.99%	1.31%	0.17%	97.37%	12.86%	75.93%
6~13	4.62%	35.11%			0.00%	31.16%	
7~01	67.78%	84.10%	1.51%	95.23%	84.57%		
7~02	83.30%	55.68%	38.95%	97.55%	12.82%	0.12%	96.24%
7~03	1.47%	75.71%	99.55%	2.94%		0.00%	
7~04	99.56%	27.73%	22.13%	8.87%	96.20%	1.04%	99.87%
7~05	59.96%	92.71%	10.22%	78.88%	48.68%	29.03%	63.18%
7~06	37.56%	67.65%	64.49%	5.60%	88.07%		0.00%
7~07	62.80%	88.89%	61.14%		0.00%		
7~08	0.00%	34.74%	50.04%	47.28%			
7~09	0.00%	100.00%	1.11%	93.61%	1.58%	6.01%	25.40%
7~10		0.00%	0.00%	41.57%	39.15%	2.53%	99.27%
7~11	91.17%	60.50%	85.93%	62.76%	4.81%	75.52%	11.07%
7~12		0.00%				0.00%	91.21%
7~13			0.00%	28.45%			
7~14	0.34%	99.80%	3.28%	8.59%	99.68%	1.77%	27.72%

Number of blank cells = 60 (either did not run or had less than 4 participating grades)

Number of normal cells = 225 (within 95% CI)

Number of abnormal cells = 51 (outside 95% CI)

Total number of cells = 336

Mann-Whitney U-test values by run

	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	30392.5	22718.5	34541.5	28039	31152.5	25022	21125.5
4~02	24920	28815.5	29286				
4~03	28795	27187.5	27378.5	27076	30547.5	23594.5	
4~04	6417	5677.5	10654.5	11068.5	10961	9918	
4~05	4868	2514.5	3349.5	2422.5	2561.5	733.5	
4~06	9637.5	11755.5	12123	14696.5	15401	12500.5	15599
4~07	7285	9913.5	14235	16991	14922.5	12945.5	18278.5
5~01	1527.5	2022	1833	2121.5	1716	1651	1580.5
5~02	4365	4317.5	6703	8497	6589	6169	4407.5
5~03	2402	3102.5	3640	5023	4346	4769.5	3707
5~04	4882	9321.5	9429	7923	5738	3528	4842.5
5~05	504.5		0	785.5	723	777.5	571
5~06	456		0	1613	2252.5	713.5	619.5
5~07	4598	4315.5	4508	9711.5	8426	8047	6460.5
5~08	1131.5	1259	2036.5	3441.5	3089.5	2985	2394.5
5~09	1624.5	563.5	513	404			
5~10	1220.5	4293	1836	1125	2049.5	2052	1646
5~11	10967.5	24089.5	16359.5	29339	29310.5	23351	15482
5~12	589.5	1445.5	1106.5	1186	885	1055.5	960
5~13	6392.5	24569	20054	27415.5	22465.5	19001.5	14204
5~14	534.5	296.5	596.5	1277	1793	1696.5	1303
6~01	0	219.5	332		0		0
6~02						0	
6~03	387	254.5	214	230.5		0	
6~04	1698	685	207.5	339			
6~05	2979	4383	3036.5	2276	1803.5	1892.5	
6~06	1376	811	384	354.5	412		0
6~07	1878	2606	2040.5	1326.5	1083.5	664.5	447
6~08	285.5	210.5		0	208		
6~09	275.5						
6~10	4637.5	10289.5	6603.5	5595.5	4072.5	3091.5	2959
6~11	900	4125	2470.5	1579	2120.5	1331.5	550.5
6~12	9216	9203	4907	3525.5	4180.5	2991.5	3018.5
6~13	158	459.5			0	171	
7~01	4239	2851.5	2315.5	2641.5	2169		
7~02	2125	1085	1544.5	994	658.5	734.5	797.5
7~03	450.5	534	584.5	416		0	
7~04	661	596	952	419.5	250.5	138	214
7~05	2151.5	1715	2183.5	1830	1086.5	838	1034.5
7~06	450	332.5	373.5	414	295.5		0
7~07	870	378	553		0		
7~08	0	1366	1113	597			
7~09	0	1500.5	2494	1188	1196	1046.5	1086.5
7~10		0	717.5	706	787	725	568.5
7~11	1039	1319.5	1284	1079	476	405	702
7~12		0				0	219.5
7~13			0	297			
7~14	1715.5	2173.5	1810	942	528	354	222

Mann-Whitney U-test probabilities by run

	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	7.78%	0.00%	2.08%	0.00%	8.20%	13.17%	31.80%
4~02	20.07%	34.32%	0.01%				
4~03	0.20%	2.24%	0.00%	0.00%	5.39%	0.02%	
4~04	4.76%	1.14%	40.87%	14.74%	25.85%	20.52%	
4~05	0.00%	0.00%	0.00%	0.00%	17.32%	0.00%	
4~06	8.00%	48.16%	0.00%	0.00%	2.46%	5.23%	0.11%
4~07	15.30%	35.83%	5.08%	10.74%	19.09%	49.90%	19.30%
5~01	1.31%	42.52%	42.63%	40.29%	34.56%	2.40%	0.14%
5~02	6.78%	42.66%	0.47%	5.07%	4.30%	16.68%	0.07%
5~03	5.27%	20.42%	25.73%	11.97%	1.56%	47.59%	0.16%
5~04	16.82%	0.00%	2.67%	0.00%	0.28%	0.00%	9.94%
5~05	44.28%		0.00%	11.38%	0.71%	2.17%	7.13%
5~06	43.53%		0.00%	0.28%	15.02%	2.61%	32.16%
5~07	19.99%	14.49%	0.08%	22.98%	0.21%	0.14%	0.01%
5~08	30.56%	3.62%	4.94%	45.15%	15.60%	13.68%	7.73%
5~09	48.01%	42.10%	43.26%	26.06%			
5~10	5.83%	15.37%	0.03%	0.00%	48.30%	38.23%	35.51%
5~11	26.47%	7.47%	0.00%	0.01%	43.29%	5.71%	0.00%
5~12	4.53%	30.11%	5.54%	33.55%	9.05%	43.25%	4.38%
5~13	46.06%	19.04%	0.02%	0.02%	0.30%	0.13%	0.00%
5~14	23.91%	0.09%	6.40%	0.59%	5.74%	2.85%	24.60%
6~01	0.00%	16.64%	6.64%		0.00%		0.00%
6~02						0.00%	
6~03	0.01%	26.05%	34.41%	24.34%		0.00%	
6~04	28.18%	24.48%	1.54%	38.30%			
6~05	43.85%	0.28%	0.40%	0.34%	0.18%	1.27%	
6~06	25.52%	14.83%	19.52%	12.91%	50.00%		0.00%
6~07	5.35%	16.48%	27.58%	25.81%	13.67%	30.86%	8.18%
6~08	0.24%	16.86%		0.00%	17.01%		
6~09	0.06%						
6~10	21.63%	24.18%	1.41%	17.43%	0.08%	5.35%	1.68%
6~11	34.07%	38.45%	24.33%	8.04%	39.84%	25.28%	27.39%
6~12	49.97%	6.76%	48.57%	0.01%	13.55%	0.20%	16.62%
6~13	2.55%	0.00%			0.00%	2.48%	
7~01	2.12%	47.80%	30.24%	15.43%	23.60%		
7~02	10.37%	0.53%	12.16%	11.08%	5.28%	38.31%	18.48%
7~03	0.68%	12.12%	20.42%	4.02%		0.00%	
7~04	0.52%	0.06%	44.98%	10.35%	49.09%	0.20%	0.13%
7~05	17.66%	2.09%	10.77%	3.31%	1.11%	44.00%	30.75%
7~06	33.59%	34.00%	21.60%	33.12%	46.80%		0.00%
7~07	34.45%	0.22%	32.10%		0.00%		
7~08	0.00%	40.51%	42.40%	38.25%			
7~09	0.00%	10.97%	34.86%	0.00%	25.21%	17.57%	31.78%
7~10		0.00%	44.69%	40.07%	39.42%	26.13%	9.27%
7~11	9.37%	4.42%	22.49%	31.74%	34.79%	44.74%	42.77%
7~12		0.00%				0.00%	50.00%
7~13			0.00%	11.41%			
7~14	3.49%	43.91%	24.22%	23.15%	16.31%	0.03%	0.00%

Number of blank cells = 60 (either did not run or had less than 4 participating grades)

Number of normal cells = 165 (within 95% CI)

Number of abnormal cells = 111 (outside 95% CI)

Total number of cells = 336

Levene test values by run

	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	0.2177	12.1824	0.7117	9.2033	0.4621	6.7522	0.1641
4~02	1.0088	13.3334	15.3362				
4~03	0.0375	3.3662	0.0760	2.7649	1.7418	4.7342	
4~04	0.0811	0.3468	4.2644	0.0816	0.0017	8.1825	
4~05	5.9641	5.3049	2.7616	0.0284	0.5592	5.7138	
4~06	0.0006	0.9660	0.6405	0.0254	5.5962	0.2349	0.2255
4~07	0.0116	0.4613	2.3298	2.6029	1.4381	3.8084	7.5615
5~01	0.0213	1.5304	1.0240	4.6138	2.4081	0.5310	0.0999
5~02	0.0698	3.0712	2.6005	0.0058	8.5848	1.1902	2.0246
5~03	1.4937	0.6115	1.1842	1.1258	0.3332	6.1836	0.2890
5~04	1.7843	0.6144	0.0195	0.0045	2.5535	2.0306	4.4218
5~05	3.7300		0.0000	0.3260	4.2307	0.2829	0.0683
5~06	1.0749		0.0000	0.1163	0.8019	1.7934	4.9440
5~07	8.6446	0.1816	0.0129	0.1846	0.7884	0.2832	0.0141
5~08	3.5849	0.0212	3.2576	6.8441	0.0006	0.0030	6.1682
5~09	0.8767	0.2748	1.6783	0.8490			
5~10	4.3792	2.7240	3.3266	0.4184	2.9381	0.0042	2.7527
5~11	13.8949	0.0308	2.1455	5.4208	3.2970	3.5283	0.0028
5~12	0.3405	0.0001	0.0598	4.3357	3.0660	0.0897	0.5937
5~13	8.1984	4.4673	17.0991	0.3777	0.1896	0.8330	0.3109
5~14	2.7401	3.6492	5.4235	0.4775	3.4820	2.2330	1.3383
6~01	0.0000	0.8806	9.5591		0.0000		0.0000
6~02						0.0000	
6~03	0.3624	0.2031	0.0861	0.5122		0.0000	
6~04	0.0030	3.0184	0.2517	0.1696			
6~05	26.2001	1.4729	0.0852	0.0017	0.5228	0.7527	
6~06	4.1664	0.0364	0.0014	0.6047	0.0016		0.0000
6~07	0.6479	2.9416	4.7098	0.0075	0.1209	0.2369	0.2802
6~08	3.7099	0.5115		0.0000	1.2005		
6~09	6.9020						
6~10	3.9502	1.1539	0.0813	2.0287	1.4495	5.6991	1.7692
6~11	1.7113	0.0018	0.3166	1.9568	0.4485	0.4329	0.0031
6~12	0.0000	8.5389	2.5658	13.3843	4.5719	1.4047	1.1325
6~13	4.3360	0.0350			0.0000	0.7208	
7~01	0.0077	1.0461	3.8479	1.6312	2.2383		
7~02	0.5706	0.2527	0.1855	2.1999	1.4122	9.7853	4.8982
7~03	4.4970	0.7933	5.1386	2.1099		0.0000	
7~04	4.5780	0.8421	1.5560	1.8064	5.4771	1.8029	2.4370
7~05	0.4158	2.0698	2.0467	0.3628	0.0563	0.5823	0.1684
7~06	0.7379	0.4042	0.0930	2.5043	0.9773		0.0000
7~07	0.0776	0.5627	0.0252		0.0000		
7~08	0.0000	0.0924	0.0000	0.0426			
7~09	0.0000	10.8076	5.8079	4.0443	9.1600	2.0449	0.8905
7~10		0.0000	7.8042	0.0657	0.0311	2.0493	2.4378
7~11	2.2790	0.2588	0.8965	0.0168	2.4165	0.5666	0.4183
7~12		0.0000				0.0000	2.7587
7~13			0.0000	0.6377			
7~14	5.6440	9.3507	4.8636	1.0997	4.0233	0.2790	0.2916

