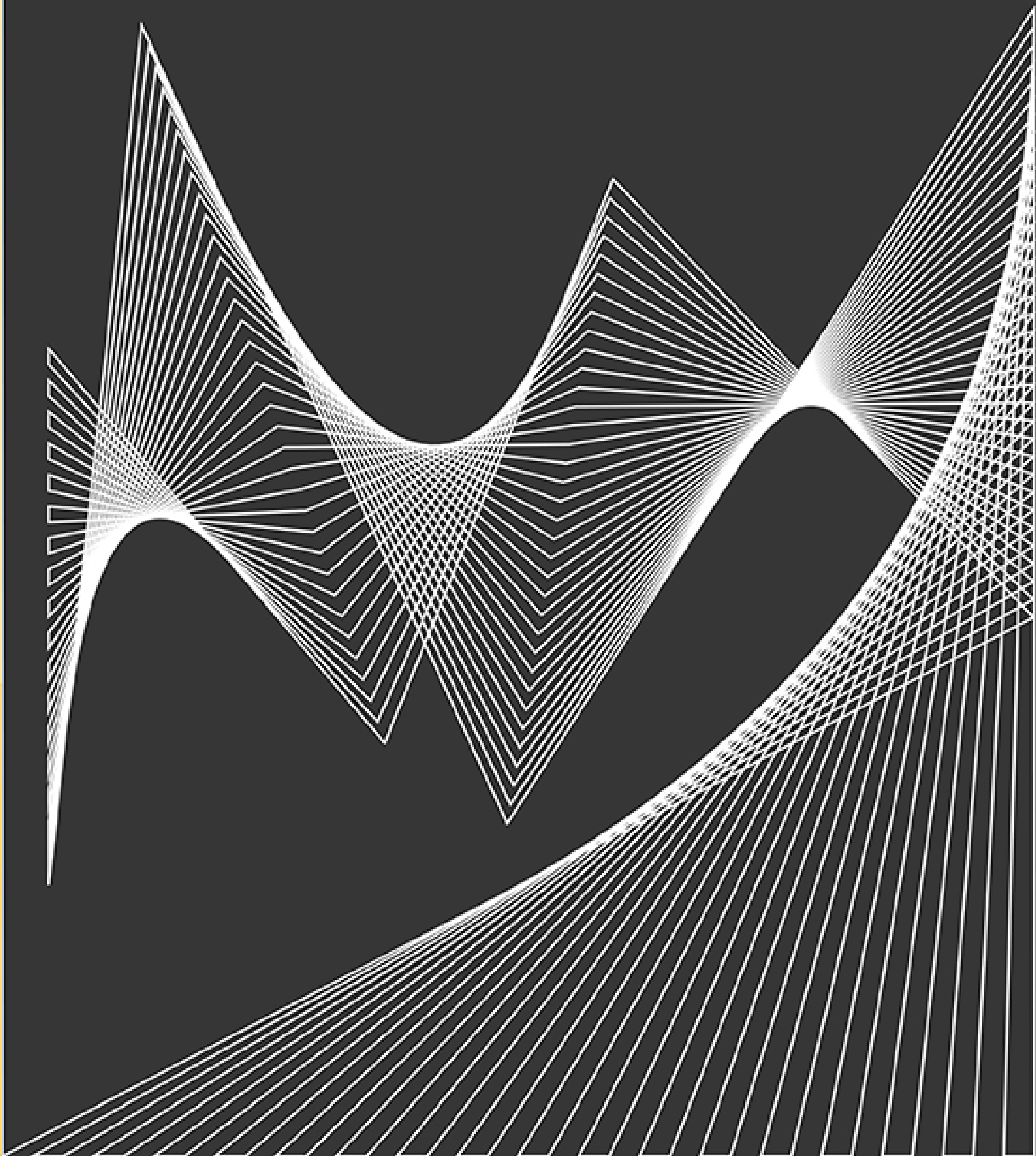
MSOR connections

Articles, case studies and opinion pieces relating to innovative learning, teaching, assessment and support in Mathematics, Statistics and Operational Research in HE.

Volume 16 No. 2



Contents

This journal is published with the support of the **sigma**-network and the Greenwich Maths Centre.

To register for submissions, notifications, and further information relating to MSOR Connections please visit https://journals.gre.ac.uk/index.php/msor

Editorial

Noel-Ann Bradshaw, Department of Mathematical Sciences, University of Greenwich, London, UK. Email: n.bradshaw@gre.ac.uk

Welcome to this edition of MSOR Connections, which once again reflects the broad range of mathematics education activity taking place across Higher Education in the UK.

I am delighted that this edition starts with a report of the important workshop held recently in memory of Professor John Blake, who is so much missed in the mathematics education community. Edited by Grove, the report contains details of the talks given at that meeting which both summarised the enormous contribution that John made to UK HE mathematics education and investigated the next steps going forward. It covers a wide range of topics including employability, maths support, assessment and the HE maths curriculum. This theme of mathematics curricula is taken up by Capes and Rowlett, who discuss the design of a first-year curriculum to aid transition from school to university. This is a subject very close to my own heart as it was my concern for issues with transition and retention which led me to set up the first Maths Arcade at Greenwich and to take a number of other initiatives to help first-year undergraduates. The Maths Arcade is the focus of the next paper, authored by students Golding and Smith from Sheffield Hallam University, who investigate how strategy games can be used to promote mathematical thinking.

We then move to maths support: Mulligan and Mac an Bhaird examine the value of having a full-time maths support worker based in maths support centres. Although this is specific to maths support at Maynooth University, Ireland there are many lessons that can be learned for maths support centres elsewhere, and this paper may be particularly helpful for anyone wishing to make a case for a full-time role at their institution. This is followed by Mitchell's paper detailing one of the roles of a maths support centre, in providing statistics help to non-maths students. Mitchell describes his method of teaching them statistical appreciation rather than trying to teach a couple of statistical methods which are unlikely to be understood. The final paper from Little shows how important statistics support is in the teaching of nurses: something that I was privileged to learn more about during my time organising meetings for sigma.

I am particularly pleased that this edition of Connections resonates so strongly with my own special interests in mathematics education in HE, as this is likely to be the last editorial I will be writing, due to my recent move from HE to industry.

I would like to end by saying how grateful I am to Duncan Lawson, Tony Croft and others for giving me the opportunity of helping resurrect MSOR Connections and, as John Blake urged, to build upon the legacy of the Maths, Stats & OR Network. I have very much enjoyed working with the rest of the editorial team – Peter Rowlett, Alun Owen and Rob Wilson – and I am very pleased to announce that Tony Mann will be joining them. I wish them every success for the future and am confident that Connections will continue to serve the needs of the HE maths community for many years to come.

WORKSHOP REPORT

Where next for mathematics education in higher education?

A one-day meeting in honour of Professor John Blake

Michael Grove, School of Mathematics, University of Birmingham, Birmingham, UK. Email: m.j.grove@bham.ac.uk
Gary Brown, Office for National Statistics, Newport, South Wales. Email: gary.brown@ons.gov.uk
Tony Croft, Mathematics Education Centre, Loughborough University, Loughborough, UK. Email: A.C.Croft@lboro.ac.uk
Stephen Hibberd, School of Mathematical Sciences, University of Nottingham, Nottingham, UK. Email: Stephen.Hibberd@nottingham.ac.uk
Joe Kyle, School of Mathematics, University of Birmingham, Birmingham, UK. Email: J.Kyle@bham.ac.uk
Jeremy Levesley, Department of Mathematics, University of Leicester, Leicester, UK. Email: l1@leicester.ac.uk
Chris Linton, Vice-Chancellor's Office, Loughborough University, UK. Email: C.M.Linton@lboro.ac.uk

Abstract

On the 10 June 2016 Professor John Blake, known to many readers of MSOR Connections for his leadership of the Maths, Stats & OR Network and his passionate support for mathematics education, passed away peacefully following a short illness. He was 69. In recognition of John's work, a number of his friends and colleagues came together to organise a one-day education meeting in his honour that comprised part of a larger event celebrating his many achievements in applied mathematics. The meeting was held in July 2017 at the University of Birmingham, a place where John spent many years of his career and established a legacy for teaching and learning from which many now benefit today. John was a man for action and the purpose of this one-day meeting was to explore the current needs and priorities of the mathematical sciences community and identify mechanisms by which we can continue to work together in a changed, and changing, higher education landscape: this article comprises a record of the thoughts and ideas of those who presented in honour of John's legacy.

Keywords: Maths, Stats & OR Network, mathematics education, teaching and learning, mathematics support.

1. Introduction

For over twenty years, and until his retirement in 2013, John Blake was Professor of Applied Mathematics at the University of Birmingham, where he held the Headship of the School of Mathematics on two occasions; he was also Dean of the Faculty of Science. In addition to his world-leading profile in mathematical research, John was amongst the first true champions of teaching and learning within higher education. He established the U.K. Mathematics Courseware Consortium (MATHWISE), a groundbreaking forerunner of later innovative uses of software in teaching and learning, and went on to lead the Computers in Teaching Initiative (CTI) Centre for Mathematics. In 2000, John became inaugural Director of the highly regarded Mathematics, Statistics & Operational Research (MSOR) Network under the auspices of the national Learning and Teaching Support Network (LTSN). As Director, he shaped a number of important initiatives through the Network (Blake, 2012) and many regard this as his greatest legacy, one that thrives to this day.

John was a passionate and committed teacher delivering a range of undergraduate courses at the University of Birmingham. In his joint foreword to the text *Transitions in Mathematics Education* (Grove et al., 2015), John commented: "*The transmission of our excitement, knowledge and* *understanding to our students is a complex challenge that demands considerable expertise, especially given the different mathematical backgrounds of our students, their diverse mathematical needs and career goals. Students will inevitably face barriers to their learning of mathematics. . .*" There is no doubt, John was fiercely committed to the enhancement of the learning and teaching of mathematics within higher education and the recognition of those involved in such endeavours.