Levene test probabilities by run

	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	64.10%	0.05%	39.93%	0.25%	49.69%	0.97%	68.57%
4~02	31.57%	0.03%	0.01%				
4~03	84.65%	6.72%	78.29%	9.69%	18.75%	3.01%	
4~04	77.61%	55.65%	3.98%	77.54%	96.68%	0.45%	
4~05	1.53%	2.22%	9.80%	86.64%	45.58%	1.83%	
4~06	98.09%	32.65%	42.40%	87.35%	1.85%	62.83%	63.51%
4~07	91.42%	49.76%	12.78%	10.75%	23.13%	5.19%	0.62%
5~01	88.42%	21.83%	31.36%	3.36%	12.34%	46.75%	75.24%
5~02	79.19%	8.13%	10.81%	93.91%	0.37%	27.64%	15.62%
5~03	22.36%	43.54%	27.80%	28.99%	56.44%	1.37%	59.15%
5~04	18.29%	43.37%	88.91%	94.67%	11.13%	15.57%	3.67%
5~05	5.78%		0.00%	56.95%	4.26%	59.61%	79.45%
5~06	30.39%		0.00%	73.35%	37.18%	18.40%	2.92%
5~07	0.37%	67.05%	90.97%	66.78%	37.53%	59.50%	90.54%
5~08	6.13%	88.45%	7.32%	0.97%	98.11%	95.61%	1.41%
5~09	35.10%	60.18%	19.98%	36.05%			
5~10	3.81%	10.04%	7.02%	51.88%	8.90%	94.86%	9.97%
5~11	0.02%	86.07%	14.37%	2.03%	7.00%	6.10%	95.79%
5~12	56.11%	99.09%	80.73%	3.99%	8.34%	76.52%	44.29%
5~13	0.45%	3.51%	0.00%	53.91%	66.35%	36.19%	57.74%
5~14	10.25%	6.06%	2.25%	49.09%	6.43%	13.76%	24.99%
6~01	0.00%	35.32%	0.31%		0.00%		0.00%
6~02						0.00%	
6~03	54.88%	65.43%	77.07%	47.80%		0.00%	
6~04	95.62%	8.62%	61.81%	68.22%			
6~05	0.00%	22.63%	77.06%	96.70%	47.09%	38.71%	
6~06	4.37%	84.90%	97.00%	44.00%	96.79%		0.00%
6~07	42.23%	8.84%	3.18%	93.12%	72.88%	62.79%	59.84%
6~08	5.85%	47.83%		0.00%	27.93%		
6~09	1.07%						
6~10	4.82%	28.36%	77.57%	15.58%	23.00%	1.81%	18.53%
6~11	19.35%	96.62%	57.45%	16.44%	50.43%	51.20%	95.60%
6~12	100.00%	0.37%	11.08%	0.03%	3.38%	23.75%	28.88%
6~13	4.34%	85.18%			0.00%	40.05%	
7~01	93.02%	30.80%	5.18%	20.35%	13.70%		
7~02	45.13%	61.62%	66.74%	14.12%	23.82%	0.25%	2.96%
7~03	3.73%	37.61%	2.64%	15.12%		0.00%	
7~04	3.52%	36.13%	21.56%	18.36%	2.40%	18.61%	12.42%
7~05	52.01%	15.26%	15.48%	54.80%	81.29%	44.76%	68.25%
7~06	39.38%	52.77%	76.15%	11.89%	32.79%		0.00%
7~07	78.12%	45.57%	87.44%		0.00%		
7~08	0.00%	76.17%	99.50%	83.71%			
7~09	0.00%	0.13%	1.72%	4.63%	0.31%	15.60%	34.78%
7~10		0.00%	0.66%	79.84%	86.06%	15.62%	12.28%
7~11	13.43%	61.20%	34.59%	89.72%	12.49%	45.47%	51.98%
7~12		0.00%				0.00%	10.45%
7~13			0.00%	42.81%			
7~14	1.90%	0.27%	2.93%	29.70%	4.89%	59.90%	59.08%

Number of blank cells = 60 (either did not run or had less than 4 participating grades)

Number of normal cells = 201 (within 95% CI)

Number of abnormal cells = 75 (outside 95% CI)

Total number of cells = 336

Brown-Forsythe test values by run

	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	0.2196	12.4068	0.6922	7.8750	0.2530	7.0746	0.0773
4~02	0.4864	9.3204	12.0719				
4~03	0.1377	1.3998	0.0252	2.9954	1.9189	4.1408	
4~04	0.0045	0.3985	3.8257	0.0307	0.0028	8.1171	
4~05	2.1288	1.8998	1.7898	0.0777	0.4922	5.6790	
4~06	0.0008	1.0680	0.7532	0.0010	4.5680	0.2275	0.2259
4~07	0.0164	0.4209	2.3320	2.5909	1.4346	3.0762	6.4215
5~01	0.0697	1.3534	0.9710	4.8339	1.9995	0.4128	0.0380
5~02	0.1350	2.7739	2.6306	0.0014	8.4004	1.1666	2.0072
5~03	1.3075	0.4308	1.1866	1.1575	0.3650	6.2084	0.1973
5~04	1.7397	0.4970	0.0731	0.0002	1.9937	1.3431	4.1269
5~05	3.4565		0.0000	0.2650	3.8137	0.2934	0.0517
5~06	0.7222		0.0000	0.1305	0.9616	1.1852	4.8700
5~07	8.6138	0.1867	0.0003	0.2858	0.5313	0.1596	0.0162
5~08	3.4042	0.0091	2.9767	6.5794	0.0045	0.0093	4.7889
5~09	0.6632	0.1912	1.0300	0.4772			
5~10	3.9785	3.0017	3.5697	0.6201	3.3023	0.0042	2.8006
5~11	12.5054	0.0338	2.1985	5.2710	2.7666	4.2505	0.0000
5~12	0.3420	0.0016	0.0678	4.2854	3.0370	0.0161	0.2793
5~13	5.1731	2.4191	11.8346	0.3359	0.1892	0.8773	0.3367
5~14	2.8949	3.6330	2.1622	0.0231	2.7731	2.4202	1.2231
6~01	0.0000	0.4988	5.9394		0.0000		0.0000
6~02						0.0000	
6~03	0.4030	0.2026	0.0432	0.2507		0.0000	
6~04	0.0909	2.5384	0.3434	0.1266			
6~05	14.7758	1.6530	0.1217	0.1677	0.2135	0.7781	
6~06	3.0546	0.0370	0.0154	0.4889	0.0008		0.0000
6~07	0.5931	2.4243	4.8084	0.0118	0.0712	0.3112	0.2968
6~08	3.4996	0.5577		0.0000	1.2876		
6~09	5.7829						
6~10	3.2068	1.0917	0.0207	1.1451	0.8093	5.5083	1.7238
6~11	1.6924	0.0281	0.3983	1.2941	0.3047	0.2685	0.0002
6~12	0.0000	4.1545	2.4622	11.2204	3.7332	1.3731	1.0655
6~13	3.0497	0.0772			0.0000	0.1748	
7~01	0.0359	1.1728	2.6460	0.9314	1.9393		
7~02	0.5734	0.2384	0.2000	2.1608	1.3617	6.0787	3.1240
7~03	4.1935	0.7244	5.0828	1.9854		0.0000	
7~04	3.6454	0.7309	1.4779	1.8264	5.4976	0.6331	0.7767
7~05	0.4828	1.8236	1.9022	0.3137	0.0724	0.2941	0.0440
7~06	0.6539	0.3493	0.1374	2.3034	0.9906		0.0000
7~07	0.1044	0.5615	0.0025		0.0000		
7~08	0.0000	0.0992	0.0015	0.0390			
7~09	0.0000	10.7638	5.6706	3.9036	9.0648	2.0178	0.4815
7~10		0.0000	6.8685	0.0745	0.0075	1.8740	2.1230
7~11	2.2291	0.2632	0.9982	0.0333	2.2475	0.5011	0.3099
7~12		0.0000				0.0000	2.5957
7~13			0.0000	0.6543			
7~14	5.6006	9.2893	4.4366	1.1185	3.8657	0.1488	0.3816

Brown-Forsythe test probabilities by run

	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	63.96%	0.05%	40.58%	0.52%	61.52%	0.81%	78.12%
4~02	48.59%	0.24%	0.06%				
4~03	71.07%	23.73%	87.38%	8.41%	16.66%	4.24%	
4~04	94.67%	52.85%	5.14%	86.11%	95.82%	0.47%	
4~05	14.59%	16.95%	18.23%	78.07%	48.40%	1.86%	
4~06	97.68%	30.22%	38.60%	97.49%	3.32%	63.37%	63.48%
4~07	89.82%	51.70%	12.76%	10.83%	23.18%	8.04%	1.17%
5~01	79.23%	24.69%	32.64%	2.97%	16.00%	52.17%	84.56%
5~02	71.37%	9.75%	10.61%	97.01%	0.41%	28.12%	15.80%
5~03	25.47%	51.25%	27.75%	28.32%	54.64%	1.36%	65.74%
5~04	18.84%	48.13%	78.71%	98.76%	15.92%	24.78%	4.35%
5~05	6.75%		0.00%	60.80%	5.39%	58.94%	82.07%
5~06	39.87%		0.00%	71.84%	32.82%	27.93%	3.04%
5~07	0.37%	66.62%	98.73%	59.34%	46.67%	68.98%	89.88%
5~08	6.81%	92.40%	8.67%	1.12%	94.64%	92.33%	3.02%
5~09	41.71%	66.33%	31.40%	49.23%			
5~10	4.80%	8.47%	6.08%	43.23%	7.15%	94.83%	9.69%
5~11	0.05%	85.43%	13.88%	2.21%	9.69%	3.98%	99.57%
5~12	56.03%	96.86%	79.50%	4.11%	8.48%	89.94%	59.84%
5~13	2.36%	12.06%	0.06%	56.25%	66.38%	34.95%	56.21%
5~14	9.35%	6.11%	14.55%	87.95%	9.83%	12.23%	27.12%
6~01	0.00%	48.38%	1.79%		0.00%		0.00%
6~02						0.00%	
6~03	52.73%	65.47%	83.63%	61.91%		0.00%	
6~04	76.35%	11.51%	56.06%	72.34%			
6~05	0.02%	20.00%	72.76%	68.28%	64.47%	37.93%	
6~06	8.34%	84.80%	90.17%	48.72%	97.80%		0.00%
6~07	44.26%	12.16%	3.01%	91.39%	79.01%	57.86%	58.77%
6~08	6.60%	45.93%		0.00%	26.28%		
6~09	1.90%						
6~10	7.48%	29.70%	88.57%	28.58%	36.94%	2.01%	19.10%
6~11	19.60%	86.71%	52.89%	25.76%	58.19%	60.54%	99.02%
6~12	100.00%	4.24%	11.82%	0.10%	5.48%	24.28%	30.35%
6~13	8.81%	78.16%			0.00%	67.79%	
7~01	85.00%	28.05%	10.61%	33.61%	16.60%		
7~02	45.02%	62.63%	65.56%	14.48%	24.67%	1.59%	8.07%
7~03	4.41%	39.76%	2.72%	16.36%		0.00%	
7~04	5.96%	39.49%	22.74%	18.12%	2.37%	43.04%	38.20%
7~05	48.83%	17.92%	17.00%	57.64%	78.85%	58.91%	83.42%
7~06	42.19%	55.71%	71.22%	13.44%	32.47%		0.00%
7~07	74.75%	45.62%	96.04%		0.00%		
7~08	0.00%	75.35%	96.90%	84.40%			
7~09	0.00%	0.14%	1.86%	5.03%	0.33%	15.87%	48.95%
7~10		0.00%	1.06%	78.56%	93.13%	17.49%	14.94%
7~11	13.86%	60.89%	32.01%	85.56%	13.87%	48.19%	57.94%
7~12		0.00%				0.00%	11.50%
7~13			0.00%	42.22%			
7~14	1.95%	0.28%	3.72%	29.30%	5.34%	70.08%	53.87%

Number of blank cells = 60 (either did not run or had less than 4 participating grades)

Number of normal cells = 214 (within 95% CI)

Number of abnormal cells = 62 (outside 95% CI)

Total number of cells = 336

Positional scores for the LILMs for each run

	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	1.0144	0.9282	1.0872	1.1183	1.0793	1.0601	1.0076	0.9451
4~02	0.9650	0.9249	0.9313	0.8456				
4~03	1.1113	1.1618	1.1397	0.9937	0.9451	0.9914	0.8609	
4~04	0.8832	0.8918	0.8083	0.8368	0.8260	0.8236	0.8592	
4~05	1.1620	1.2947	0.9739	1.2123	1.0522	0.9873	1.4071	
4~06	0.9791	0.9895	1.0065	1.1936	1.1398	1.1660	1.2399	1.0813
4~07	0.9187	0.9393	0.9757	0.9332	0.9713	0.9537	0.9841	0.9767
5~01	1.0322	0.9587	0.9126	0.9216	0.8884	0.8926	0.8691	0.8363
5~02	1.1232	0.9587	0.9990	0.8922	0.9923	0.9746	0.9489	0.9799
5~03	0.7815	0.9063	0.8701	0.9819	1.0343	0.9273	0.9809	0.9571
5~04	1.0398	0.8985	1.0901	1.1753	0.9944	1.1631	0.9725	0.9490
5~05	1.0906	1.0396	0.9069	0.9722	0.9461	0.8507	0.9120	0.9574
5~06	1.1008	1.0898	0.9845	1.0458	1.0368	1.0413	1.1152	1.0257
5~07	0.9672	0.9994	1.0111	0.9522	0.9723	0.9244	1.0304	1.0743
5~08	0.9550	0.9358	0.9767	0.9401	0.9423	0.9519	0.9048	0.9780
5~09	1.0350	0.9758	0.9856	0.9150	0.9499	0.7848	0.9915	
5~10	1.0597	1.1599	1.1380	1.1831	1.0443	1.0295	1.1528	1.0592
5~11	1.0061	1.0641	0.9961	0.9050	0.9679	1.0077	0.9997	1.0510
5~12	1.0230	0.8340	0.8936	0.9068	0.9806	0.9061	0.9551	0.9884
5~13	0.9524	1.0526	1.0022	1.1101	1.0794	1.0685	1.0498	1.0332
5~14	0.9762	1.0604	1.1988	1.1203	1.0655	1.1853	1.0865	0.9936
6~01	1.1303	0.9734	1.1077	1.0107				
6~02	0.8983	0.9607	1.1159	0.8627	1.0061	0.9513		0.9387
6~03	0.7737	1.0154	1.0206	0.9002	0.9321	1.0161	1.0856	1.0690
6~04	1.1978	0.9943	0.9334	0.9190	0.8506	0.8716	0.8774	1.0430
6~05	0.8288	0.9276	1.0628	0.9143	1.0393	1.2119	1.1858	0.9506
6~06	1.0393	1.0000	1.0294	1.0374	0.9866	0.9959	1.1316	1.0024
6~07	1.0148	0.9828	0.9682	0.9850	1.0672	0.9659	0.9655	0.9956
6~08	0.9760	1.0442	1.0048	0.9960	1.0419	1.0891	1.0543	1.1731
6~09	1.0131	1.0442	0.9154	1.1499	0.9773	0.9378	0.9301	0.8843
6~10	1.0985	0.9900	1.0355	1.1017	1.1311	0.9965	0.9712	1.0376
6~11	1.0041	1.0415	0.9440	0.9941	0.8639	0.9456	0.9676	0.9390
6~12	1.5759	1.0562	0.9863	0.9861	0.9045	0.9335	0.8643	0.8596
6~13	0.8847	1.0273	0.6492	0.9894				
7~01	0.8779	0.9307	0.9196	0.8845	0.9633	0.9916		
7~02	0.9337	0.9331	1.0060	0.9644	0.8924	0.9445	0.9409	0.9667
7~03	1.0374	0.9322	1.1357	0.9769	0.9879	1.1886	0.8765	1.0509
7~04	1.1844	1.1651	0.9899	0.9631	1.0066	1.0291	1.1895	1.1244
7~05	0.9583	0.9553	0.9507	0.9421	0.9760	0.9153	0.8781	0.9049
7~06	1.0615	1.0291	0.9968	0.9645	1.0231	0.9987	1.1145	0.9418
7~07	0.8458	0.8322	0.9707	0.9492	0.9552	0.9116	1.0849	1.0025
7~08	0.9402	0.9903	0.9667	1.0231	0.9416	1.0112	0.8936	1.0573
7~09	0.9695	1.0587	1.0263	1.0523	1.1290	1.0521	1.0308	1.0947
7~10	1.0391	1.0820	1.0099	1.0302	0.9816	0.9936	1.0267	1.0448
7~11	0.9759	0.9000	0.9843	1.0142	0.9747	0.9703	0.9000	0.9570
7~12	1.0335	1.1600	0.9772	1.0267	1.0438	0.9002	1.0447	1.0190
7~13	1.0692	1.1231	0.9626	1.0623	1.0539	1.2486	1.1042	1.1062
7~14	1.1431	1.0386	1.0360	1.0526	1.0067	0.9808	1.0962	0.8930

Learning intervention location metrics (LILM) for each run

	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	-0.0862	0.1590	0.0311	-0.0390	-0.0193	-0.0525	-0.0625
4~02	-0.0401	0.0064	-0.0857				
4~03	0.0506	-0.0221	-0.1460	-0.0486	0.0464	-0.1305	
4~04	0.0086	-0.0835	0.0285	-0.0108	-0.0024	0.0356	
4~05	0.1327	-0.3208	0.2384	-0.1601	-0.0649	0.4198	
4~06	0.0104	0.0170	0.1871	-0.0537	0.0261	0.0739	-0.1586
4~07	0.0205	0.0365	-0.0426	0.0382	-0.0176	0.0304	-0.0074
5~01	-0.0736	-0.0460	0.0090	-0.0332	0.0042	-0.0235	-0.0328
5~02	-0.1645	0.0404	-0.1069	0.1002	-0.0178	-0.0257	0.0310
5~03	0.1248	-0.0362	0.1117	0.0525	-0.1070	0.0536	-0.0238
5~04	-0.1413	0.1916	0.0852	-0.1809	0.1687	-0.1906	-0.0235
5~05	-0.0510	-0.1327	0.0654	-0.0261	-0.0954	0.0612	0.0454
5~06	-0.0111	-0.1052	0.0612	-0.0089	0.0045	0.0740	-0.0895
5~07	0.0322	0.0117	-0.0589	0.0201	-0.0478	0.1060	0.0438
5~08	-0.0192	0.0409	-0.0366	0.0022	0.0097	-0.0471	0.0732
5~09	-0.0592	0.0098	-0.0706	0.0349	-0.1651	0.2066	
5~10	0.1002	-0.0218	0.0451	-0.1389	-0.0147	0.1233	-0.0936
5~11	0.0580	-0.0680	-0.0911	0.0629	0.0398	-0.0080	0.0513
5~12	-0.1890	0.0596	0.0131	0.0739	-0.0746	0.0490	0.0333
5~13	0.1003	-0.0504	0.1079	-0.0307	-0.0109	-0.0186	-0.0166
5~14	0.0842	0.1384	-0.0785	-0.0548	0.1199	-0.0989	-0.0929
6~01	-0.1569	0.1343	-0.0970				
6~02	0.0624	0.1552	-0.2532	0.1435	-0.0548		-0.1674
6~03	0.2417	0.0052	-0.1204	0.0319	0.0841	0.0695	-0.0167
6~04	-0.2035	-0.0610	-0.0143	-0.0685	0.0210	0.0059	0.1655
6~05	0.0988	0.1352	-0.1485	0.1250	0.1726	-0.0261	-0.2352
6~06	-0.0393	0.0294	0.0080	-0.0509	0.0094	0.1357	-0.1292
6~07	-0.0320	-0.0146	0.0168	0.0822	-0.1013	-0.0003	0.0300
6~08	0.0682	-0.0394	-0.0088	0.0459	0.0472	-0.0348	0.1188
6~09	0.0312	-0.1288	0.2345	-0.1726	-0.0395	-0.0077	-0.0458
6~10	-0.1085	0.0456	0.0661	0.0294	-0.1346	-0.0253	0.0664
6~11	0.0373	-0.0974	0.0501	-0.1302	0.0816	0.0220	-0.0286
6~12	-0.5198	-0.0699	-0.0001	-0.0816	0.0290	-0.0692	-0.0047
6~13	0.1426	-0.3782	0.3402				
7~01	0.0528	-0.0110	-0.0351	0.0787	0.0283		
7~02	-0.0006	0.0729	-0.0416	-0.0719	0.0521	-0.0037	0.0258
7~03	-0.1052	0.2035	-0.1588	0.0110	0.2007	-0.3121	0.1744
7~04	-0.0194	-0.1752	-0.0267	0.0435	0.0225	0.1604	-0.0650
7~05	-0.0030	-0.0046	-0.0086	0.0339	-0.0606	-0.0372	0.0267
7~06	-0.0323	-0.0324	-0.0322	0.0586	-0.0245	0.1158	-0.1728
7~07	-0.0136	0.1385	-0.0215	0.0061	-0.0436	0.1733	-0.0825
7~08	0.0501	-0.0237	0.0565	-0.0815	0.0696	-0.1177	0.1637
7~09	0.0892	-0.0324	0.0260	0.0767	-0.0769	-0.0213	0.0640
7~10	0.0429	-0.0721	0.0203	-0.0486	0.0120	0.0331	0.0181
7~11	-0.0759	0.0843	0.0299	-0.0394	-0.0045	-0.0703	0.0570
7~12	0.1265	-0.1828	0.0495	0.0171	-0.1436	0.1445	-0.0257
7~13	0.0539	-0.1605	0.0997	-0.0084	0.1947	-0.1444	0.0019
7~14	-0.1046	-0.0026	0.0166	-0.0459	-0.0259	0.1154	-0.2032

Number of blank cells = 20 (either did not run or had less than 4 participating grades)

Number of normal cells = 300 (within 95% CI)

Number of abnormal cells = 16 (outside 95% CI)

Total number of cells = 336

Positional scores for the LISM_{Mean} values for each run

	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	1.1233	1.1257	1.0267	0.9864	1.0518	1.0138	0.8887	0.8720
4~02	0.8634	0.9145	0.8511	0.7174				
4~03	0.9299	0.7716	0.8232	0.9615	0.9011	0.8272	0.7346	
4~04	1.1241	1.2459	1.3972	1.2092	1.0395	1.0657	1.2969	
4~05	0.8955	1.1684	1.0737	1.2164	1.0818	1.1113	0.8666	
4~06	0.8983	0.8665	0.8875	0.9524	0.8474	0.9885	0.9406	1.0344
4~07	1.1572	1.1761	1.2048	1.2131	1.1770	1.1577	1.2485	1.0889
5~01	1.1774	1.2217	1.1427	1.2505	1.0261	0.9081	1.0430	1.0830
5~02	0.9036	1.2217	1.1126	0.9929	0.9376	1.1446	1.3553	1.1662
5~03	1.0284	1.2845	1.3181	1.3004	1.2728	1.3133	0.9961	0.9693
5~04	1.0068	1.1902	1.1977	1.4079	1.3732	1.4822	1.1962	1.5463
5~05	0.8648	1.0335	0.9569	1.1559	1.2248	0.8974	1.0423	0.9705
5~06	0.8831	1.2469	1.3712	1.3039	1.1497	1.1410	1.0086	0.7185
5~07	1.1417	0.9072	1.0674	0.9327	1.0865	1.1477	1.1058	1.0271
5~08	1.2307	1.1356	1.1872	1.3164	1.1687	1.2317	1.2219	1.3432
5~09	0.8745	1.2367	0.9722	1.6021	1.3228	1.1837	1.0861	
5~10	0.5514	0.9028	0.9787	0.9478	0.7278	0.9456	0.9609	0.7234
5~11	1.1564	0.7045	0.7843	0.6468	0.8391	0.6909	0.8242	0.7667
5~12	1.0034	1.2659	0.9496	1.1905	1.3783	1.0958	1.1917	1.1723
5~13	1.0442	0.6813	0.6968	0.5318	0.5203	0.5395	0.5738	0.5272
5~14	0.7011	1.3482	0.8249	1.3039	1.2654	1.0181	1.1729	1.2935
6~01	0.9333	0.8733	1.1829	0.4123				
6~02	1.2470	2.0731	2.2409	2.0639	1.9625	1.7919		0.9652
6~03	0.9578	1.1474	0.9436	1.4214	1.2723	1.4752	1.2359	1.1146
6~04	1.1380	1.5552	1.0881	1.0488	0.9047	1.2882	1.0131	0.8169
6~05	1.5550	0.9999	1.3193	1.4508	1.1552	1.1206	0.9439	0.8335
6~06	1.0405	0.9195	1.0721	1.0339	1.2678	1.1975	0.9420	0.8391
6~07	0.7131	1.0621	0.8861	1.0699	0.9147	0.8689	0.8886	0.9794
6~08	0.8267	0.6605	1.0006	0.6396	0.4254	0.7592	1.3445	0.9583
6~09	0.9894	0.6605	0.9919	0.7340	1.4368	1.6979	1.2329	1.6713
6~10	0.7682	0.8855	0.7317	0.7407	0.9058	0.7185	0.9744	1.0226
6~11	0.9104	0.9019	1.3082	0.8490	1.1027	1.0748	0.9372	0.9249
6~12	0.4009	1.0305	0.8098	0.7837	0.8868	0.9370	1.0202	1.1125
6~13	1.0715	1.1451	0.9910	2.4257				
7~01	1.1073	1.1005	1.1843	1.3257	0.9713	1.0471		
7~02	1.2221	1.0284	1.4297	1.2957	0.8048	0.8915	0.9504	0.9650
7~03	0.6912	1.1101	0.9025	0.7032	0.6785	0.8145	0.3289	0.3910
7~04	1.2230	1.1276	1.5411	1.0850	1.9117	1.1779	1.6338	0.8226
7~05	0.8474	1.1056	1.1725	1.0850	1.1034	0.8263	0.6734	0.8770
7~06	0.3644	0.3572	0.4625	0.5180	0.5962	0.5166	0.3842	0.7035
7~07	1.1971	1.1275	1.1409	0.7985	0.6115	1.3956	1.3161	1.0979
7~08	2.4505	0.9804	1.0554	0.8769	1.2769	0.8924	0.9966	1.2589
7~09	0.7153	0.9843	0.6295	0.7845	0.4555	0.7777	0.7230	1.0253
7~10	0.9178	0.6512	0.7087	1.2065	1.3257	1.4082	1.4622	1.1362
7~11	1.4136	1.1465	1.0988	0.9553	1.0938	1.3989	1.2451	1.1767
7~12	1.2103	0.8411	1.1131	1.3824	1.5456	0.8744	2.9697	1.1420
7~13	1.0303	0.9760	1.3825	0.9260	1.1576	1.7160	0.8143	0.7034
7~14	0.7348	1.1419	0.8818	1.0558	1.1614	0.9671	1.0187	1.1675