In the now 18 years since the Maths, Stats & OR Network was established, there has been much activity to support learning and teaching enhancement, and innovation, within the mathematical sciences. John was vocal in urging the community to build upon the legacy of the Maths, Stats & OR Network: *"…may you continue to provide leadership in learning and teaching in your respective institutions in the UK, and more widely on the international stage. Do not let our legacy disappear!*" (Blake, 2012). But in recent times, higher education has undergone a period of significant change: with the recent introduction of the Teaching Excellence Framework (TEF), and changes to (i.e. the removal of) the national infrastructure to support discipline-based learning and teaching and the sharing of effective practices, there is now a need for the mathematical sciences community itself to consider the next steps and priorities for mathematics education in higher education and how the community itself may realise these. In his honour, the School of Mathematics at the University of Birmingham, with support from the Institute of Mathematics and Applications, along with John's friends and colleagues organised this one-day meeting to explore what these 'next steps' might be.

What follows are the words of those individuals who contributed to this meeting.

2. The expanding reach of the mathematical sciences and its implications for undergraduate curricula (Chris Linton)

What are university mathematics courses for? For those who have some responsibility for the education of mathematics undergraduates, I think it is valuable to consider this question periodically. Of course, there is no single answer, but nor is it the case that a university education serves only to impart knowledge to students. The answer is also time dependent.

Amongst other things, we need university courses which produce graduates with the appropriate knowledge and skills to fill the jobs that will be in demand over the coming years (a moving target that is difficult to predict) and which reflect the fact that mathematics is becoming pervasive in an ever-increasing number of areas.

There have been many reports written about mathematics over the past twenty years or so. In terms of looking forward, I would recommend the 2013 report *The Mathematical Sciences in 2025* (US National Research Council, 2013). The authors argue that the mathematical sciences enterprise in the early 21st century is qualitatively different to that of the latter part of the 20th.

One of the report's key observations is that the value of the mathematical sciences would be heightened if the number of mathematical scientists who share the following characteristics could be increased:

- They are knowledgeable across a broad range of the discipline, beyond their own area(s) of expertise;
- They communicate well with researchers in other disciplines;
- They understand the role of the mathematical sciences in the wider world of science, engineering, medicine, defence, and business; and,
- They have some experience with computation.

Of course this will only happen if they are exposed to these things. This begs the question: are our undergraduate curricula really suitable for the 21st century?

I think it is vital that students studying mathematics-related degree programmes gain a broad and up-to-date understanding of the uses to which mathematics is being put in the modern world. There are many challenges to doing this successfully. Given the rapid pace of change and the new uses that are constantly being found for mathematics, how do we ensure that lecturers are sufficiently aware to be able to impart this knowledge? How many of us could talk knowledgably about the role mathematics plays in, for example, social science networks; protein folding; climate modelling; computational biology; artificial intelligence; public health; metamaterials; compressed sensing – all areas identified in *Mathematics 2025* as ones where mathematics has a significant role to play.

As an example, consider PageRank, the algorithm used by Google to rank its search engine results. This was developed in the 1990s, bringing together ideas from linear algebra, probability theory and graph theory as well as a lot of computer science. The PageRank algorithm solves a problem which did not exist until recently. I obviously didn't learn about it as an undergraduate in the 1980s. How many of today's mathematics undergraduates learn about PageRank, something that has had a profound effect on their lives? My guess is not many, certainly not enough.

It is an example of a broader problem. Not enough effort is spent on giving mathematics undergraduates a broad and, critically, up-to-date understanding of the expansive reach of the mathematical sciences. If mathematics graduates don't know what mathematics is being used for in today's world, who is going to spread the knowledge to funders, policy makers, schoolchildren?

One of the primary causes of the massive increase in reach of the mathematical sciences over the past few decades has been the exponential increase in computing power. The pervasive nature and power of computer technology has transformed the world and we can now apply mathematics where previously there was little value.

Do we do enough to educate students as to how computers are used to apply mathematics? Do we teach students how to write mathematical software? Do we embrace phenomenology-driven enquiry, in addition to the theorem-proof paradigm? I would argue that the mathematical sciences should more thoroughly embrace computing.

Above all, mathematics graduates need to understand the link between mathematics, the power of computers, and real world problems.

3. Where next for mathematics support? (Tony Croft)

From the early 1990's the phrase "the mathematics problem" came into common usage amongst university academics. The term referred to student under-preparedness for the mathematical demands of their university courses, particularly in engineering and the physical sciences. Students were struggling with university-level mathematics and their lecturers were struggling to deal with, let alone rectify, the situation. Numerous professional bodies and learned societies reported the problem (for example LMS, 1995; Sutherland and Pozzi, 1995).

Universities were making attempts to alleviate this situation and by far the most common response was the development of mathematics support (here 'mathematics support' will be used as a short-hand for the more correct, but longer, 'mathematics and statistics support'.) The term 'mathematics support' means provision which is additional to normal lectures, tutorials and problems classes and the aim of which is to improve the performance of students, particularly but not exclusively those in danger of failing. Mathematics support can take various forms including drop-in centres, appointment-based services, pre-sessional courses and on-line resources. In recent years, the drop-in centre has become the dominant form of provision. Coventry University (1991) and Loughborough University (1996) were two of several institutions to establish large-scale mathematics support facilities. Working together, they have provided models and resources that have been adopted or adapted at many other universities throughout the UK and beyond.

It was around the year 2000 that John Blake, as Director of the LTSN MSOR Network, began his collaboration with the nascent mathematics support community. I, along with Duncan Lawson (Coventry), was a recipient of Network funding to undertake an initial survey of mathematics support provision (Lawson, Halpin and Croft, 2001; Lawson, Halpin and Croft, 2002a) and produce a guide *Good Practice in the Provision of Mathematics Support Centres*, published by the Network (Lawson, Halpin and Croft, 2002b). John's enthusiasm for supporting this community was unreserved. Under John's leadership, the Network re-published resources, for example, the very widely used *Facts & Formulae* leaflet (and others which would follow) distributing these to tens of thousands of students at UK universities. John lent his authority and gravitas to the search for funding to establish a virtual mathematics support centre, which became **math**centre in 2003 (http://www.mathcentre.ac.uk/). Today **math**centre is still a widely used resource base for both university staff and students, housing well over 1000 individual items, including 100+ hours of video tutorials designed to ease the transition into mathematics at university. Moreover, the site is also used by the academic community as a repository for a much broader collection of mathematics teaching and learning resources.

In 2002, the MSOR Network established the MathsTEAM project which produced numerous guides and case studies aimed at addressing the mathematics problem in various ways. One strand of this work was concerned with mathematics support (see http://www.sigma-network.ac.uk/ltsn-mathsteam-project-guides/. Other strands were Diagnostic Testing, and Teaching Mathematics within an Engineering/Science Context). When the **sigma** CETL (Centre for Excellence in University-wide Mathematics and Statistics Support) was established in 2005, which later became the **sigma** Network, John continued to encourage and actively sustain the mathematics support community. During the most recent period of Higher Education Funding Council for England (HEFCE) funding (2013-2016) John, notwithstanding being formally retired, chaired the Advisory Group of the Network until illness prevented him continuing in the final months of the project.

There is little doubt that John's legacy in the field of mathematics support will continue to address the serious and ongoing mathematics problem in higher education. Indeed, the problem has broadened significantly from the days when mathematics support was conceived. Newer reports continue to flag the challenges, increasingly in disciplines other than engineering and the physical sciences. The biosciences form an area where increasing quantification of the discipline is causing concern (ABPI, 2008; BBSRC, undated). Social Science is another field in which quantitative demands are increasing and students need to be better prepared for competing internationally (British Academy, 2012).

As for the future of mathematics support, the case for its continuation has been made above. But more needs to be done. Younger tutors, and especially postgraduates, need to be made fully aware of the background to the mathematics problem so that they appreciate the challenges faced by many of today's students, their diversity of expertise and interests, especially those from non-mathematical background who nevertheless need to become proficient in mathematical and statistical skills; that many are intimidated by the prospect of doing any mathematics at all at university. For these reasons and more, my view is that support centre tutors should be experienced teaching staff who want to devote time and energy to the enterprise; centres should not be solely reliant on postgraduates. Our students deserve more. There should be good career opportunities for those engaged in this valuable work. All too often, though less so nowadays, mathematics support has been a Cinderella service – something seen as 'nice to have' but not essential. Work must continue to educate university senior managers of the need for mathematics support to address the systemic issues highlighted above – to enlighten them just as John was enlightened. Finally, there is room for improvement in university mathematics teaching and if such improvement were forthcoming the need for mathematics support would surely reduce. Knowing John's commitment to the improvement of teaching through his work in the Maths, Stats & OR Network, I am sure this is something he would endorse wholeheartedly.

3. Graduate Employability – where is its place in the mathematical sciences? (Stephen Hibberd)

Graduate outcomes from University courses are emerging as a high priority area and a core feature of the Teaching Excellence Framework (TEF) (HEFCE, 2017). An aim of the TEF is to raise the profile and visibility of teaching within providers along with helping ensure graduates achieve better employability outcomes. Expectations are that the evolving TEF, currently at a developmental stage of subject level (DfE, 2017), will influence academic degree course accreditation and consequently how this might relate to employability. In this context a notable distinction must be made between 'employment' and 'employability'. The former relates to the short-term employment outcomes (see for example UNISTATS, 2017) based upon graduate employment data obtained six months after graduation. Employability on the other hand, relates to enhancing the longer-term ability and resilience of graduates to achieving their evolving career aspirations; something which is better evidenced by data showing more longer-term trends (see for example, WONKHE, 2017).

The proportion of UK school leavers going to university has now increased towards 50% and as highlighted by Waldock and Hibberd (2015), the curricula of many mathematics programmes remain designed primarily for the preparation of undergraduates for further study and research. Such degree outcomes may not sufficiently incorporate the primary aspirations of the increased and varied cohort taking mathematics degree to boost their broader career prospects. Institutions are gradually evolving from a traditional approach to higher education, with the principal learning outcomes being focussed on the acquisition of subject knowledge, skills and their application. Often additional skills are available from extra-curricular sources, such as those provided by the university careers services or other optional components. Other institutions may embrace the concept of developing student employability to support student recruitment and potentially more directly the local needs of the business and the industrial community.

Globally universities are increasingly expected to provide a broad range of outcomes in graduates to meet the expectations of students, their parents, employers and government. Correspondingly individual mathematics courses are constantly evolving to meet the local strengths and expectations of their students and their own university frameworks; in this latter regard, graduate employability is increasingly becoming an important and potentially distinguishing factor. Examples of professional competencies that could be more comprehensively developed and incorporated within a mathematics degree are provided within Table 1. Identified are four key areas of graduate competency, together with potential sample activities, that could be incorporated within a curriculum at different levels and stages; the final column identifies the personal outcomes a student might experience or attain.

Connecting academic study with skills development is a growing expectation for the transition between higher education and employment. A core emphasis within an undergraduate mathematics degree is commonly on individual student activity through lectures, problem workshops and assessment dominated by examination. This approach provides an efficient and effective mathematics subject-specific approach encouraging routine core understanding and skills. More recently, graduate expectations are increasingly focused on broader and enhanced employability skills and the subsequent expectation of significant personal career progression. Meeting such a changing emphasis encourages extending the learning outcomes for mathematics degree students.

Project activities are widely identified as a valuable skills-rich component of a mathematics degree programme but are often restricted by concerns over implementation issues such as extensive provision of individual projects, assessment and feedback. Concerns may seem intensified for group projects, however the potential for enhancement of skills and peer learning are considerable and appropriate to most career environments. At the University of Nottingham *Vocational Mathematics* modules are synoptic modules developed to enhance subject-specific knowledge and mathematical skills in areas of applied, financial or statistical modelling. Group project tasks build on core studies obtained in the earlier years of a mathematics degree in tandem with developing graduate level team tasks. Development plenary and strand workshops give exposure to skills awareness and group project activities offer essential insight into tackling unseen problems while adhering to tight deadlines. Project tasks also require relevant mathematical background research, integrating the use of mathematical and presentational software to successfully communicate quantitative ideas orally and in compact reports (posters, oral presentations and written reports) to a professional standard. Key assessment features of the modules are the development of clear grading templates, equally available for student group self-assessment to guide their project development and formal submission, and for academic staff to allow comparative assessment and provision of detailed and informative detailed feedback. For further details of this approach at the University of Nottingham, see Hibberd (2011).

	Definition	Sample Activities	Desired Outcomes
Co-ordinating with Others	Working inclusively and effectively together to collectively achieve a common goal	• Group projects • Seminars • Group presentations	• Understanding your impact and contribution within the team • Recognising the contributions of others • Commitment to achieving a common goal
Digital Capabilities	• Information, data and media literacies • Digital learning and development • Digital creation, problem solving and innovation, • Digital communication, collaboration and participation	• Projects that involve the use of appropriate presentation and statistical software • Using digital resources • Presenting and interacting over digital platforms	• Understanding effective uses of digital technologies • Confident to interact with others through digital platforms • Agility and willingness to embrace and use digital communications and technologies
Professional Communication	Ability to communicate effectively and appropriately through a variety of means, including oral presentations and communication in other languages	• Briefing report • Project dissertation • Placements • Blogs • Poster • Presentations	• An understanding of professional expectations • Ability to communicate in a clear, positive and impactful way to different audiences • How to structure and deliver relevant (technical) content • Effective delivery (e.g. use of visual aids, use of time)
Reflection	Consideration in order to develop enhanced understanding and insight in relation to professional outcomes and areas/ opportunities for further enhancement	• Post activity reflection • Self-evaluation • Action planning	• Greater self-awareness • Willingness to receive and act upon feedback • To be a reflective practitioner for self improvement

Figure 1: Professional competencies applicable to a mathematical sciences curriculum

The universal provision of substantial project activities remains an underdeveloped area of the undergraduate mathematics curriculum. Such activities have the potential of promoting individual study, research and employability skills in students and of highlighting the versatility of mathematics graduates. This is particularly relevant at a time when a number of external influences are indicating that degree specifications should embrace an extended range of subject specific and wider skills together with more feedback to students on assessments. Following the recommendation of the *Wakeham Review* (Wakeham, 2016) for professional bodies to support universities to deliver high-level STEM skills, appropriate mathematics specific modules with a strong employability element are encouraged and may now also formally contribute as part of the mathematics content towards Institute of Mathematics and its Applications (IMA) Programme Accreditation (IMA, 2017).

4. Scholarship – what's the point? (Michael Grove)

For many years the mathematical sciences community benefitted from a series of national initiatives, projects and networks each with the aim of enhancing the quality of the student learning experience in mathematics and disciplines where components of mathematics are taught. John Blake was involved in some way with all of them: through establishing, and then leading, the Maths, Stats & OR (MSOR) Network (2000-2012), something which many regard as his greatest legacy, defining the mathematical activities that would comprise part of the National HE STEM Programme (2009-2012), and his most recent Chairing of the Advisory Committee of the **sigma** Network (2012-2016).

John was not only an excellent teacher, but was committed to the creation and dissemination of effective learning and teaching practices; most significantly, he strongly believed in providing discipline-based opportunities for others to do the same. All who are involved in teaching and learning within higher education should be *scholarly*, that is collecting data and evidence for use by themselves to verify and enhance their own teaching knowledge and practice. Many extend this to *scholarship,* collecting data and evidence which is then shared with a like-minded (disciplinary or institutional) group who help verify this knowledge and use it to bring about enhancement and improvement. Others make the transition to *research* where data and evidence is collected to inform a much wider, and often public audience, and where verification may take place outside of the context of a single discipline.

There are many reasons for an individual to engage in scholarship. At the heart is the commitment to improving student learning and enhancing educational quality, but for an individual it can help them become part of a community, build an academic identity, and receive recognition for their endeavours. Being part of a community also benefits our institutions: through the collaborations that develop, and the sharing of ideas and effective practices. This will be particularly important with the Teaching Excellence Framework (TEF) being piloted in 2017/18 at a disciplinary level; if it continues to progress in this way, we will all be stronger as departments by working together.

The mathematical teaching and learning initiatives, projects and networks that existed from the late 1980s helped individuals make the transition to scholarship and become part of a community; in particular this growing appetite of those within higher education to collect data and evidence was something that I, working with Tina Overton, was able to witness first-hand through the National HE STEM Programme in particular (Grove & Overton, 2013). However since 2016, the external funding that helped sustain such communities and networks has ceased. If we wish to maintain and grow the learning and teaching community that so many of us continue to value, we need to organise ourselves and move forward without external support. Maintaining visibility is important because we need to ensure that we, as a community, are ready to capitalise on any national funding opportunities that may arise in the future. So just how do we move forward?

To begin, we need a way of communicating information effectively with all who are interested in teaching and learning in the mathematical sciences and in a way that allows newcomers to join. But there is no need to start from scratch, for example **sigma** has established a national network for those involved in mathematics and statistics support and the Institute of Mathematics and its Applications (IMA) has its Academic Representatives Network. However there are things we can readily do, for example, a JiscMail email distribution list for mathematics teaching and learning more generally is an easy and quick place to start.

We need to think about how we bring people together to share ideas and build collaborations. There is a role for our learned societies: the London Mathematical Society already offers an annual 'Education Day' and the IMA is offering support to departments to run teaching and learning workshops in 2018 through a dedicated initiative. More broadly, departments could run their own teaching and learning events and open these up to those from outside. For over 10 years the annual CETL-MSOR Conference has been a mainstay in the mathematics teaching and learning calendar; due to community support from those within England, Scotland, Wales and Ireland, it will return to its traditional September timeslot in 2018.

Publication is an important part of scholarship. MSOR Connections, the hugely popular and accessible teaching and learning journal established by the MSOR Network has restarted and is now distributed through the University of Greenwich. Both the MSOR Network and **sigma** published occasional guides on topical issues; these might be continued on an ad hoc basis as a community-led initiative and published electronically or made available through print-on-demand services, possibly by utilising support from the educational grant schemes of the learned societies to assist with design costs. At the very least, we must ensure the vast array of legacy publications remain available.

Scholarship is also about aiding and supporting the development of others, particularly as they begin their careers; this was something at which John himself excelled. The *Induction Course for New Lecturers* and the workshops for *Postgraduate Students Who Teach Mathematics and Statistics* were both important and highly valued activities of the MSOR Network. We need to find a way to re-establish and sustain them, and in doing so ensure they remain 'owned' by the community, much as **sigma** has done with its workshop for mathematics support tutors. Even further, with teaching focused routes becoming increasingly recognised across the sector we need to help those looking to extend their scholarship into educational research.

Here I have only outlined a start; there is much more that needs to be done, but there are already things upon which we can work together to ensure the legacy of John, and his work, is maintained.

5. Assessing the Value of Assessment (Jeremy Levesley)

When I first started to engage with the mathematics education community 25 or so years ago, it was no time before I was introduced to John Blake, Director of the LTSN MSOR Network. He was one of the few people who managed to span the research and teaching divide in university mathematics, and did so very successfully. I have looked up to few people in my time (and of course the number decreases with my age and general cynicism; see below) but John was one of those. When our 'parents' die, we become the grown ups, so now we are faced with the unenviable task of trying to continue his legacy.