Learning intervention spread metrics – mean for each run

	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	0.0024	-0.0990	-0.0404	0.0654	-0.0380	-0.1251	-0.0167
4~02	0.0512	-0.0634	-0.1337				
4~03	-0.1582	0.0515	0.1383	-0.0604	-0.0739	-0.0926	
4~04	0.1218	0.1513	-0.1880	-0.1696	0.0262	0.2312	
4~05	0.2729	-0.0947	0.1427	-0.1346	0.0295	-0.2447	
4~06	-0.0317	0.0209	0.0649	-0.1050	0.1411	-0.0479	0.0937
4~07	0.0189	0.0287	0.0083	-0.0361	-0.0193	0.0908	-0.1596
5~01	0.0443	-0.0791	0.1079	-0.2244	-0.1180	0.1349	0.0400
5~02	0.3181	-0.1092	-0.1197	-0.0553	0.2070	0.2107	-0.1891
5~03	0.2561	0.0336	-0.0177	-0.0277	0.0405	-0.3172	-0.0268
5~04	0.1835	0.0075	0.2102	-0.0348	0.1090	-0.2860	0.3501
5~05	0.1687	-0.0766	0.1989	0.0689	-0.3274	0.1449	-0.0718
5~06	0.3637	0.1244	-0.0673	-0.1542	-0.0087	-0.1324	-0.2901
5~07	-0.2345	0.1601	-0.1346	0.1537	0.0612	-0.0419	-0.0787
5~08	-0.0951	0.0516	0.1292	-0.1477	0.0630	-0.0098	0.1213
5~09	0.3623	-0.2645	0.6299	-0.2793	-0.1391	-0.0977	
5~10	0.3515	0.0759	-0.0309	-0.2200	0.2177	0.0154	-0.2376
5~11	-0.4519	0.0798	-0.1375	0.1923	-0.1481	0.1332	-0.0574
5~12	0.2625	-0.3163	0.2409	0.1878	-0.2825	0.0960	-0.0195
5~13	-0.3630	0.0155	-0.1650	-0.0115	0.0192	0.0344	-0.0466
5~14	0.6471	-0.5233	0.4791	-0.0385	-0.2473	0.1548	0.1206
6~01	-0.0600	0.3096	-0.7706				
6~02	0.8261	0.1679	-0.1770	-0.1014	-0.1706		-1.3014
6~03	0.1896	-0.2039	0.4778	-0.1491	0.2028	-0.2393	-0.1213
6~04	0.4172	-0.4671	-0.0393	-0.1441	0.3835	-0.2751	-0.1962
6~05	-0.5551	0.3194	0.1315	-0.2956	-0.0346	-0.1767	-0.1104
6~06	-0.1210	0.1526	-0.0382	0.2339	-0.0703	-0.2555	-0.1029
6~07	0.3490	-0.1760	0.1838	-0.1552	-0.0458	0.0197	0.0909
6~08	-0.1663	0.3401	-0.3610	-0.2141	0.3338	0.5852	-0.3862
6~09	-0.3289	0.3315	-0.2579	0.7028	0.2611	-0.4650	0.4384
6~10	0.1174	-0.1538	0.0090	0.1651	-0.1872	0.2558	0.0482
6~11	-0.0085	0.4063	-0.4592	0.2537	-0.0280	-0.1376	-0.0123
6~12	0.6296	-0.2207	-0.0261	0.1031	0.0502	0.0832	0.0923
6~13	0.0737	-0.1541	1.4347				
7~01	-0.0067	0.0837	0.1414	-0.3544	0.0758		
7~02	-0.1936	0.4013	-0.1340	-0.4909	0.0867	0.0589	0.0145
7~03	0.4189	-0.2075	-0.1993	-0.0247	0.1360	-0.4856	0.0621
7~04	-0.0954	0.4135	-0.4561	0.8267	-0.7338	0.4559	-0.8111
7~05	0.2582	0.0668	-0.0874	0.0183	-0.2771	-0.1529	0.2036
7~06	-0.0072	0.1053	0.0555	0.0782	-0.0796	-0.1324	0.3193
7~07	-0.0696	0.0134	-0.3424	-0.1870	0.7842	-0.0795	-0.2181
7~08	-1.4701	0.0750	-0.1785	0.3999	-0.3844	0.1042	0.2623
7~09	0.2690	-0.3548	0.1550	-0.3290	0.3222	-0.0547	0.3023
7~10	-0.2666	0.0575	0.4978	0.1191	0.0826	0.0539	-0.3260
7~11	-0.2671	-0.0477	-0.1434	0.1385	0.3051	-0.1538	-0.0684
7~12	-0.3692	0.2721	0.2693	0.1632	-0.6712	2.0952	-1.8277
7~13	-0.0544	0.4066	-0.4565	0.2316	0.5584	-0.9017	-0.1109
7~14	0.4071	-0.2601	0.1740	0.1056	-0.1944	0.0516	0.1488

Number of blank cells = 20 (either did not run or had less than 4 participating grades)

Number of normal cells = 298 (within 95% CI)

Number of abnormal cells = 18 (outside 95% CI)

Total number of cells = 336

Positional scores for the LISM_{Median} values for each run

	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	1.1189	1.1014	1.0168	0.9822	1.0623	1.0070	0.8959	0.8714
4~02	0.8601	0.9146	0.8711	0.7187				
4~03	0.9423	0.7962	0.8373	0.9547	0.8978	0.8258	0.7298	
4~04	1.1319	1.2419	1.3941	1.2010	1.0449	1.0553	1.2847	
4~05	0.9010	1.2072	1.0664	1.2416	1.0738	1.1100	0.8553	
4~06	0.8911	0.8573	0.8787	0.9695	0.8423	1.0100	0.9471	1.0387
4~07	1.1463	1.1690	1.1914	1.2039	1.1721	1.1538	1.2486	1.0851
5~01	1.1949	1.2153	1.1440	1.2394	1.0434	0.9106	1.0619	1.0547
5~02	0.9130	1.2153	1.1060	0.9878	0.9301	1.1300	1.3672	1.1744
5~03	1.0269	1.2874	1.3112	1.2900	1.2715	1.3073	0.9903	0.9671
5~04	1.0122	1.1785	1.1976	1.4241	1.3976	1.4975	1.1921	1.5366
5~05	0.8698	1.0345	0.9642	1.1349	1.2241	0.8887	1.0368	0.9718
5~06	0.8848	1.2627	1.3553	1.2770	1.1818	1.1652	1.0027	0.7253
5~07	1.1271	0.8978	1.0640	0.9319	1.0749	1.1627	1.0997	1.0273
5~08	1.2297	1.1450	1.2023	1.3221	1.1648	1.2138	1.2050	1.3608
5~09	0.8576	1.2720	0.9477	1.6480	1.3423	1.2036	1.1597	
5~10	0.5208	0.9043	0.9638	0.9243	0.7274	0.9442	0.9607	0.7646
5~11	1.1662	0.7168	0.7977	0.6459	0.8382	0.6907	0.8231	0.7647
5~12	0.9550	1.2527	0.9587	1.1863	1.3233	1.0997	1.2206	1.1250
5~13	1.0912	0.6756	0.7005	0.5255	0.5158	0.5365	0.5691	0.5285
5~14	0.7026	1.3418	0.8249	1.3533	1.2843	1.0203	1.1711	1.2870
6~01	0.8754	0.8695	1.3477	0.4070				
6~02	1.3769	2.0956	2.2758	2.0712	2.0056	1.9093		0.9471
6~03	0.9438	1.1296	0.9394	1.4586	1.2814	1.4269	1.2199	1.1180
6~04	1.1305	1.5799	1.0991	1.0390	0.8919	1.3402	1.0183	0.8050
6~05	1.5526	1.0209	1.2950	1.4468	1.2441	1.1271	0.9519	0.8468
6~06	1.0259	0.9189	1.0655	1.0289	1.3210	1.1896	0.9541	0.8428
6~07	0.7254	1.0609	0.8782	1.0525	0.9024	0.8709	0.9112	0.9981
6~08	0.8094	0.6762	0.9880	0.6383	0.4091	0.7311	1.3655	0.9587
6~09	1.0088	0.6762	1.0114	0.7978	1.4403	1.6397	1.1389	1.6782
6~10	0.7697	0.8820	0.7310	0.7383	0.8969	0.7063	0.9788	1.0219
6~11	0.8890	0.8885	1.3045	0.8478	1.1197	1.0781	0.9146	0.9087
6~12	0.4009	1.0317	0.8296	0.7900	0.8426	0.9328	1.0151	1.1109
6~13	1.1423	1.1501	0.8763	2.4568				
7~01	1.1194	1.0889	1.1384	1.3652	0.9733	1.0426		
7~02	1.2126	1.0294	1.4442	1.2894	0.8121	0.8952	0.9279	0.9698
7~03	0.6740	1.1052	0.8842	0.7011	0.6982	0.8848	0.3273	0.4118
7~04	1.2131	1.1588	1.5691	1.2223	1.9445	1.1706	1.6866	0.8347
7~05	0.8729	1.1045	1.1522	1.0584	1.0914	0.8058	0.6621	0.8738
7~06	0.3591	0.3597	0.4554	0.5165	0.6030	0.5197	0.3711	0.6920
7~07	1.2311	1.0616	1.0954	0.8869	0.6313	1.3588	1.3499	1.0451
7~08	2.5225	0.9816	1.0644	0.8476	1.2661	0.8900	1.0266	1.2764
7~09	0.7454	0.9787	0.6278	0.7767	0.4533	0.7796	0.7109	1.0424
7~10	0.9183	0.6397	0.7131	1.2086	1.3140	1.4065	1.5231	1.1341
7~11	1.3772	1.1382	1.1314	0.9266	1.1017	1.3817	1.2456	1.1640
7~12	1.1932	0.8641	1.1870	1.3252	1.5046	0.8808	2.9175	1.1346
7~13	1.0475	1.0044	1.4338	0.9500	1.1668	1.8532	0.8687	0.6873
7~14	0.7188	1.1613	0.8610	1.0177	1.1352	0.9507	0.9848	1.1558

Learning intervention spread metrics – median for each run

	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19
4~01	-0.0175	-0.0846	-0.0346	0.0801	-0.0553	-0.1111	-0.0245
4~02	0.0546	-0.0435	-0.1524				
4~03	-0.1462	0.0411	0.1174	-0.0569	-0.0720	-0.0959	
4~04	0.1100	0.1522	-0.1930	-0.1562	0.0104	0.2294	
4~05	0.3062	-0.1408	0.1751	-0.1678	0.0362	-0.2547	
4~06	-0.0338	0.0214	0.0908	-0.1272	0.1677	-0.0629	0.0916
4~07	0.0227	0.0224	0.0124	-0.0317	-0.0184	0.0948	-0.1634
5~01	0.0204	-0.0713	0.0954	-0.1959	-0.1329	0.1513	-0.0072
5~02	0.3023	-0.1093	-0.1182	-0.0578	0.1999	0.2372	-0.1928
5~03	0.2605	0.0237	-0.0212	-0.0185	0.0359	-0.3171	-0.0232
5~04	0.1663	0.0191	0.2264	-0.0264	0.0998	-0.3053	0.3445
5~05	0.1647	-0.0703	0.1708	0.0892	-0.3354	0.1481	-0.0649
5~06	0.3779	0.0926	-0.0783	-0.0952	-0.0167	-0.1625	-0.2773
5~07	-0.2293	0.1662	-0.1321	0.1429	0.0878	-0.0630	-0.0724
5~08	-0.0847	0.0573	0.1198	-0.1573	0.0491	-0.0088	0.1557
5~09	0.4144	-0.3243	0.7004	-0.3058	-0.1387	-0.0438	
5~10	0.3835	0.0595	-0.0395	-0.1969	0.2169	0.0165	-0.1961
5~11	-0.4494	0.0809	-0.1518	0.1923	-0.1475	0.1324	-0.0584
5~12	0.2977	-0.2940	0.2276	0.1370	-0.2237	0.1210	-0.0956
5~13	-0.4156	0.0249	-0.1750	-0.0097	0.0207	0.0326	-0.0406
5~14	0.6392	-0.5169	0.5284	-0.0690	-0.2640	0.1509	0.1159
6~01	-0.0059	0.4782	-0.9407				
6~02	0.7187	0.1802	-0.2047	-0.0656	-0.0963		-1.3195
6~03	0.1858	-0.1902	0.5192	-0.1772	0.1455	-0.2070	-0.1019
6~04	0.4494	-0.4808	-0.0600	-0.1471	0.4483	-0.3219	-0.2133
6~05	-0.5318	0.2741	0.1519	-0.2028	-0.1170	-0.1752	-0.1052
6~06	-0.1070	0.1466	-0.0365	0.2921	-0.1314	-0.2355	-0.1112
6~07	0.3355	-0.1827	0.1743	-0.1501	-0.0315	0.0402	0.0869
6~08	-0.1333	0.3118	-0.3497	-0.2292	0.3220	0.6344	-0.4068
6~09	-0.3326	0.3353	-0.2137	0.6425	0.1994	-0.5008	0.5393
6~10	0.1123	-0.1510	0.0073	0.1586	-0.1906	0.2724	0.0431
6~11	-0.0004	0.4160	-0.4567	0.2719	-0.0416	-0.1635	-0.0059
6~12	0.6308	-0.2021	-0.0396	0.0526	0.0902	0.0823	0.0958
6~13	0.0078	-0.2738	1.5805				
7~01	-0.0305	0.0495	0.2268	-0.3918	0.0693		
7~02	-0.1833	0.4149	-0.1548	-0.4773	0.0832	0.0327	0.0419
7~03	0.4312	-0.2210	-0.1831	-0.0030	0.1866	-0.5575	0.0844
7~04	-0.0542	0.4103	-0.3468	0.7222	-0.7739	0.5159	-0.8519
7~05	0.2317	0.0477	-0.0938	0.0331	-0.2856	-0.1437	0.2117
7~06	0.0005	0.0957	0.0611	0.0864	-0.0833	-0.1486	0.3209
7~07	-0.1695	0.0338	-0.2085	-0.2556	0.7276	-0.0089	-0.3048
7~08	-1.5408	0.0828	-0.2168	0.4185	-0.3761	0.1366	0.2498
7~09	0.2333	-0.3509	0.1489	-0.3234	0.3263	-0.0687	0.3315
7~10	-0.2787	0.0734	0.4955	0.1054	0.0924	0.1167	-0.3890
7~11	-0.2390	-0.0068	-0.2048	0.1751	0.2800	-0.1360	-0.0816
7~12	-0.3290	0.3229	0.1381	0.1794	-0.6238	2.0367	-1.7828
7~13	-0.0431	0.4294	-0.4838	0.2168	0.6864	-0.9845	-0.1813
7~14	0.4424	-0.3003	0.1567	0.1175	-0.1845	0.0341	0.1710

Number of blank cells = 20 (either did not run or had less than 4 participating grades)

Number of normal cells = 296 (within 95% CI)

Number of abnormal cells = 20 (outside 95% CI)

Total number of cells = 336

Screen shots from the LYW4.life website

Appendix 26.1 Login/Registration

Figure 18 - Opening screen



Figure 19 - Registering with LYW4

Register Now

Agreement to participate

- I have read the information on the [About page](#) about this study.
- I have had an opportunity to ask questions and discuss this study. Click [here](#) to send an email.
- I have received satisfactory answers to any of my questions.
- I have received enough information about this study.
- I understand that I am free to withdraw from this study:-
 - at any time (until such date as this will no longer be possible, of which I will be told).
 - without giving a reason for withdrawing.
 - without it affecting my grades or my future at the University of Greenwich.
- I agree to take part in this study.

☒ I agree ☐ I decline (Please check one of these two boxes)

For everyone

Your data will not be made available in any format that would identify you. Please contact my lead supervisor, Prof. Lachlan MacKinnon if you encounter problems. His email address is lmackinnon@gre.ac.uk. His telephone number is +44 (0)20 8331 8559.

Insert your pass phrase

☐ Show pass phrase

Re-enter your pass phrase

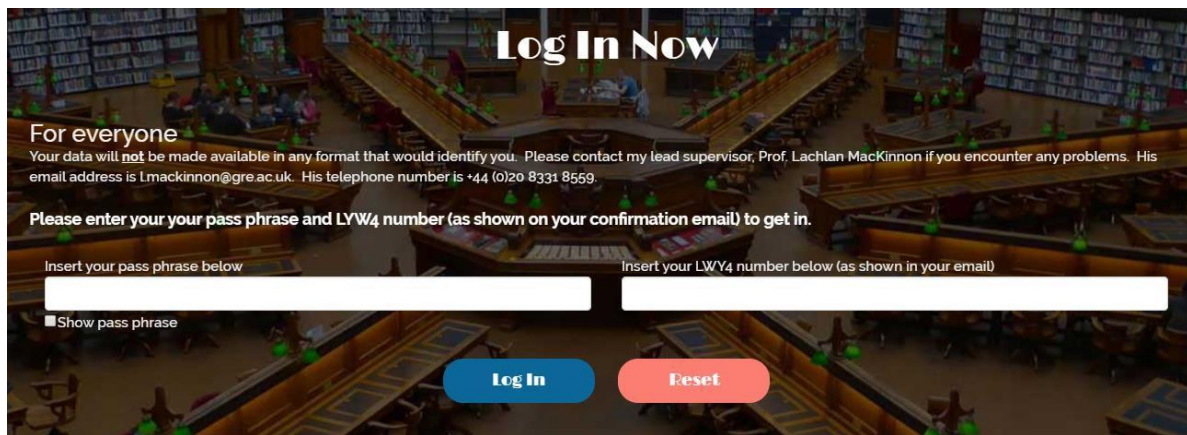
☐ Show pass phrase

Insert your email address below

Repeat your email address below

Register **Reset**

Figure 20 - Logging in to LYW4



Log In Now

For everyone
Your data will not be made available in any format that would identify you. Please contact my lead supervisor, Prof. Lachlan MacKinnon if you encounter any problems. His email address is Lmackinnon@gre.ac.uk. His telephone number is +44 (0)20 8331 8559.

Please enter your your pass phrase and LYW4 number (as shown on your confirmation email) to get in.

Insert your pass phrase below

Insert your LYW4 number below (as shown in your email)

☐ Show pass phrase

Log In **Reset**

Appendix 26.2 The Home Page

Figure 21 - The home page once logged in

ADMINISTRATION
SETTINGS
PAPERMAKER
USAGE
LOG OFF

My Portfolio
Enrol on a run
Open a run
Overview of runs

Enrol on a course
Open a course
Overview of courses

Enrol on a task
Open a task
Overview of tasks

My Teaching
Create a run
Enrol a student
Student progress
Add a tutor
Amend a tutor
Edit a run
Run results
Move student

My Creations
Find a task
Clone a task
Create a task
Edit a task
Task status

Find a course
Clone a course
Create a course
Edit a course

Settings
Change consent
Change email add.
Change pass phrase
Change background/font
Suggest feature

Other Items
PaperMaker
Site usage
Log off

My Office - (LYW4 No.: 1)

How do I use LYW4.life?

Clicking on either **Administration** in the menu bar or the **LYW4 logo** in front of My Office above will return you to these instructions.

Everyone is a student, teacher and content creator. It is up to you how use this site. Please see the courses LYW4 Everyone, LYW4 Students, LYW4 Tutors and LYW4 Content Creators for the details.

Definitions

Task : A small, single item of learning. You can see it as text on screen, as a video, listen to it or download the PDF if you prefer to read things quietly. There is a short quiz at the end which you can take to make sure that you understood this. Passing the test gains you a certificate and scores points.

Topic : A collection of tasks. This will usually be set up by someone who wants to create a collection of materials for their learners. The order of the tasks is the order in which they are likely to be taught, but there is nothing to stop you working ahead.

Run : A collection of topics created for a single group of students. Tutors will use this to create a group which they can help. It is a run of a course and so can be repeated year after year for different groups. This is a taught course and can only be taken if you are enrolled at a learning institution.

Please select a run
wa02-1556-2018/19

Video - Welcome to LYW4.life
Video - How to use LYW4.life

Success Rate for wa02-2018/19

Maximum Score = 666 (40% of Maximum Score = 266) (30% of Maximum Score = 200)

Green shows a level of understanding needed for the coursework. Orange shows a level of understanding needed to attempt the coursework.

Score	No. of Students with this Score
666	4
630	0
595	0
561	0
528	0
496	0
465	0
435	0
406	0
378	0
351	0
325	0
300	1
276	1
253	0
231	1
210	2
190	1
171	0
153	0
136	0
120	0
105	3
91	1
78	0
66	1
55	0
45	1
36	0
28	6
21	3
15	3
10	5
6	4
3	10
1	3
0	46

15.59% completed

(c) 2018, 2019 LYW4.life

Safe browsing
Explicit redactions
Automatic Encryption

Appendix 26.3 Enrolment

Figure 22 - Entering the code for the run



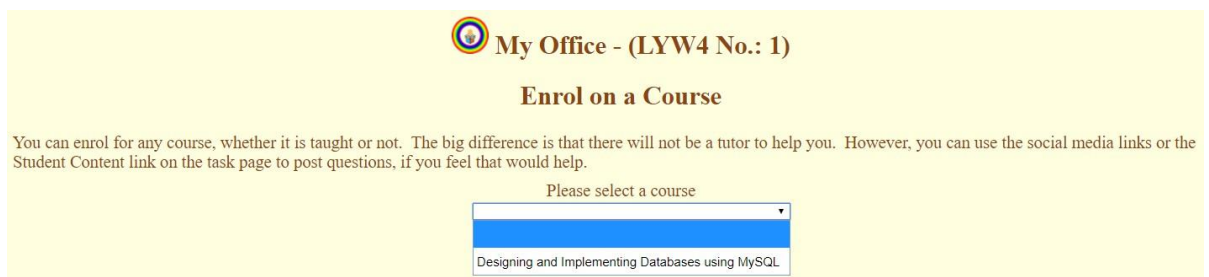
My Office - (LYW4 No.: 1)

Enrol on a Course Run

Your teacher/lecturer should have provided you with a code (e.g. *wa02-2018/19*). Please ask for one from them if you do not know it. Enter that course code in the box below when you have it. You can then ask your tutor to accept your enrolment when you see them next time in class.

Enrol on this

Figure 23 - Applying for a run without enrolling



My Office - (LYW4 No.: 1)

Enrol on a Course

You can enrol for any course, whether it is taught or not. The big difference is that there will not be a tutor to help you. However, you can use the social media links or the Student Content link on the task page to post questions, if you feel that would help.

Please select a course

- Designing and Implementing Databases using MySQL

Appendix 26.4 A Task

Figure 24 - Opening a module run

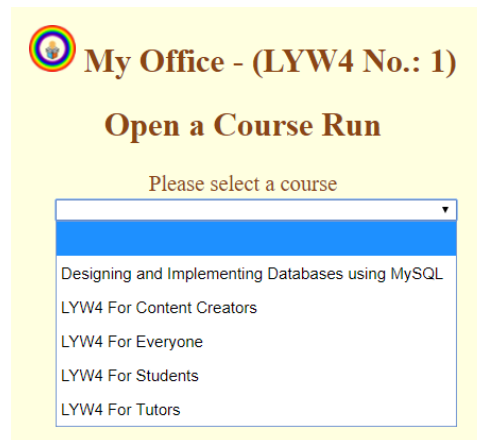


Figure 25 - Selecting a task

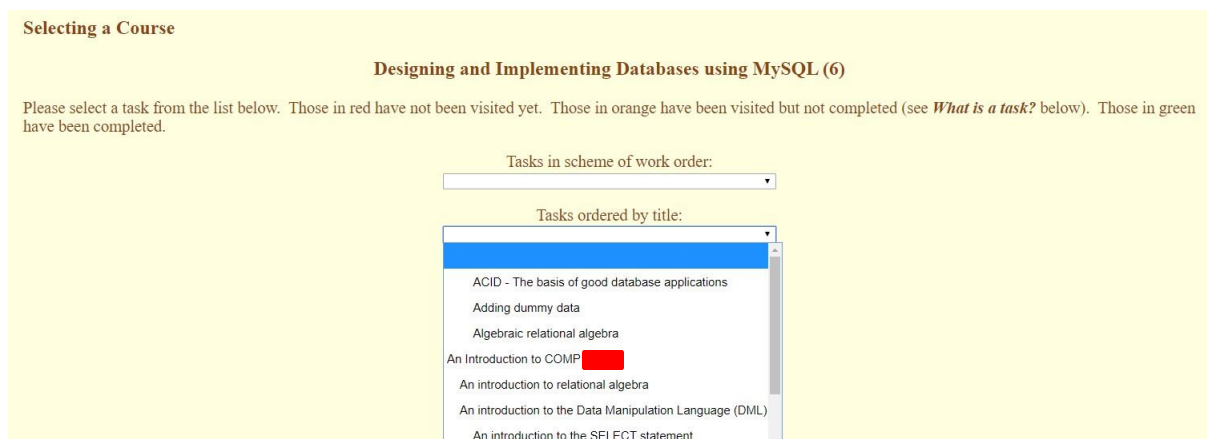


Figure 26 - A task

MY OFFICE

MATERIALS

QUIZ

COURSEWORK

LINKS

FEEDBACK

My Office

Materials

Main contents

Video, MP3 and PDF

Student generated content

Relies on tasks

External links

Knowledge Check

Quiz

Coursework

Additional materials

Feedback

Searching

Resources

File Storage

Social Media

An Introduction to (2)

A different task?

View this task

This is a level 4 (2nd year undergraduate) course in database design.

Here you will learn about why a database has to be normalised, how to design a sustainable database and how to use SQL to query and manipulate a database.

Video

MP3

PDF

Introduction

Welcome to This is a level 4 (second-year undergraduate) course which will cover the following three areas:-

1. The design of sustainable databases

2. The creation of databases filled with dummy data

3. The use of real-world SQL

The design of sustainable databases

We will be using object-oriented relational database design (OORDBD) which has some significant advantages: you do not need to go through the three levels of normalisation, your database will be in fourth normal form (4NF), writing the SQL is easier and making changes when the client says, "I've had an idea ..." does not require renormalising your database.

There is a six-step process, the last of which is adding the fields which would be the last step in conventional normalisation.

The creation of databases filled with dummy data

You will see how you can create all the tables and their fields without typing a single line of code! You can do this by creating your ERDs in an on-line package called [LucidChart](#). There is a link which, when clicked, shows all the SQL needed to create the tables. You just copy that those commands into your DBMS, run it and the tables are created. This is free for academic users.

You can then use a data generation package, such as [GenerateData.com](#) or [Mockaroo.com](#), to populate these tables with hundreds of rows of dummy data. This makes testing your queries MUCH easier. But (and there is always a "but"), you do have to work out the order in which the tables have to be populated because the PKs and FKs must be synchronised - and that requires planning.

The use of real-world SQL

Most people on this level, of course, will have used the CREATE TABLE and SELECT commands. However, there is much more that SQL can do. This course takes you through the more complex areas that are used in real situations.

How will you be assessed?

You have a coursework to complete. Most of the coursework is done in pairs - you find a buddy and team up. You register with whom you are working by telling your tutor.

The coursework presents you with a scenario. The first task is to create a conceptual ERD (along with any sub-classes). You then create the physical ERD which you then use to create the tables in MySQL. The next step is to create a plan of how the tables are to be populated and get the data in there. These steps form the bulk of the coursework and are completed with your pair.

The final steps are to create some SQL queries (there are two sets and you decide who will do which set). Finally, you evaluate the work you and your pair have produced in an academic way.