I would like to talk about assessment, as I believe it is the tail which wags the dog of education in our society. This is not just a problem for education, but this will be the place where we experience it most keenly. TEF is a system by which the government can understand whether or not it has delivered the task of 'higher education'. REF (the Research Excellence Framework) is an assessment of whether or not 'research' has been done. The notion of assessment has been propagated (I believe) by a bureaucratic class which creates more and more opportunities for assessment, thus proliferating itself. This tendency has infected universities and their administration.

Assessment is the measurement tool by which we understand whether or not we have completed our task. We face the challenge of giving feedback to students so that they may understand whether or not they have completed their task. The main reason that students continually complain about feedback is that there is no authentic ‘task’. An authentic task is one in which someone can say for themselves whether or not they completed it. For instance, I know I have cleared the high-jump bar because it does not fall off. The question to my coach is how I can learn to jump higher. Since an exam does not have any internal mark scheme, I cannot tell a student how to get 70%. Our assessment tasks need to be broken down and aligned with easy to elaborate learning outcomes. Then we can say “Yes”, “No”, or “Nearly”, as conclusions of assessment.

We heard an interesting anecdote on the day of the workshop. A non-university colleague was talking about how they had given students some data and to come back with some sort of report on their conclusions. Some of them were fine, and others were not. To those that were not, he asked them to make them better. The university in question asked where the mark scheme was, and that there should be more detailed feedback. The external in question may not bother to work with that university again. If the students do not care enough about the task to understand what ‘a good report’ meant, all the feedback in the world is not going to help.

Our current education system, with its fixation on feedback, is infantilising all of the learners. They should be learning to develop their own quality compass, and that should be one of the main things that we are wishing to see. We should not be hoping that they manage to hurdle a bar that has come from my imagination, but should be being encouraged to set their own expectations, and to learn to shoot for the moon. In this world they will fail continuously, as we all do, in striving for the best. I have students who know that to get their first they can afford to get 52% in my module, and so they do. I do not blame them for this. This is a problem that we have carefully colluded in creating. Should we continue, or should we aim to address the problem?

What might a solution look like? First of all we should jettison the classification system, and develop a set of real tasks that students can complete. Make a computer programme which solves nonlinear equations. If it does not work, you do not pass. There are no ‘method marks’ because these are for early stage learners where the process is more important than the outcome. I do not get method marks for my papers, or for the delivery of my classes.

We need a deep set of agreed capabilities, decided at a national level, so that we have a standard. I believe that industry should help provide this (they have members of staff already who work to recruit our graduates into the workplace). The degree should have multiple opportunities for students to attempt skills, and the assessors can be anyone qualified (no need to be an academic or even a member of the university). Of course, those involved in quality assurance are going to say, *“We cannot do this, where are the standards?”* I am going to reply, prove to me that all of your paperwork has created any proper standard! The view of the world of work on the resilience of students leaving university gives the lie to this objection.

We need to make a stand against the quality assurance community. It is the child of the unproven notions of management for performance that have been eschewed by the business world to a large extent. Let us seek their input on how to run the universities. We need to push back on the creeping illness that, in the guise of Ofsted (the Office for Standards in Education, Children's Services and Skills), has stolen the spirit of our secondary teachers, and in the camouflage of TEF, will take ours too.

6. Even bigger challenges than Big Data (Gary Brown)

My work at the Office for National Statistics (ONS) is to ensure official statistics conform to the principles of the *Code of Practice for Official Statistics* (UK Statistics Authority, 2009), especially principle 4 - sound methods and assured quality. The data revolution has provided all analysts - in public and private industry - with a wealth of data previously unimaginable. Using these data for public good, successful developments such as Uber, Google Translate, and smart cars have transformed many areas of our lives - my job at the ONS is to assess whether they can also transform the statistics used by government and industry to measure the social and economic health of the nation.

We can definitely use the new data science techniques to help us innovate - data scientists analyse data to find a question (through induction) and statisticians find which data can help answer that question (deduction). Both these approaches are needed - exploration and confirmation - as they complement each other perfectly, "*having only one is madness*" (Tukey, 1980).

Big data offer a different view on the world, but to use them in the production of official statistics requires old views to be assured - value can only follow veracity, regardless of volume, variety and velocity. This requires data scientists and official statisticians working closely together - which we are doing - and ensuring the mathematics community value both disciplines, and provide the training needed for the future. We are actively involved in many courses - to ensure their relevance - and encourage the whole community to focus on these future needs, and help us target and train the workforce for the future.

The value of the academic community and mathematics teaching profession cannot be over-emphasised in the future needs of official statistics. Only by working in partnership with higher education can we fully exploit the data revolution, and ensure our statistics are fit for the 21st century.

7. Blue Skies Thinking in a Cloudy Climate – Making Sense of it All (Joe Kyle)

Had I said to John Blake that I was about to give a talk with the above title, he would have asked what it was about, probably approved (with a few caveats) and then offered the strong encouragement to find a better title – one less journalistic and cliché-ridden. And how I wish I still could rely on that blunt, kind and generous man for good, no-nonsense helpful advice; my work would be so much the better for it.

Having said that, let me turn to my reflections on the topics that exercised me in our recent meeting. The greatest piece of Blue Skies thinking I currently indulge in is a desire to turn the clock back – a not uncommon wish in people of certain age. But I don't mean simple nostalgia. Nor, as mathematicians are often heard to expound, do I think there was ever a 'golden age' in the teaching and learning of mathematics. No, I'm thinking of something far more structural. I was fortunate to have enjoyed undergraduate study at a time when education was regarded as an investment any civilised nation undertook to develop a better society (at much the same time, on the other side of the world, John was embarking on his own undergraduate journey). Now education has become a commodity to be bartered and used as a vehicle to burden young people with debt. This absurd use of 'the market' has resulted in recent graduates facing interest charges of some ten to twelve times the Bank of England base rate. I can see no justification for this and reversing this situation would be the one Blues Skies thought I would love to see come true.

7.1. Problem-solving – "oh, that's when it's in words."

These days, we hear talk of "problem-solving" so often, in so many walks of life, that there's a real danger that it will become a term devoid of real meaning, in my view. Within the realm of the mathematical sciences, though, it is the beating heart of our discipline and there are authors who take a scholarly and serious-minded approach to the topic. But all too often it is reduced to such banal strategies as "read the problem twice" (why not thrice?) and "underline the question". To be fair, I was never convinced that the greatly admired Pólya (1945) offers much substance: 'formulate a plan and carry it out' seems a little lacking in depth, to my eyes.

Given all this, I don't blame the young teacher I recently talked to who, when asked what she understood by problem-solving replied: "oh, that's when it's in words." At least she had formulated something that she could take into the classroom. My second Blue Skies ambition would be to see problem-solving discussed in more mature terms, and not as if it were some advertising slogan.

7.2. Experts and expertise

John Blake was an expert; both as an academic and as an expert witness for legal proceedings. The idea that "people [in this country or elsewhere] have had enough of experts" would be quite simply belief-beggaring.

The great C. P. Scott (1921), a towering figure as editor of the Manchester Guardian, where his tenure straddled both the nineteenth and twentieth centuries, formulated the most important dictum for all reporters: "*Comment is free, but facts are sacred*". My concern is that the cloudy climate of our times has produced an atmosphere where comment is sacred and facts are 'fake'. But my Blue Skies optimism leads me to believe that we will emerge from this dark age of debased discourse and will, as a civilisation, learn from the experience. Our students are exposed to much information and data; we have an obligation, through the mathematical sciences provision we deliver, to equip them with the skills and abilities to enable them to form their own judgements on not only what is real and 'fake', but more importantly to ask questions and challenge perceptions of what is valid and reliable. Our graduates will be a key part of the 'data science' revolution that currently seems to be taking place and such skills will be essential.

7.3. A Concrete Proposal

As we know, John Blake's supervisor at Cambridge was the precocious, and in many ways eccentric, Professor Sir James Lighthill – founder and first President of the Institute of Mathematics and its Applications (IMA). It is now nearly 40 years ago since the publication, under the aegis of the IMA with Lighthill as editor, of *Newer Uses of Mathematics* (Lighthill, 1978). The preface tells us that the purpose of the book was to introduce new applications to "*quicken the interest*" of the intended readership; the intention was to allow research and teaching to benefit from each other – principles that guided John Blake in all his endeavours. Might it not be a fitting memorial to John Blake to round up the leading applied mathematicians of the day to produce a sequel? After all, a common theme that has emerged from this meeting is how we might need to rethink learning outcomes for the mathematical sciences in light of ensuring our graduates are adequately prepared for the diverse range of careers they now enter in the 21st century.

8. Conclusion

While those of us who knew John, and regarded him as a friend, no doubt wish that this meeting might have been held under very different circumstances, we hope that we did the great man proud. Who knows what he would have made of our presentations and the subsequent discussions. He was a modest man but one who had his own views and was not afraid to share them, just like those who presented and attended in his honour. While we can only speculate as to whether John would have approved, we know that he would certainly have supported our spirit for sharing our ideas and for our commitment to enhancing the profile of teaching and learning within the mathematical sciences in higher education!

Although there are too many to list, we are enormously grateful to everyone who attended or who enabled this fitting tribute to John to take place, but particular thanks are due to Dave Smith from the

University of Birmingham for his support and willingness to include an educational event as part of a much larger celebration of John's academic work.

6. References

ABPI (2008). *Skills needed for biomedical research*. London, UK: Association of the British Pharmaceutical Industry.

Blake, J.R. (2012). Reflections from a past Director (2000-8). *MSOR Connections*, 12(2), pp. 34-35.

BBSRC (undated). *The Age of Bioscience: Strategic Plan 2010-2015*. Swindon, UK: Biotechnology & Biological Sciences Research Council.

British Academy (2012). *Society Counts: Quantitative Skills in the Social Sciences and Humanities*. London, UK: British Academy.

DfE (2017). *Teaching Excellence Framework: Subject-level pilot specification*. London, UK: Department for Education.

Grove, M.J., Croft, A.C., Kyle, J. and Lawson, D.A. (Eds.) (2015). *Transitions in Undergraduate Mathematics Education*. Birmingham, UK: University of Birmingham and Higher Education Academy.

Grove, M.J. and Overton, T.L. (Eds.) (2013). *Getting Started in Pedagogic Research within the STEM Disciplines*. Birmingham, UK: University of Birmingham and The Higher Education Academy.

HEFCE (2017). *The TEF.* Available via http://www.hefce.ac.uk/lt/tef/ (last accessed 30 December 2017).

Hibberd, S. (2011). Integrative Use of Group Projects. In Robinson, M., Challis, N.,Thomlinson, M., Eds. *Developing Graduate Skills in HE Mathematics Programmes - Case Studies of Successful Practice: Reflections on experience, practice and provision*. Birmingham, UK: Math, Stats and OR Network, pp. 18-19.

IMA (2017). *University Degree Programme Accreditation.* Available via https://ima.org.uk/university-degree-programme-accreditation/ (last accessed 31 December 2017).

Lawson, D., Halpin, M. and Croft, T. (2001). After the diagnostic test: what next? Evaluating and enhancing the effectiveness of mathematics support centres. *MSOR Connections*, 1(3), pp. 19-23.

Lawson, D., Halpin, M. and Croft, T. (2002a). After the diagnostic test: what next? Evaluating and enhancing the effectiveness of mathematics support centres part 2. *MSOR Connections*, 2(1), pp. 23-26.

Lawson, D., Croft, T., Halpin, M. (2002b). *Good Practice in the Provision of Mathematics Support Centres*. Birmingham, UK: The Maths, Stats & OR Network.

Lighthill, J. (Ed.) (1978). *Newer Uses of Mathematics*. London, UK: Penguin Books

LMS (1995). *Tackling the Mathematics Problem.* London, UK: London Mathematical Society, Institute of Mathematics and Its Applications and the Royal Statistical Society.

National Research Council (2013). *The Mathematical Sciences in 2025*. Washington, DC: The National Academies Press. https://doi.org/10.17226/15269.

Pólya, G. (1945). *How to Solve It*. New Jersey, USA: Princeton University Press.

Scott, C.P. (1921). *A Hundred years*. Available via https://www.theguardian.com/sustainability/cp-scott-centenary-essay (last accessed 31 December 2017).

Sutherland, R. and Pozzi, S. (1995). *The Changing Mathematical Background of Undergraduate Engineers.* London: The Engineering Council.

Tukey, J.W. (1980). We Need Both Exploratory and Confirmatory. *The American Statistician*, Vol. 34, No. 1.

UK Statistics Authority (2009). *Code of Practice for Official Statistics* (Edition 1.0). London, UK: UK Statistics Authority, pp. 23 – 25.

UNISTATS (2017). *UNISTATS: The official website for comparing UK higher education course data*. Available via https://unistats.ac.uk (last accessed 30 December 2017).

Wakeham, W. (2016). *Wakeham Review of STEM Degree Provision and Graduate Employability.* Available via https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/518582/ind-16-6-wakeham-review-stem-graduate-employability.pdf (last accessed 30 December 2017).

Waldock, J. and Hibberd, S. Chapter (2015). Employability for the workplace. In Grove, M.J., Croft, A.C., Kyle, J. and Lawson, D.A., Eds. *Transitions in Undergraduate Mathematics Education.* Birmingham, UK: Higher Education Academy and University of Birmingham.

WONKHE (2017). Graduate salaries 5 years after graduation (2008/9 cohort) – Mathematical Sciences. Available via http://wonkhe.com/wp-content/uploads/2017/06/Screen-Shot-2017-06-13-at-16.12.06.png (last accessed 31 December 2017).

RESEARCH ARTICLE

Essential Concepts for the First Year of Study for BSc Mathematics

Alexander Capes, Sheffield Hallam University, Sheffield, United Kingdom. Email: alex.capes@hotmail.co.uk.
Peter Rowlett, Sheffield Hallam University, Sheffield, United Kingdom. Email: p.rowlett@shu.ac.uk.

Abstract

To inform discussion about content for the first year of undergraduate mathematics, a study was completed which reviewed: the A-level Mathematics specification; published literature on the transition from A-level to university mathematics; the second and third year curricula of modules at three English universities with different foci. This aimed to investigate what students might reasonably be expected to have covered when they arrive at university, what happens in practice at the transition to university, and the role of the first year as preparation for later study. Content suggestions focus on calculus, linear algebra and analysis as core topics. There is also evidence of the need to focus on students' understanding of where formulae and solutions originated as well as their ability to produce pieces of academic and mathematical writing. Findings also include suggestion that what happens in the first year, while similar between institutions, does depend on the overall focus of the degree programme.

Keywords: undergraduate, mathematics education, transition, Further Education, Higher Education.

1. Introduction

Although a reasonable amount of research exists regarding issues faced by students transitioning from Further to Higher Education in the field of mathematics, there is little discussion of concepts that are key during the first year of study for a BSc Mathematics course. This article is designed to provide some discussion in this area, through analysis of the Department of Education GCE Mathematics specification (DoE, 2016), review of relevant literature, and examination of BSc Mathematics courses at three UK universities, selected to provide a varied basis for discussion. Readers should note the anglocentricism of this article, as the university courses selected for discussion are based in England, and the Further Education analysis uses only the GCE A-level specification, in lieu of potential alternatives such as the International Baccalaureate.

2. Method

The research for this project aimed to examine three key elements that must be considered when discussing BSc Mathematics first year curricula: the material studied directly prior to the first year of study (in the UK, Further Education most commonly takes the form of A-levels); the differences in material, focus, student independence and other factors between further level mathematical education, and higher level; and the curricula at three UK universities, specifically selected for maximum contrast in their focuses on pure and applied topics in mathematics, covering as broad a scope as possible without focusing too heavily on this area.

2.1. Review of A-level syllabus

Preliminary research involved an in-depth analysis of the Department for Education specification for A-level Mathematics, to ascertain the level of knowledge that can be assumed of university starters. A-levels are the UK standard qualification of Further Education (typically studied for between the ages of 16 and 19), and the vast majority of BSc Mathematics courses at UK universities require an

A-level (or equivalent) qualification in Mathematics, ordinarily at grade C minimum. This analysis formed a basis for the rest of the research undertaken, as it highlighted the areas of mathematics that are studied in depth, and areas that potentially require additional support during the first year of undergraduate study.

2.2. Literature Review: Issues in Transition from Further to Higher Education

A literature review covers material from over 50 years of research into the difficulties faced by students of mathematics at the further and higher levels, providing an overview of opinions of some experts in this field. Additionally, some examination of the differences between Further and Higher Education is included to contextualise any comparisons thereof.

2.3. Review of Second and Third Year Curricula at Three UK Universities

The third part of this article is a comparison of the second and third year curricula at three UK universities for their respective BSc Mathematics courses, the aim of which is to identify some mathematical topics and skills that are required to be covered in the first year of higher level study as preparation for the second and third years. The QAA Benchmark Statement (QAA, 2015) defines mathematics degrees on a spectrum from “theory-based” to “practice-based” and allows a broad range of options for the focus of a degree. The universities chosen will remain unnamed, and shall be hereon referred to as University A, B and C, respectively. University A is a research-intensive, pre-1992 university, with Mathematics courses ranked in the top five on entry tariff (Guardian League Table, 2018), whose BSc Mathematics course features a strong focus on pure mathematics. University B is a research-intensive, pre-1992 university with Mathematics courses ranked around 20th on entry tariff (Guardian League Table, 2018), whose BSc Mathematics course aims to balance the focus between pure and applied topics. Finally, University C is a post-1992 university with Mathematics courses ranked bottom 10 by entry tariff (Guardian League Table, 2018), whose Mathematics course focuses strongly on applied mathematics.

3. A-level Syllabus Review

The A-level syllabus splits the content of the Pure Core modules into ten discrete sections (DfE, 2016), which are as follows: Proof, Algebra and Functions, 2-Dimensional Coordinate Geometry, Sequences and Series, Trigonometry, Exponentials and Logarithms, Differential Calculus, Integral Calculus, Numerical Methods, and Vectors.

In addition to the basic concepts of Proof, the syllabus names proof by exhaustion, deduction, contradiction, and disproof by counterexample, as areas of study to be covered. Although students must learn the principles of these techniques, it could be suggested that more emphasis on rigorous and sound mathematical reasoning when constructing proofs would benefit those students continuing into higher level mathematics.

The Algebra and Functions section of the syllabus aims to build some foundations of algebraic manipulation methods required for further study of calculus and analysis; the syllabus covers a wide breadth of topics, including surds, algebraic manipulation of polynomials and first- and second-degree inequalities, laws of exponents, proportional relationships, partial fractions, and the manipulation of functions and their graphs.

In Co-ordinate Geometry, students can be expected to study straight-line equations, circular geometry, and parametric equations. These topics provide foundations for work using ordinary and partial differential equations, and general work in areas of analysis, such as measure theory.

The Sequences and Series section provides essential introductions to concepts such as sums to infinity (used in a wide variety of contexts in analysis), the general idea of infinite series/sequences, and iterative formulae. Additionally, students are introduced to sigma notation, and other classical notation.

The Trigonometry section of the syllabus goes beyond the basics of sines, cosines, and tangents, and covers their reciprocals and inverses, various trigonometric identities, and double angle formulae. Some proofs involving trigonometric functions are also included on the syllabus. This understanding of trigonometry is integral in the study of any topic involving waves, signals, and/or analytical methods.