(c) 2018, 2019 LYW4Life

Safe browsing

Exploit mitigations

Machine Integrity

191

Appendix 26.5 A Quiz

Figure 27 - The quiz for a task

An Introduction to [REDACTED] (2)

1) Which of these is covered on [REDACTED]? (Choose 2)

- ☒ Normalisation
- ☒ Differentiation of records
- ☐ OORDBD
- ☐ Integration of records
- ☐ The 12 rules of RDBs

2) How will you be assessed? (Choose 1)

- ☒ By coursework
- ☐ By examination
- ☐ By examination in pairs
- ☐ By coursework in pairs
- ☐ By coursework in small groups

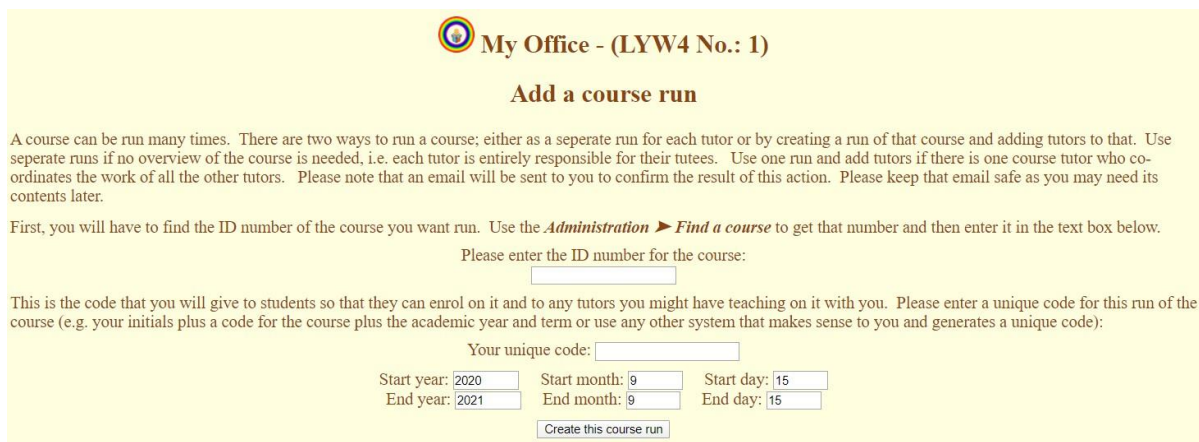
3) Why is dummy data needed? (Choose 2)

- ☒ To ensure that only intelligent people can use the DB.
- ☒ To allow the SQL to be tested.
- ☐ Because thorough testing is good practice.
- ☐ Because dummy data is needed to seed the system.
- ☐ Because dummy data clears the DB.

[Score this quiz](#)

Appendix 26.6 Run Administration

Figure 28 - Creating a run for a module



My Office - (LYW4 No.: 1)

Add a course run

A course can be run many times. There are two ways to run a course; either as a separate run for each tutor or by creating a run of that course and adding tutors to that. Use separate runs if no overview of the course is needed, i.e. each tutor is entirely responsible for their tutees. Use one run and add tutors if there is one course tutor who coordinates the work of all the other tutors. Please note that an email will be sent to you to confirm the result of this action. Please keep that email safe as you may need its contents later.

First, you will have to find the ID number of the course you want run. Use the **Administration ► Find a course** to get that number and then enter it in the text box below.

Please enter the ID number for the course:

This is the code that you will give to students so that they can enrol on it and to any tutors you might have teaching on it with you. Please enter a unique code for this run of the course (e.g. your initials plus a code for the course plus the academic year and term or use any other system that makes sense to you and generates a unique code):

Your unique code:

Start year: Start month: Start day:
 End year: End month: End day:

Figure 29 - Accepting a student onto a run



COMP1556 - Database Applications Technology

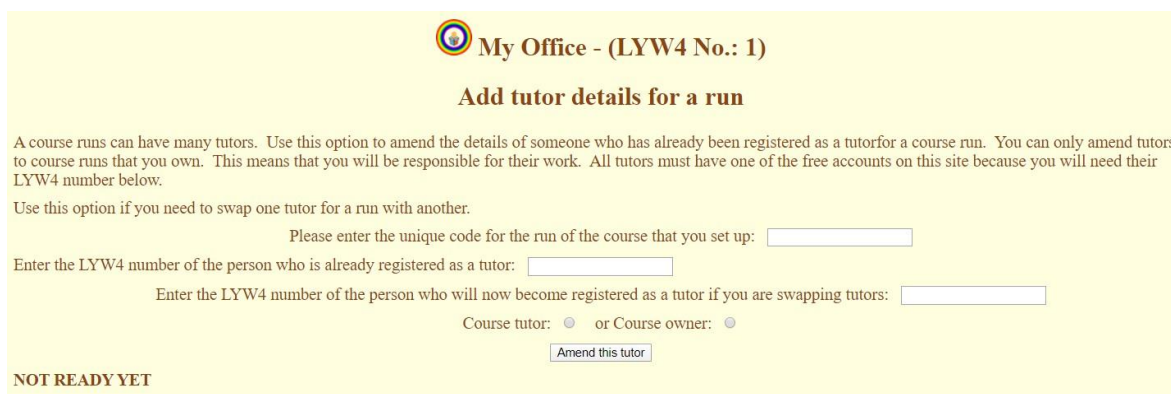
2019-01-15 to 2019-04-15

Please only enrol those students who are your tutees (i.e. appear on your register). You could approve any applicant for this run, but they may then become your responsibility. All students can access the materials whether they have been accepted or not, so they are not being disadvantaged by being asked to go to the correct tutorial group to be accepted.

Please select a student id from the list below

(c) 2018, 2019 LYW4.life

Figure 30 - Adding a tutor to a run



My Office - (LYW4 No.: 1)

Add tutor details for a run

A course runs can have many tutors. Use this option to amend the details of someone who has already been registered as a tutor for a course run. You can only amend tutors to course runs that you own. This means that you will be responsible for their work. All tutors must have one of the free accounts on this site because you will need their LYW4 number below.

Use this option if you need to swap one tutor for a run with another.

Please enter the unique code for the run of the course that you set up:

Enter the LYW4 number of the person who is already registered as a tutor:

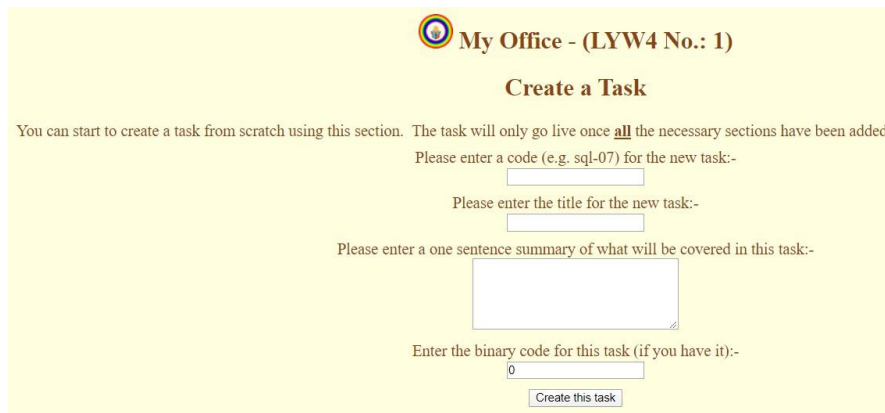
Enter the LYW4 number of the person who will now become registered as a tutor if you are swapping tutors:

Course tutor: ☐ or Course owner: ☐

NOT READY YET

Appendix 26.7 Creating and Editing Tasks

Figure 31 - Creating a task outline



My Office - (LYW4 No.: 1)

Create a Task

You can start to create a task from scratch using this section. The task will only go live once **all** the necessary sections have been added.

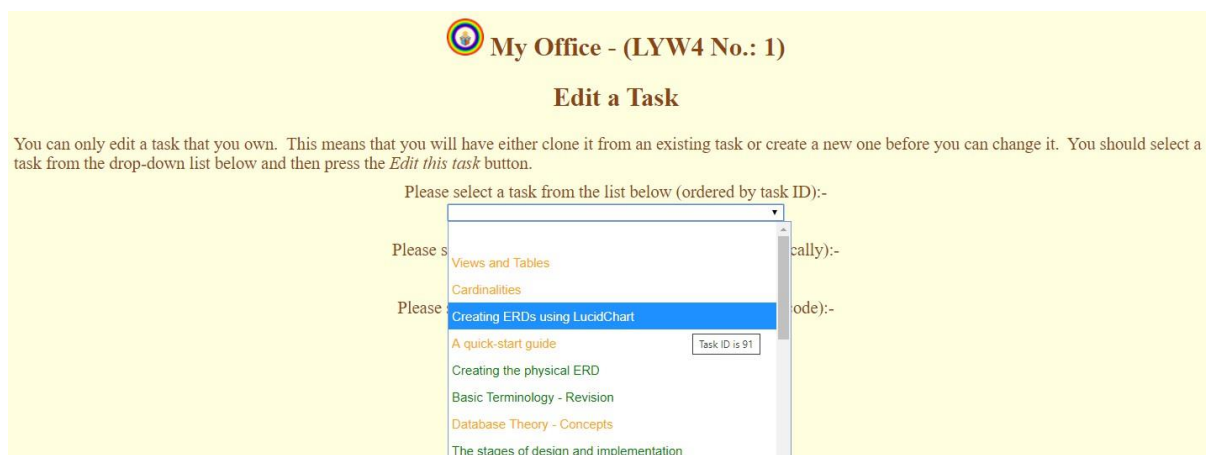
Please enter a code (e.g. sql-07) for the new task:-

Please enter the title for the new task:-

Please enter a one sentence summary of what will be covered in this task:-

Enter the binary code for this task (if you have it):-

Figure 32 - Editing the task



My Office - (LYW4 No.: 1)

Edit a Task

You can only edit a task that you own. This means that you will have either clone it from an existing task or create a new one before you can change it. You should select a task from the drop-down list below and then press the *Edit this task* button.

Please select a task from the list below (ordered by task ID):-

Please s
Views and Tables
Cardinalities
Please
Creating ERDs using LucidChart
A quick-start guide
Creating the physical ERD
Basic Terminology - Revision
Database Theory - Concepts
The stages of design and implementation

Task ID is 91

Figure 34 - Editing the task summary

Change the Task Summary

Creating ERDs using LucidChart

Use this section when wanting to change the wording of the summary for this task.

Please alter the summary if you wish:-

This task shows you how to quickly create an ERD which looks professional.

Figure 35 - Adding video, sound and PDF

Add/Change the Alternative Media Links

Creating ERDs using LucidChart

Use this section add/alter the address of the video, audio file or PDF associated with this task. Audio and PDF files must be uploaded before they can be added.

Please enter the **URL of the video** for this task:-

https://youtu.be/mK_h9lD

Current audio file for this task: 63

Please select the **audio file** for this task (.mp3, .wav or .ogg):-

No file chosen

Current PDF file for this task: 64

Please select the **PDF file** for this task (.pdf only):-

No file chosen

Figure 36 - Adding links to other sites

Add these External Links

Creating ERDs using LucidChart

Use this section add/alter the list of links to other sites associated with this task.

Please enter the URL for the external link that you would like to add (e.g. <http://www.lyw4.life> name this link as **Learning YOUR Way**):-

	name this link as	
	name this link as	
	name this link as	
	name this link as	
	name this link as	

Figure 37 - Adding extra materials to a task

Add these Associated Materials

Creating ERDs using LucidChart

Use this section add/alter the list of associated materials for this tasks, e.g. notes, case study, diagrams, etc.

Please note that you must have uploaded the materials before using this section using the *Upload Task Materials* button on the **My Office** page. The select the materials that you would like to add from the following drop-down list:-

Already associated with this task

You have not attached any materials to this task ... yet.

Figure 38 - Adding links to prerequisite tasks

Add these Prerequisite Tasks

Creating ERDs using LucidChart

Use this section add/alter the list of tasks which would preceed this one and help the student to understand this content.

Please enter the Task ID for the prerequisite tasks that you would like to add:-

Figure 39 - Creating quiz questions

MY OFFICE TASK DETAILS RUN MATERIALS DIFFERENT TASK

My Office

Task Items

Task title

Task code

Summary

Contents

Task status

Video, audio, PDF

External Links

Extra materials

Prerequisites

Quiz

Run materials

Different task

Add these Quiz Questions

Creating ERDs using LucidChart

This task shows you how to quickly create an ERD which looks professional.

Use this section create the quiz questions for this task. There should be at least three questions for each task. Each question can have multiple correct answers. Click the check box next to the correct answers.

Please enter question 1:-

What are the first TWO this

☒ Select the Entity Relations

☒ Drag a three-bar entity ont

☒ Link the entities together

☒ Go into LucidChart.

☒ Decide on the list of object

Please enter question 2:-

Which of these should a w

☒ A blank central column.

☒ An underlined PK.

☒ An italicised or emboldene

☒ Either PK or FK in the first

☒ All fields formatted with str

Please enter question 3:-

Links are created by

☒ Highlighting both entities, r

☒ Dragging from the one-enc

☒ Dragging from the many-e

☒ Placing the cursor over the

☒ Placing the cursor over the

Please enter question 4 (optional):-

☒

☒

☒

☒

Figure 40 - Seeing an overview of tasks created

Task Status - (LYW4 No.: I)															
111e 6Sh be0Ware presented in task code order so that you'll can see related groups of tasks.															
TaskID	TaskCode	An Lntroduction to-- Task Title	Words	https://youtu.be/...	Audio	PDF	Link	Ops	Link	Link	Kn...	ord	Staps	trivs	Comnle
S7	1556-02	Using the coursework template	3487	https://youtu.be/...	13	12	41	0	0	0	0	0	13		
S	DBDes-000	The world of databases	13*	https://youtu.be/...	7	S	41	0	0	0	0	0	13		
88	DBDes-001	Basic Technology: Revision	17961	https://youtu.be/...	0			0	0	0	0	0	13		
7	DBDes-002	Views of DBs	38833	https://youtu.be/...	44	6IS	41	0	0	0	0	0	13		
86	DBDes-002	The stages of design and implementation	SO1	https://youtu.be/...	16	4	41	4	0	0	0	0	13		
9	DBDes-004	Using a data generator: Overview	698	https://youtu.be/...	20	19	4	0	0	0	0	0	13		
8	DB-003	What is SQL?	529	https://youtu.be/...	48	17	41	0	0	0	0	0	13		
82	DBDes-005	The relational database	1926	https://youtu.be/...	24	53	41	4	0	0	0	0	13		
11	DBDes-007	Getting Lucidchart for 111e	1356	https://youtu.be/...	35	36	41	0	0	0	0	0	13		
13	DBDes-009	Stage 1 - Intro to the objects	8169	https://youtu.be/...	69	23	47	0	0	0	0	0	13		
15	DBDes-010	Stage 3: Adding the attributes	15444	https://youtu.be/...	57	55	41	0	0	0	0	0	13		
16	DBDes-011	Stage 4 - Adding the primary keys	20518	https://youtu.be/...	52	52	41	0	0	0	0	0	13		
17	DBDes-012	Stage 5: Adding the foreign keys	39425	https://youtu.be/...	51	49	4	0	0	0	0	0	13		
18	DBDes-013	Stage 6: Adding the fields	31678	https://youtu.be/...	48	46	41	0	0	0	0	0	13		
19	DBDes-014	Making changes to the design	80913	https://youtu.be/...	45	43	41	0	0	0	0	0	13		
20	DBDes-015	Using a data generator	38924	https://youtu.be/...	45	43	41	0	0	0	0	0	13		
21	DBDes-016	Using a data generator	6173	https://youtu.be/...	42	40	481	0	0	0	0	0	13		
89	DBDes-017	Using a data generator	11609	https://youtu.be/...	39	11	41	0	0	0	0	0	13		
91	DBDes-018	Using a data generator	27570	https://youtu.be/...	63	61	1	0	0	0	0	0	13		
22	DBDes-019	Using a data generator	14375	https://youtu.be/...	68	66	41	0	0	0	0	0	13		
23	DBDes-020	Using a data generator	95325	https://youtu.be/...	68	66	41	0	0	0	0	0	13		
24	DBDes-021	Using a data generator	40888	https://youtu.be/...	70	68	41	0	0	0	0	0	13		
25	DBDes-022	Using a data generator	5158	https://youtu.be/...	71	69	41	0	0	0	0	0	13		
26	DBDes-023	Using a data generator	25014	https://youtu.be/...	71	69	41	0	0	0	0	0	13		
27	DBDes-024	Using a data generator	51530	https://youtu.be/...	72	70	41	0	0	0	0	0	13		
87	DBDes-026	Using a data generator	0	https://youtu.be/...	41	0	0	0	0	0	0	0	13		
47	DBDes-027	Using a data generator	0	https://youtu.be/...	41	0	0	0	0	0	0	0	13		
48	DBDes-028	Using a data generator	0	https://youtu.be/...	41	0	0	0	0	0	0	0	13		
49	DBDes-029	Using a data generator	0	https://youtu.be/...	41	0	0	0	0	0	0	0	13		
50	DBDes-030	Using a data generator	0	https://youtu.be/...	41	0	0	0	0	0	0	0	13		
51	DBDes-031	Using a data generator	0	https://youtu.be/...	41	0	0	0	0	0	0	0	13		
52	DBDes-032	Using a data generator	325	https://youtu.be/...	41	0	0	0	0	0	0	0	13		
53	DBDes-033	Using a data generator	0	https://youtu.be/...	41	0	0	0	0	0	0	0	13		
54	DBDes-034	Using a data generator	0	https://youtu.be/...	41	0	0	0	0	0	0	0	13		
55	DBDes-035	Using a data generator	0	https://youtu.be/...	41	0	0	0	0	0	0	0	13		
56	DBDes-036	Using a data generator	467	https://youtu.be/...	33	31	41	0	0	0	0	0	13		
59	LYW4-02	The Words - Tasks, Courses and Runs	802	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
90	LYW4-03	The Words - Tasks, Courses and Runs	1150	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
75	LYW4-04	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
80	LYW4-05	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
81	LYW4-06	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
82	LYW4-07	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
83	LYW4-08	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
84	LYW4-09	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
58	LYW4-10	The Words - Tasks, Courses and Runs	825	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
60	LYW4-11	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
61	LYW4-12	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
62	LYW4-13	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
63	LYW4-14	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
64	LYW4-15	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
85	LYW4-16	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
65	LYW4-17	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
66	LYW4-18	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
67	LYW4-19	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
68	LYW4-20	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
69	LYW4-21	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
70	LYW4-22	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
71	LYW4-23	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
72	LYW4-24	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
73	LYW4-25	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
74	LYW4-26	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
75	LYW4-27	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
76	LYW4-28	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
77	LYW4-29	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
78	LYW4-30	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
79	LYW4-31	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
80	LYW4-32	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
81	LYW4-33	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
82	LYW4-34	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
83	LYW4-35	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
84	LYW4-36	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
85	LYW4-37	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
86	LYW4-38	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
87	LYW4-39	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
88	LYW4-40	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
89	LYW4-41	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
90	LYW4-42	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
91	LYW4-43	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
92	LYW4-44	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
93	LYW4-45	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
94	LYW4-46	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
95	LYW4-47	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
96	LYW4-48	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
97	LYW4-49	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
98	LYW4-50	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
99	LYW4-51	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
100	LYW4-52	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
101	LYW4-53	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
102	LYW4-54	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		
103	LYW4-55	The Words - Tasks, Courses and Runs	0	https://youtu.be/...	41	0	0	0	0	0	0	0	12		

Appendix 26.8 Checking Student Progress

Figure 41 - Monitoring the progress of a student


My Office - (LYW4 No.: 1)

Student Progress

If you are a registered tutor on a course then you view the progress of those you have accepted onto it. Please select a course from the drop-down list below. Another drop-down box will then appear which shows a list of those who are registered to you. You will then see the results for that course and that student. For comparison, the averages for your tutees and those of the course as a whole will be shown.

Please select a run from the drop-down below: [Redacted]

Please select a student LYW number : 59

[Check the progress for this student](#)

Figure 42 - The student report for a run

[Redacted] - Database Applications Technology [wa02-1556-2018/19]

Designing and Implementing Databases using MySQL [55]

My Office

Run Progress

Logoff

This course covers a revision of the terminology, the design of sustainable databases (using OORDBD) as well as the use of enough SQL to enable you to create, populate and manipulate the data. You will also cover the use of dummy data.

[Joined: 2019-01-15 15:13:51] ~ [Last Accessed: 2019-03-24 22:20:51] ~ [No. of Logins: 12] ~ [No. of Clicks: 0]

2	1556-01	An Introduction to [Redacted]	0	2019-02-01 11:12:00
6	1556-02	Creating pairings	0	2019-02-01 11:14:06
57	1556-03	Using the coursework template	0	2019-02-01 11:18:18
5	DBDes-000	The world of databases	0	2019-02-01 11:28:04
88	DBDes-001	Basic Terminology - Revision	0	2019-02-01 11:29:33
86	DBDes-002	The stages of design and implementation	0	2019-02-01 11:38:42
9	DBDes-004	Using a data generator - Overview	0	2019-02-19 12:00:09
10	DBDes-005	Querying a database	0	2019-02-19 12:04:28
8	DBDes-003	What is SQL?	0	2019-02-19 12:06:15
20	DBDes-015	Types of ERDs	0	2019-02-19 12:10:27
12	DBDes-006	Creating an object-oriented relational database	0	
13	DBDes-007	Getting LucidChart for academics	0	2019-01-27 20:36:31
91	DBDes-018	Creating ERDs using LucidChart	0	2019-02-19 12:17:28
29	DBDes-008	Stage 1 - Finding the objects	0	2019-01-27 20:38:55
14	DBDes-009	Stage 2 - Linking the objects	0	2019-01-27 20:42:29
15	DBDes-010	Stage 3 - Adding the cardinalities	0	2019-01-27 20:43:44
16	DBDes-011	Stage 4 - Adding the primary keys	0	2019-01-27 20:59:47
17	DBDes-012	Stage 5 - Adding the foreign keys	0	2019-02-19 11:48:12
18	DBDes-013	Stage 6 - Adding the fields	0	2019-02-19 11:49:23
19	DBDes-014	Making changes to the design	0	2019-02-19 11:52:52
21	DBDes-016	Using sub-classes in ERDs	0	
89	DBDes-017	Creating the physical ERD	0	
22	DBSetup-001	Creating a database in MySQL	0	
23	DBSetup-002	Downloading and installing EasyPHP	0	
26	DBSetup-005	Planning to use a data generator	0	
28	DBSetup-007	Generating the dummy data	0	
30	SQL-002	The Data Definition Language (DDL)	0	
31	SQL-003	An introduction to the Data Manipulation Language (DML)	0	
32	SQL-004	An introduction to the SELECT statement	0	
33	SQL-005	Using wild cards in SELECT statements	0	
34	SQL-006	Using BETWEEN and IN in SELECT statements	0	
35	SQL-007	Using ORDER BY in a SELECT statement.	0	
36	SQL-008	Linking several tables	0	
37	SQL-009	Using built-in functions in MySQL	0	
38	SQL-010	Using variables and last insert_id	0	
39	SQL-011	Using GROUP BY and HAVING	0	

(c) 2018, 2019 LYW4Lib6

[Safe browsing](#) [Export metadata](#) [Keywords description](#)

Figure 43 - An overview of all students on a run

Database Applications Technology Iwa021-S56-2018/1/1

Designing and Implementing Databases using MySQL

This course covers: 1) how to design and implement a database using MySQL; 2) how to use MySQL to create, update, delete, and insert data; 3) how to use MySQL to create, update, delete, and insert data. You will also cover the use of dummy data.

LYWIG	Accepted	Last Accessed	Logins	Committed	SQL
1	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
2	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
3	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
4	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
5	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
6	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
7	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
8	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
9	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
10	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
11	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
12	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
13	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
14	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
15	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
16	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
17	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
18	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
19	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
20	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
21	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
22	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
23	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
24	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
25	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
26	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
27	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
28	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
29	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
30	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
31	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
32	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
33	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
34	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
35	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
36	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
37	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
38	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
39	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
40	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
41	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
42	2019-01-29 02:37:14	2019-04-17 18:34:36	46	14 completed out of 36	Score=105
43	2019-01-29				

No. Enrolled • 98

Completions 545 out of 3.528 (15.4%)

Appendix 26.9 View Progress

Figure 44 - Progress report - The student view


My Office - (LYW4 No.: 1)

Overview of all my Runs

You can monitor your progress on all your runs below. Tutors cannot see this page. They can only see the runs for which you have been accepted by them. They cannot see your results for any other courses.

Please select a course run

See this run

Figure 45 - Progress so far - The student view

My Office

Logoff

Database Applications Technology [wa02-1556-2018/19]

Designing and Implementing Databases using MySQL [1]

This course covers a revision of the terminology, the design of sustainable databases (using OORDBD) as well as the use of enough SQL to enable you to create, populate and manipulate the data. You will also cover the use of dummy data.

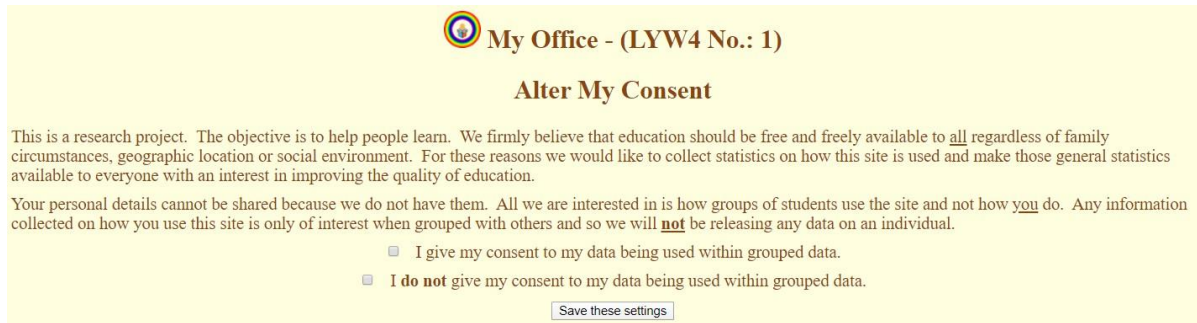
2	1556-01	An Introduction to	2018-12-22 16:02:20
6	1556-02	Creating pairings	2019-01-31 11:47:54
57	1556-03	Using the coursework template	2019-01-23 06:28:09
5	DBDes-000	The world of databases	2019-02-04 15:41:47
88	DBDes-001	Basic Terminology - Revision	
86	DBDes-002	The stages of design and implementation	
9	DBDes-004	Using a data generator - Overview	
10	DBDes-005	Querying a database	
8	DBDes-003	What is SQL?	
20	DBDes-015	Types of ERDs	
12	DBDes-006	Creating an object-oriented relational database	
13	DBDes-007	Getting LucidChart for academics	
91	DBDes-018	Creating ERDs using LucidChart	
29	DBDes-008	Stage 1 - Finding the objects	
14	DBDes-009	Stage 2 - Linking the objects	
15	DBDes-010	Stage 3 - Adding the cardinalities	
16	DBDes-011	Stage 4 - Adding the primary keys	
17	DBDes-012	Stage 5 - Adding the foreign keys	2019-02-13 06:32:22
18	DBDes-013	Stage 6 - Adding the fields	
19	DBDes-014	Making changes to the design	
21	DBDes-016	Using sub-classes in ERDs	
89	DBDes-017	Creating the physical ERD	
22	DBSetup-001	Creating a database in MySQL	
23	DBSetup-002	Downloading and installing EasyPHP	
26	DBSetup-005	Planning to use a data generator	
28	DBSetup-007	Generating the dummy data	
30	SQL-002	The Data Definition Language (DDL)	
31	SQL-003	An introduction to the Data Manipulation Language (DML)	
32	SQL-004	An introduction to the SELECT statement	
33	SQL-005	Using wild cards in SELECT statements	
34	SQL-006	Using BETWEEN and IN in SELECT statements	
35	SQL-007	Using ORDER BY in a SELECT statement.	
36	SQL-008	Linking several tables	
37	SQL-009	Using built-in functions in MySQL	
38	SQL-010	Using variables and last insert_id	
39	SQL-011	Using GROUP BY and HAVING	

(c) 2018, 2019 LYW4 Ltd

Safe browsing
Exploit mitigations
Securely Encrypted

Appendix 26.10 Personalisation

Figure 46 - Consent for research data



 My Office - (LYW4 No.: 1)

Alter My Consent

This is a research project. The objective is to help people learn. We firmly believe that education should be free and freely available to all regardless of family circumstances, geographic location or social environment. For these reasons we would like to collect statistics on how this site is used and make those general statistics available to everyone with an interest in improving the quality of education.