The main direct precursors to undergraduate level calculus and analysis (and any topics containing differentials and integrals) on the syllabus are the Exponentials and Logarithms, Differential Calculus, and Integral Calculus sections. Students are taught the basics of differentiation and integration, including the chain, product, and quotient rules of differentiation, and integration by parts and substitution. The syllabus also mentions the Fundamental Theorem of Calculus. Linking this knowledge with exponential functions and logarithmic laws gives students the basis to study most higher topics with strong calculus components. Continuing the work on integration, students are introduced to numerical methods of solution for definite integrals (using Simpson's Rule and the Trapezium Rule), widely used in a variety of applied topics at undergraduate level.

Finally, the syllabus covers Vectors in 2- and 3-dimensions; although the syllabus sticks to the basics of vectors (running through vector notation, component and magnitude/direction forms, the geometric implications of position vectors, and basic vector arithmetic), these concepts are imperative for any later study involving the complex plane, and/or vector calculus.

4. Literature Review: Issues in Transition – Further to Higher

Questions may be raised about the efficiency and depth of the previously discussed curriculum, when considering the importance of concepts behind proof and calculus (Prendergast et al., 2017), and the changing standards of assessment at A-level. For example, Epstein (2013) notes that assigning work based on derivations of formulae and justifications thereof, rather than algebraic drills, has proven to greatly improve students' understanding of the topic of calculus, and demonstrated benefits for students who have continued into tertiary study. Epstein also notes that - while a direct cause-and-effect relationship cannot be established between the assignment of these tasks and an increase in technical prowess - due to the strong positive correlation, further discussion is warranted. By contrast, it is noted that some students view mathematics as a "rote learning activity" (Nardi and Steward, 2003; p. 362); a problem also highlighted in school inspections (Ofsted, 2012).

Lawson (1997) notes that in a study carried out by Coventry University, in which undergraduate students were asked a series of questions based on principles that are covered in the A-level curriculum, the attainment levels of students with A, B and C grades were virtually indistinguishable from one another. This raises the question of whether A-level grades are a worthwhile indicator of students' later ability to study mathematics at higher level.

As noted in *Tackling the Mathematics Problem* (LMS, IMA and RSS, 1995), issues surrounding A-level programmes and their efficacy in preparing students for undergraduate study have been of concern for governmental (Burghes, 1990) and non-governmental (Osmon, 2013) organisations since the 1990s. Moore (1994) suggests that the minimal focus on the fluent mathematical writing is detrimental to students' abilities to construct formal proofs, which makes a strong case for the introduction of some measure to remedy this imbalance. Additionally, *What Maths do you need for University?* (Osmon, 2010) indicates that the absence of linear algebra in the A-level curriculum puts students at a clear disadvantage, given the range of mathematical topics requiring its usage.

Burton and Haines (1997) suggest that the burden of fixing the issues in transition lies with the designers of the A-level specification, implying that a higher level of contextualisation of techniques would aid students in their general understanding of topics such as calculus and analysis. A strong understanding of differential calculus is noted by Biza et al. (2016) as a vital prerequisite for undergraduate level study of mathematics. Abdulwahed, Jaworski and Crawford (2012) support this, suggesting that earlier introductions of contextualised calculus problems would potentially give students a more meaningful understanding of these subject areas, whilst also granting them greater insight into the sort of mathematics covered at undergraduate level.

Kalajdzievska (2014) showed that students creating their own questions and mark-schemes for areas of calculus demonstrated a lack of understanding of the fundamental reasoning behind the uses of calculus in applied mathematics. This suggests that giving students a greater focus on the applications of calculus at an earlier point in their education would potentially benefit them by the time they reach undergraduate level. Ellis et al. (2015) highlighted the need for a keen understanding of the principles of calculus in first-year undergraduate mathematics study, due to the prominence of calculus throughout BSc Mathematics courses; this lends credence to the notion that calculus must form a substantial element of the first year of any BSc Mathematics course, due to the reliance on it in most further module areas.

Another issue in transition is that of students' lack of technical and mathematical fluency in their exams and written assignments. The results of a study by Stylianou, Blanton and Rotou (2015) into undergraduates' understanding of proof and its surrounding concepts posit that students do understand the importance of proof and well-written mathematical arguments, however they are not encouraged enough through the means of assessment to improve their own mathematical communication skills. Solomon (2006) suggests that a lack of contextualised material at A-level hinders students' understanding of the importance of mathematical communication; for example, due to the lack of a coursework element, students are almost never required to construct long-form mathematical arguments, supported by their own mathematical assertions, as is the case at degree level. It is also suggested in a paper by Hoyles, Newman and Noss (2001) that critical evaluation of mathematics problems is an area of poor understanding for many undergraduate mathematics students, which demonstrates the need for change at some level to address this issue.

Iannone and Simpson (2011, 2012) have posited that there is a fundamental lack of students' ability to explain their mathematical ideas in an academic manner, which is suggestive of a requirement for more universities to place emphasis on these issues as early as possible in their BSc Mathematics programmes. This assertion is supported by Sofronas et al. (2015), who note that calculus is used as a unifying thread in the first year of undergraduate teaching, and the contextualisation of calculus as a pure mathematical concept is directly beneficial to students in this regard. This ties back into the aforementioned requirement for students' to possess a keen understanding of the applications of pure mathematics and the theories behind calculus and analysis, as noted by Jaworski, Mali and Petropoulou (2017), who state the importance of explaining the reasoning mathematical principles to students.

It should be noted that all of the studies with experimental components mentioned here contained relatively small sample sizes, rendering their validity somewhat questionable in this context. Additionally, due to the inherently subjective nature of topics in pedagogy, there is likely to be confirmation bias present in some of the sources named. In spite of these elements, there is still clearly a strong case for a need for change at some level to address the growing concerns surrounding the transition from A-level study, to degree level study.

5. Examination of Selected University Curricula

As stated previously, part of this study involves an examination of the second- and third-year modules of the BSc Mathematics courses at three UK universities, selected for the different focuses of their courses, in order to provide a strong and clear contrast. In spite of this contrast, it should be clear that there are several common required skills and technical areas of knowledge across the three courses. It should be noted that the sample size has been kept deliberately small, to enable a more thorough examination of each university in the time available for this study.

Firstly, the BSc Mathematics course at University A focuses heavily on pure mathematics, with students expected to carry out rigorous and extensive proofs of a wide variety of subject areas, and utilise both analytical and numerical methods of solution for the real-world contexts, when they are mentioned. More so than with the other mentioned university courses, students of this course must clearly possess a strong level of competence in terms of differential and integral calculus, as well as confidence in proof construction, algebraic discourse, a variety of forms of mathematical notation, and academic writing with a substantial mathematics component as a whole.

The BSc Mathematics course at University B features a more balanced programme in terms of pure and applied mathematics, with students having the option to determine how much they want to lean in either direction; this element of choice is particularly noteworthy in this context, as the university must ensure students are equally capable in a variety of areas during their first year of study, such that they can excel in their second and third years of study in more specialised areas. The majority of modules on this course require a clear understanding of the principles of calculus and analysis, along with a strong focus on producing professional-looking work with a sound basis in fact and mathematical reasoning. There is a slight emphasis on the study of mathematical physics, with one core module in each of the second and third years of study centring around topics in it. In accordance with typical scientific modules at universities, these modules require strong analytical skills, and excellent objective writing abilities, backed up by well-grounded mathematical arguments, presented in a clear, objective manner. Finally, there is also an emphasis on the proper construction of proofs, such that several core modules contain elements of proof construction as either a focus, or a sidebar. These proof-based tasks clearly also require the ability to construct strong, sound mathematical arguments and present them in a way that is professional and clear to the reader.

Finally, the BSc Mathematics course at University C has a strong focus on applied mathematics over pure mathematics. Nonetheless, students are still required to produce work demonstrating a clear and present understanding of analytical methods of solution, as well as utilising differential, integral and vector calculus in an applied mathematics context - such as in final year modules focused on fluid mechanics and modelling using partial differential equations. Additionally, students are required to produce professional and academically credible pieces of mathematical writing, which clearly necessitates a keen understanding of the principles of mathematical fluency in communication – a common theme throughout all the modules analysed, with the exception of the final year project, which can of course take very different directions on a student-by-student basis.

6. Discussion

This article aimed to inform discussion of curriculum developments at first year undergraduate level by examining what students who studied A-level can be expected (in principle) to have covered before they arrive at university, what happens (in practice) at the transition to university and the role of the first year as preparation for later study.

While the A-level syllabus covers a wide range of important material, this investigation suggests that too little emphasis is placed on ensuring that students understand where formulae and solutions originated, with too much focus on learning by rote; arguably (e.g. Epstein, 2013) a major cause of students' difficulties in making the transition to university level mathematics from A-level.

The examination of later years of study at three universities highlights the variety of practice in UK higher education mathematics degrees. All three courses clearly meet the standards set out by the QAA Benchmark Statement (QAA, 2015), despite having considerable differences. This invites the suggestion that the first year of a university maths degree should differ according to the nature of the degree course. Despite differences, it is noted that students studying for any of these courses would be required to have strong understanding of calculus, linear algebra and analysis. The evidence examined here also highlights the need to produce pieces of academic and mathematical writing. It may be concluded from this study that these topics are considered key elements of a wide variety of mathematics undergraduate courses. Good (2011) reports on an exercise where representatives of UK mathematics departments were asked to "list the top five topics they felt a mathematics undergraduate must not graduate without knowing". Note that Good's focus is on the whole degree, not just the first year. The most common mathematical choices reported by Good were topics around analysis, calculus and linear algebra, and the exercise reported also highlighted "the ability to communicate mathematics" among other valued graduate attributes. The findings of this present study, then, support those reported by Good.

It should be noted that the study here is based on examination of a small number of university courses and a review of some relevant literature. However, a picture has emerged of measures that could be taken to assist students with the transition to becoming university mathematics students. The evidence reviewed in this study leads to the conclusion that the first year of an undergraduate curriculum could place greater emphasis on creating and communicating clear mathematical arguments. This would aid students with key issues around transition and provide a strong basis for later study.

7. References

Abdulwahed, M., Jaworski, B. and Crawford, A., 2012. Innovative approaches to teaching mathematics in higher education: a review and critique. *Nordic Studies in Mathematics Education,* 17(2), pp. 49-68.

Biza, I., Giraldo, V., Hochmuth, R., Khakbaz, A., Rasmussen, C., 2016. *Research on Teaching and Learning Mathematics at the Tertiary Level.* http://dx.doi.org/10.1007/978-3-319-41814-8

Burghes, D., 1990. A-level Mathematics: A Time for Change? *Teaching Mathematics and its Applications,* 9(3), pp. 97-101. http://dx.doi.org/10.1093/teamat/9.3.97

Burton, L., Haines, C., 1997. Innovation in teaching and Assessing Mathematics at University Level. *Teaching in Higher Education,* 2(3), pp. 273-293. http://dx.doi.org/10.1080/1356215970020308

Department for Education, 2016. *Mathematics AS and A Level Content*. London: Department for Education.

Ellis, J., Hanson, K., Nuñez, G., Rasmussen, C., 2015. Beyond Plug and Chug: an Analysis of Calculus I Homework. *International Journal of Research in Undergraduate Mathematics Education,* 1(2), p. 268-287. http://dx.doi.org/10.1007/s40753-015-0012-z

Epstein, J., 2013. The Calculus Concept Inventory – Measurement of the Effect of Teaching Methodology in Mathematics. *Notices of the American Mathematical Society*, 60(8), pp. 1018-1026. http://dx.doi.org/10.1090/noti1033

Good, C., 2011. 'We can't let them graduate unless...' In: P. Rowlett, ed., 2011. *HE Mathematics Curriculum Summit*. Birmingham, U.K.: Maths, Stats and OR Network. pp. 15-16.

The Guardian, 2017. *Guardian University League Table 2018.* [online] Available at: https://www.theguardian.com/education/ng-interactive/2017/may/16/university-league-tables-2018 [Accessed 15 October 2017].

Hoyles, C., Newman, K., Noss, R., 2001. Changing patterns of transition from school to university mathematics. *International Journal of Mathematical Education in Science and Technology,* 32(6), pp. 829-845. http://dx.doi.org/10.1080/00207390110067635

Iannone, P. and Simpson, A., 2011. The summative assessment diet: how we assess in mathematics degrees. *Teaching Mathematics and its Applications*, 30(4), pp. 186-196. http://dx.doi.org/10.1093/teamat/hrr017

Iannone, P., Simpson, A., 2012. Oral assessment in mathematics: implementation and outcomes. *Teaching Mathematics and its Applications*, 31(4), pp. 179-190. http://dx.doi.org/10.1093/teamat/hrs012

Jaworski, B., Mali, A., Petropoulou, G., 2017. Critical Theorising from Studies of Undergraduate Mathematics Teaching for Students' Meaning Making in Mathematics. *International Journal of Research in Undergraduate Mathematics Education*, 3(1), pp. 168-197. http://dx.doi.org/10.1007/s40753-016-0044-z

Kalajdzievska, D., 2014. Taking Math Students From "Blah" to "Aha": What can we do? *PRIMUS*, 24(5), pp. 375-391. http://dx.doi.org/10.1080/10511970.2014.893937

Lawson, D., 1997. What can we expect from A-level Mathematics Students? *Teaching Mathematics and its Applications,* 16(4), pp. 151-156 http://dx.doi.org/10.1093/teamat/16.4.151

London Mathematical Society (LMS), Institute of Mathematics and its Applications (IMA), Royal Statistical Society (RSS), 1995. *Tackling the Mathematics Problem*. Available at: http://mei.org.uk/files/pdf/Tackling_the_Mathematics_Problem.pdf [Accessed 15 October 2017].

Moore, R., 1994. Making the transition to formal proof. *Educational Studies in Mathematics,* 27(3), pp. 249-266. http://dx.doi.org/10.1007/BF01273731

Nardi, E. and Steward, S., 2003. Is Mathematics T.I.R.E.D? A Profile of Quiet Disaffection in the Secondary Mathematics Classroom. *British Educational Research Journal*, 30(3), pp. 345-367. http://dx.doi.org/10.1080/01411920301852

Ofsted, 2012. *Mathematics: made to measure. Messages from inspection evidence* (Report No. 110159). Manchester: Office for Standards in Education, Children's Services and Skills.

Osmon, P., 2010. What maths do you need for university? *Proceedings of the British Society for Research into Learning Mathematics,* 30(2), pp. 41-46. Available at: http://www.bsrlm.org.uk/wp-content/uploads/2016/02/BSRLM-IP-30-2-08.pdf [Accessed 15 October 2017].

Osmon, P., 2013. A-level mathematics reform: satisfying the requirements of university courses across the range of mathematical subjects. *Proceedings of the British Society for Research into Learning Mathematics,* 33(1), pp. 43-48. Available at: http://www.bsrlm.org.uk/wp-content/uploads/2016/02/BSRLM-IP-33-1-08.pdf [Accessed 15 October 2017].

Prendergast, M., Faulkner, F., Breen, C., Carr, M., 2017. Mind the gap: an initial analysis of the transition of a second level curriculum reform to higher education. *Teaching Mathematics and its Applications*, 36(1), pp. 1-15. http://dx.doi.org/10.1093/teamat/hrw024

QAA, 2015. Subject Benchmark Statement: Mathematics, Statistics and Operational Research. Available at: http://www.qaa.ac.uk/en/Publications/Documents/SBS-Mathematics-15.pdf [Accessed 15 October 2017].

Sofronas, K., DeFranco, T., Swaminathan, H., Gorgievski, N., Vinsonhaler, C., Wiseman, B., Escolas, S., 2015. A Study of Calculus Instructors' Perceptions of Approximation as a Unifying Thread of the First-Year Calculus. *International Journal of Research in Undergraduate Mathematics Education,* 1(3), pp. 386-412. http://dx.doi.org/10.1007/s40753-015-0019-5

Solomon, Y., 2006. Deficit or Difference? The Role of Students' epistemologies of Mathematics in their Interactions with Proof. *Educational Studies in Mathematics*, 61(3), pp.373-393. http://dx.doi.org/10.1007/s10649-006-6927-1

Stylianou, D., Blanton, M., Rotou, O., 2015. Undergraduate Students' Understanding of Proof: Relationships Between Proof Conceptions, Beliefs, and Classroom Experiences with Learning Proof. *International Journal of Research in Undergraduate Mathematics Education*, 1(4), pp. 91-134. http://dx.doi.org/10.1007/s40753-015-0003-0

8. Acknowledgements

This work resulted from Alex Capes' undergraduate dissertation, supervised at Sheffield Hallam University by Peter Rowlett. Funding for Alex to adapt his dissertation into this article was provided by an Education Grant from the Institute of Mathematics and its Applications (EG12/2017), for which we are grateful.

RESEARCH ARTICLE

Use of board games in higher education literature review

Edward Smith, Sheffield Hallam University, Sheffield, UK. Email: b4010674@my.shu.ac.uk
Lewis Golding, Sheffield Hallam University, Sheffield, UK, Email: b5020351@my.shu.ac.uk

Abstract

There is a long standing history of mathematical games and puzzles being used to support mathematical thinking and concepts. The Maths Arcade is a recent initiative designed to provide an environment for students to play games in order to communicate with fellow students whilst developing a range of skills. The purpose of this study was to determine how these skills are developed and how they link to different games, we have found the evidence for this through reviewing a variety of sources focussing on game-play. We found that the playing of board games is fundamentally linked to the development of mathematical thinking and skills. These are developed in many ways including trial and improvement and discussion.

Keywords: games, mathematical thinking, strategic thinking, problem solving, university.

1. Introduction

The Maths Arcade was designed as a drop in session for universities, to provide a safe space for students to get to know other students and lecturers, whilst also providing them with the opportunity to develop strategic thinking and problem solving skills. Students had the chance to play various strategic and challenging board games, designed to develop these skills (Bradshaw, 2011). For more information on the Maths Arcade see Bradshaw and Rowlett (2012). This review was carried out in order to argue the connection that the playing of strategic board games helps to develop these skills. Using our findings there is great potential to further develop the Maths Arcade to include games created solely for enhancing these skills.

Two or more player strategy games can be traced back to before 2000 BC, and are still being created today (Averbach & Chein, 2000). Whilst most games are created to be a fun past time, there are an increasing number of board games which include mathematical skills. It is hard to distinguish between games which involve a mathematical way of thinking and those which don’t, however it is clear that some games are more mathematical than others (Silva, 2011). For the Maths Arcade games were chosen which display these mathematical qualities in order to improve the strategic thinking and problem solving skills of students. This article will discuss the need for these skills, and the evidence linking the playing of board games to the development of these skills.

2. Methodology

Relevant papers on games and their use in higher education to develop mathematical/logical thinking were found through searches in the Sheffield Hallam University library gateway and Google Scholar. In an attempt to ensure that relevant studies weren't missed, the keywords "games or puzzles or board games or strategy games" and "logical thinking or strategic thinking or mathematical thinking or problem solving", and "higher education or tertiary education or university" were used.

Additionally, as each paper was reviewed, the references it cited were studied and papers which cite it, found using Google Scholar, were reviewed.

In total 20 relevant papers were identified and these were all read and reviewed. It was found that some of these also were not relevant due to focussing on game theory, which is not the topic of interest here, and these were rejected. Papers that were rejected were looking at online or video games rather than physical games. Papers focused at school level were included only if they focused specifically on problem-solving or strategic thinking.