Your personal details cannot be shared because we do not have them. All we are interested in is how groups of students use the site and not how you do. Any information collected on how you use this site is only of interest when grouped with others and so we will **not** be releasing any data on an individual.

☒ I give my consent to my data being used within grouped data.

☐ I do **not** give my consent to my data being used within grouped data.

[Save these settings](#)

Figure 47 - Altering the email address



 My Office - (LYW4 No.: 1)

Update My EMail Address

We need your email address so that we can send you certificates and verify that you are a real person. We will **not** be releasing your email address at any stage, even for valid research.

Your current main email address is: **lord.wicks@gmail.com**

Please enter your new email address:

Please re-enter your new email address again:

[Save this email address](#)

Figure 48 - Changing the background colours



 My Office - (LYW4 No.: 1)

Change the Colour and Font Size

Some people may prefer a different set of colours and/or font sizes. Please select your preferences from the drop-down lists below. Just keep changing them until you have something that works for you.

[Save these settings](#)

-
- I prefer this
- I prefer this
- I prefer this
- I prefer this
- I prefer this
- I prefer this
- I prefer this
- I prefer this
- I prefer this
- I prefer this
- I prefer this
- I prefer this
- I prefer this
- I prefer this

Figure 49 - Changing the font size

My Office - (LYW4 No.: 1)

Change the Colour and Font Size

Some people may prefer a different set of colours and/or font sizes. Please select your preferences from the drop-down lists below. Just keep changing them until you have something that works for you.

I prefer this
I prefer this
I prefer this
I prefer this
I prefer this
I prefer this
I prefer this
I prefer this
I prefer this
I prefer this

Figure 50 - Feature request form

My Office - (LYW4 No.: 1)

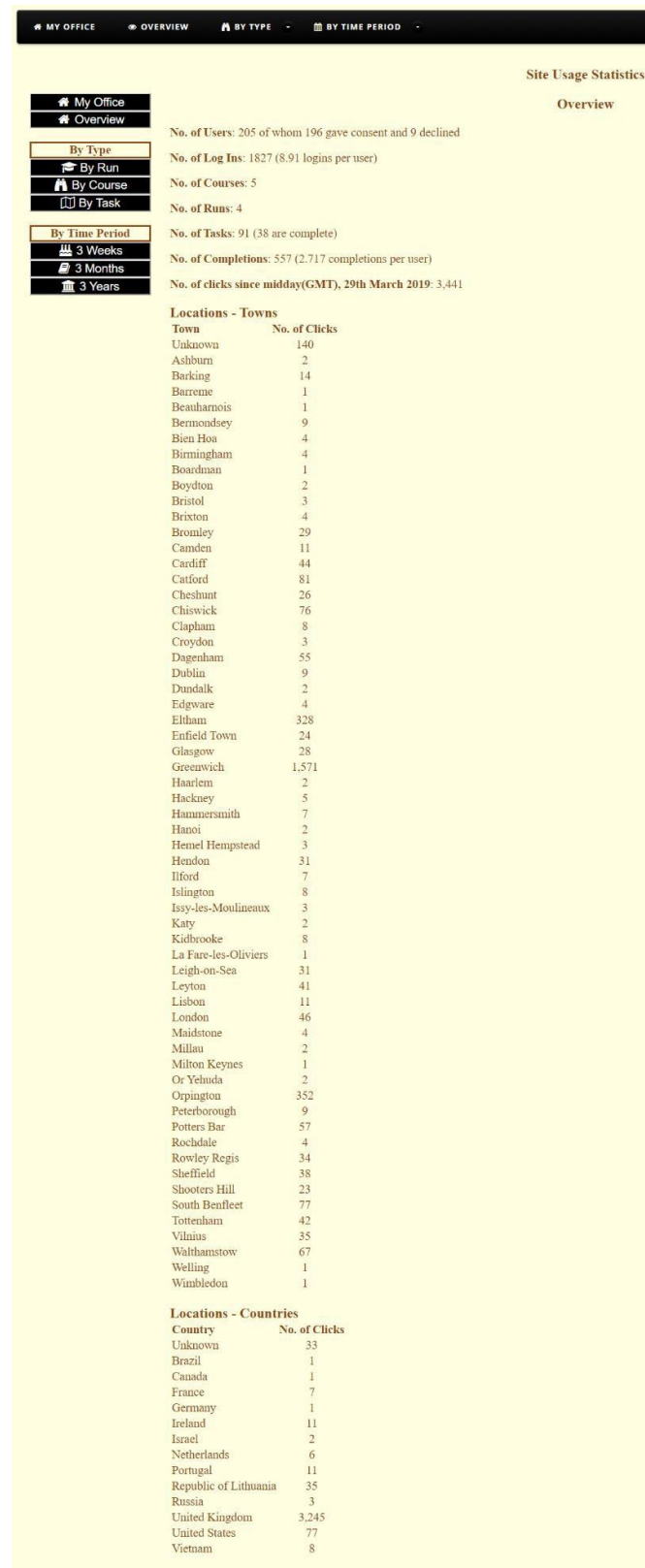
Request to add a feature or report an error

This site is still in its development phase, so please help by suggesting features you feel might be helpful to others. You can also use this area to tell me of any mistakes that you have spotted. **All help is gratefully accepted.** Please include lots of detail and even screen shots to help me add what you need or locate an error.

Please enter your suggestion for amendment.

Appendix 26.11 Usage Statistics

Figure 51 - Usage statistics



Ethical approval confirmation letter

Andrew Wicks
Department of Computing & Information Systems
Faculty of Architecture, Computing & Humanities
University of Greenwich
QM420, Queen Mary Court
Greenwich Campus

Direct Line 020 8331 8842
Direct Fax 020 8331 8824
Email researchethics@gre.ac.uk
Our Ref UREC/16.4.5.1
Date: 5th December 2017

Dear Andrew,

University Research Ethics Committee - Minute 16.4.5.1

TITLE OF RESEARCH: Content factors which influence the success of a VLE - Parts 1 & 2

I am writing to confirm that the above application has been **approved** by Chair's Action on behalf of the Committee and that you have permission to proceed.

I am advised by the Committee to remind you of the following points:

You must notify the Committee immediately of any information received by you, or of which you become aware, which would cast doubt upon, or alter, any information contained in the original application, or a later amendment, submitted to the Committee and/or which would raise questions about the safety and/or continued conduct of the research;

You must comply with the Data Protection Act 1998;

You must refer proposed amendments to the protocol to the Committee for further review and obtain the Committee's approval thereto prior to implementation (except only in cases of emergency when the welfare of the subject is paramount).

- You are authorised to present this University of Greenwich Research Ethics Committee letter of approval to outside bodies in support of any application for further research clearance.

On behalf of the Committee may I wish you success in your project.

Yours sincerely



John Wallace
Secretary, University Research Ethics Committee

cc. Professor Liz Bacon - FACH
Professor Lachlan Mackinnon - FACH
Dr Cos Ierotheou - FACH

University of Greenwich
Greenwich Campus
Old Royal Naval College
Park Row
London SE10 9LS
Telephone: +44 (0)20 8331 8000

Index

A

Abbreviations i, xviii, xxi
 Abstract 4
 Acknowledgements 3
 Adaptive learning iii
 Aesop 21
 Aims , xi, 3, 154
 Alternative education iii
 Anderson-Darling test 36
 Ausubel, Davis 25, 160

B

Binet, Alfred 17, 160
 Bloom, Benjamin
 Mastery learning 10, 11, 24, 31, 72, 163
 Bruner, Jerome 28, 71, 161, 168
 Brusilovsky, Peter 33, 81, 160, 161

C

Carroll, John 10, 12, 93, 161
 Chabris, Christopher and Simons, Daniel
 invisible gorilla 161
 Chou, Shih-Wei
 shovelware 1
 Computer-based immersive education iii
 Conclusions ii, iv, vi, vii, xi, 8, 46, 67, 90,
 145, 148, 149
 Connectivism and the Medici effect , 29
 Connolly, Thomas 32, 161

D

de Bono, Edward 30
 de Bra, Paul 33, 81
 Deep/surface learning iii
 Design and development 164
 Deterding, Sebastian 32, 76, 162
 Dewey, John 28, 30, 162, 164
 Dose-response curve xx, 122, 124, 126,
 129, 152
 Dunn, Rita and Dunn, Kenneth .18, 20, 162
 Dweck, Carol 21, 78, 81, 162

E

Entwistle, Noel 22, 162, 163
 Ethics 6, 8, 50, 51, 138, 139
 Experiential learning , xv, 16
 Experimental group , 83

F

Felder, Richard and Silverman, Linda 20
 Festinger, John 2, 21, 78, 163
 Feuerstein, Reuven 18, 81, 163
 Figures i, xv
 Flavell, John
 Metacognition ii, xxii, 2, 16, 18, 20, 29, 75, 82,
 163
 Fleming, Neil and Mills, Colleen 18
 Flipped classroom iii
 Flynn, James 18, 161, 163

G

Gagné, Robert
 Conditions of learning 11, 73
 Gee, Paul 32, 164
 Gregorc, Anthony 20, 39, 164

H

Hattie, John
 Visible learning 160, 164, 168
 Hermann, Ned 20
 Hill, Napoleon 20, 164
 Holt, John 11, 23, 164
 Honey, Peter and Mumford, Alan 18, 20,
 39

I

Illich, Ivan 11, 23, 164
 Intelligence , 17, 167

K

Kazimoglu, Cagin 32, 56, 164

Keller, Fred ...11, 12, 13, 24, 30, 55, 70, 72, 165
 Keller, Fred
 Personalised system of instruction 11
 Keller, Fred
 Personalised system of instruction 24
 Keller, Fred
 Personalised system of instruction 72
 Key terms xxii, 4
 Khan, Salman.....5, 31, 46, 165
 Kolb, David
 Experiential learning....., xv, 15, 16, 28, 165
 Kruger, Justin and Dunning, David...16, 23, 79

L

Lave, Jean and Wenger, Etienne 28
 Learning intervention impact metric ii, viii, 7, 91, 155, 156
 Learning intervention impact metric (LIIM)
 ii, x, xii, xix, xxi, 7, 118, 131, 147, 149, 155, 156, 170
 Learning intervention location metric (LILM) ... 4, viii, ix, x, xiii, xv, xviii, xix, xxi, xxii, xxiii, 100, 102, 103, 104, 105, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 123, 124, 125, 131, 132, 133, 135, 137, 141, 142, 143, 144, 145, 147, 149, 151, 152, 153, 156, 170, 194
 Learning intervention spread metric
 Mean (LISM_{Mean}). 4, viii, ix, x, xiii, xv, xviii, xix, xx, xxi, xxii, xxiii, 105, 107, 108, 111, 112, 113, 114, 115, 116, 117, 118, 124, 125, 126, 127, 131, 132, 133, 134, 135, 137, 141, 147, 149, 151, 152, 153, 154, 156, 170, 196
 Median (LISM_{Median}).4, viii, x, xiii, xv, xviii, xix, xx, xxi, xxii, xxiii, 109, 110, 117, 118, 127, 128, 129, 130, 131, 133, 134, 147, 149, 151, 152, 156, 170, 199
 Learning styles 163
 Literature review 5, 10

M

MacKinnon, Lachlan and Bacon, Liz .29, 71, 72, 75
 Marton, Ference and Säljö, Roger.....22
 Maslow, Abraham.....20, 166
 McGregor, Douglas.....20, 166

Methodology ...ii, iv, 5, 6, 49, 50, 121, 139, 166
 Meyer, Jan and Land, Ray.....26, 74
 Michalos, Alex.....22, 166, 168
 Montessori, Maria 23, 166
 Motivation and cognitive dissonance iii
 Myers, Isobel and Briggs, Katherine .18, 20

N

Nat, Muesser Cemal 19, 73, 161
 Normal distribution36, 38, 42, 43, 94, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 131, 134, 137, 142, 150, 152

O

Objectives xi, 3, 4, 154

P

Pask, Gordon 19, 20, 72, 166
 Perkins, David 26, 166
 Piaget, Jean.....15, 81, 166
 Postman, Neil.....23, 166

R

Riener, Cedar and Willingham, Daniel..18, 39, 73
 Role play iii

S

Self-paced learning iii
 Seligman, Martin 21, 167
 Shapiro-Wilk test.....36
 Shen, Luxi and Hsee, Christopher 21, 76, 78, 86
 Siemens, George and Downes, Stephen29, 72
 Socrates
 Socratic method 23
 Spearman, Charles.....17, 167

T

Tables.....i, xviii
 Technologically enhanced learning (TEL) ii, vi, xxi, xxiii, 1, 3, 4, 5, 7, 13, 19, 25, 27, 29, 32, 33, 46, 48, 49, 50, 51, 53, 54, 57,

69, 70, 71, 73, 78, 81, 84, 90, 149, 153,
154, 155, 157, 158, 159, 161
Threshold concepts iii

V

Vygotsky, Lev
Zone of proximal development iii, xv, 25, 26, 27,
71, 74