As papers were reviewed, common themes became apparent, and quotes showing evidence for these skills being developed through the playing of games were recorded against these themes as appropriate. The remainder of this paper shares these themes.

3. History of games being used to educate/develop skills

The use of games for educational purposes can be traced to the use of war games in the 1600s, where, according to Gredler (2004), their purpose was to improve the strategic planning of armies and naval forces. Gredler recognises how games are used less, as students move up through education, quoting a remark by Rieber that "although educational games are accepted in elementary school, teacher and parent interest in their use declines in the later grades". This can be observed in higher education at degree level, where the style of learning is often replaced by more traditional lectures. Whilst it may be the case that, as more complex topics are studied, it becomes more difficult to incorporate this into games, there is no evidence to suggest that they can’t still be used as a method in order to develop other skills and techniques.

4. Why mathematical/strategic thinking is important

It is widely recognised that the teaching in UK schools often focuses on students' ability to compute mathematical solutions in an exam style situation (Ofsted, 2012) rather than on their ability to understand and apply the mathematics being taught. This can lead to a reduced emphasis on strategic and mathematical thinking. Stein, et al. (1996) summarise this by saying "increased emphasis is being placed not only on students' capacity to understand the substance of mathematics but also on their capacity to "do mathematics". This has an effect when students reach higher levels of study as these skills are not as developed as much as is necessary for the advanced topics studied in higher education. This was recognised by Burton (1984) who says “few pupils leave the school system with mathematical success as measured by examinations, and those who do consistently surprise their university tutors by their lack of facility in thinking mathematically”. Students need to develop these skills when they reach a higher level of study as they are required for further study and graduate-level employment. The Maths Arcade provides the opportunity for students to do this whilst in a fun and enjoyable environment. An article by Forsyth (2012) discusses a news report in 2012 about the then Norwegian minister for international development, Heikki Holmas who is a regular and experienced player of board games. This highlights how the playing of board games has helped him to develop the strategic thinking skills required in his job. This is further evidence suggesting that the playing of board games is a valuable exercise, developing the skills needed for employment - many job interviews in fact ask for evidence of good problem solving skills.

5. How strategic thinking is developed

The use of game-play allows participants to develop a strategy for success over the long term through the outcomes of the previous games played. Creation of a winning pattern can often include subconscious aspects. There are various generic factors involved when developing a winning strategy during game-play. Games that do not share many common similarities often affect separate strategy making skills, depending on the level of skill involved. The use of trial and improvement is key to strategy development and tabulating results and routes taken is effective to analyse and improve the quality of game-play. Oldfield (1991b) notes that “the value of some form of record keeping is quickly apparent, and pupils are encouraged to develop a system". Trial and improvement skills are directly coupled with proving skills, justifying the level of success of previous actions. The player becomes more aware of which decisions lead to the highest probability of success as experience and comprehension of the game increase. Reasoning is one skill that can be developed through game-play, where one devises a system to consider all viable options available at the current time in order to gain the most benefit out of their turn. This is largely considered a background process that is not often verbally discussed without prompting. Klein and Freitag (1991) studied the effects of using an instructional game on motivation and performance, they found that whilst using instructional games as a method of delivering practice enhanced the motivation of students' it did not contribute to their enhanced performance when compared with traditional methods of practise. This shows that motivation alone is not strong enough to significantly enhance the performance of a student, but greater results may be achieved by combining the effect of instructional gameplay with the enhancement of other skills such as communication skills.

6. How game-play enables communication skills

Discussion of mathematical ideas enables the thought processes involved to be expressed, analysed and assessed. Stimulating such discussion through the means of game-play allows us to construct comparisons for how different games influence different aspects of mathematical thinking. Oldfield (1991a) states that in traditional classroom settings "pupils were struggling to express what they were doing", which shows us that there is a need for a different approach in the way pupils are taught to express mathematical thoughts and ideas. Mathematical understanding can be developed through the discussions between students whilst playing the games (Bragg, 2003) indicating that the playing of games is an effective alternative to the normal teaching methods. It could be argued therefore that the use of games should be implemented in a learning environment, in order to encourage outward communication of mathematical thoughts and to develop discussion skills.

The type of game can offer alternative forms of discussion and therefore potentially different mathematical thinking mechanisms. Working as a group allows communication throughout the course of the game between participants whereas working alone against other individuals will more commonly stimulate discussion after the game is complete (Oldfield 1991a). The mathematical thinking and discussion involved during the game is different to that involved retrospectively. This is because the discussion after gameplay will be influenced by what could have been done differently, rather than the discussion during a game focussed on what should be done presently. If two groups are working on the same problem, this could stimulate discussion in a more competitive sense. This was noted by Herbert and Pierce (2004) who comment that "competition was an important factor in the success of the program. Students formulated ever more complex strategies as they strived to win games".

7. Mathematical thinking in an academic environment

Students in an academic setting who have become accustomed to the typical teaching methods of a learning environment are an excellent set of participants for studying their mathematical thinking, especially so for those that already have a great baseline knowledge of mathematics. Herbert and Pierce studied the effect of game-play amongst a select set of gifted lower secondary school students in a controlled environment. They noted that gifted students are not necessarily highly motivated and so still require positive reinforcement and engagement in order for a long-term gain of skill development to be made. When students were tasked with creating their own games, it was shown that prior game-play experience was vital in applying new concepts. It was noted that allowing the students to create their own games enabled them to demonstrate their ability to use the ideas found in other games to develop an original, functioning game (Herbert & Pierce, 2004). This serves as direct evidence that experience is correlated to the learning of concept and strategy. Herbert and Pierce found like Oldfield that game-play includes heavy incorporation of an "if-then analysis" of many possible alternative moves showing that repetition of this analysis will solidify such skill (Oldfield 1991b). Fitting with our ideas that game-play enhances mathematical concepts Herbert and Pierce note that strategy games have the potential to extend students' higher order thinking.

8. Games being used for younger age groups

There is evidence to suggest that using games enables students to develop problem solving and a deeper understanding of concepts. Ernest (1986) discusses a teaching experiment by Biggs in 1985 which involved the use of games in the education of 7-13 year olds. "Dr Biggs observed not only that these children's conceptual understanding deepened, but that their problem solving abilities grew." It should be noted that the games included in this experiment by Biggs may not be of the same style as the games included in university Maths Arcades, but there is no evidence to suggest that these findings would not be the same for different types of games. Ernest concludes his article by suggesting that games should be included in the mathematics curriculum and that they "have a vital part to play in aiding pupils' achievement and success in mathematics." Whilst games are clearly influential and effective in the education of younger children, the evidence relating this to higher levels of education is limited. The Maths Arcade can be viewed as a tool to use this evidence from lower education and apply this in a higher education context in order to develop these skills further. Ernest (1986) also discusses the work of two Americans, Bright and Kraus, who discovered that the playing of mathematical games can stimulate problem solving skills which can in turn help to develop them through the development of strategy as the games are played. In the study, it is found that this is the case for both experienced and inexperienced problem solvers.

9. Choice of games

The games chosen as part of the Maths Arcade are games which have little to no element of chance or luck. The benefit of this is that to be successful, "players must develop better strategies than their opponents" (Herbert & Pierce, 2004). The competiveness amongst peers playing the games leads to discussions and development of strategy as the games are played more and more over time. When observing students playing board games in the library, Alvarez notes that students leave the table each week strategizing more successful outcomes and "looking forward to the next challenge" (Alvarez, 2017). The games used as part of the Maths Arcade challenge the students across a wide range of skills, enabling all of these to be developed. In fact Alvarez notes that critical thinking, problem solving, oral and written communication and analysing information can all be developed during the playing of games. Some, if not all of these, are essential to students whilst still in education and when in employment after education.

10. Conclusion

We have confirmed through the analysis of the works found, that the playing of board games enhances skills and attributes needed in the study of mathematics. The Maths Arcade houses a variety of games needed to develop these skills. These include problem solving skills, trial and improvement, strategic thinking and mathematical discussion. We have found that game-play allows these skills to develop and improve. There is an abundance of evidence to show that playing games at a young age can help with learning and understanding but this is seldom at a higher level of education. Strategic thinking and problem solving are essential to accessing higher levels of employment but often at higher education these skills are found to be lacking. The Maths Arcade is intended to provide an ideal opportunity for development of mathematical thinking in an enjoyable environment.

11. Acknowledgements

The work of the authors as undergraduate student researchers was funded by Sheffield Hallam University through a Teaching Enhancement Grant. The work was supervised by Peter Rowlett and Claire Cornock.

12. References

Alvarez, V., 2017. Engaging Students in the Library Through Tabletop Gaming. *Knowledge Quest,* 45(4), pp. 40-48.

Averbach, B. & Chein, O., 2000. Games of Strategy for Two Players. In: *Problem Solving through Recreational Mathematics.* Mineola, New York: Dover Publications, Inc, p. 213.

Bradshaw, N., 2011. The University of Greenwich Maths Arcade. *MSOR Connections,* 11(3), pp. 26-29.

Bradshaw, N. & Rowlett, P. eds., 2012. *Maths Arcade: stretching and supporting mathematical thinking.* Birmingham: Maths, Stats & OR Network.

Bragg, L., 2003. *Children's perspectives on mathematics and game playing.* Geelong, Victoria, MERGA Inc. , pp. 160-167.

Burton, L., 1984. Mathematical Thinking: The Struggle for Meaning. *Journal for Research in Mathematics Education,* 15(1), pp. 35-49.

Ernest, P., 1986. Games. A Rationale for Their Use in the Teaching of Mathematics in School. *Mathematics in School,* 15(1), pp. 2-5.

Forsyth, E., 2012. Learning Through Play: Games and Crowdsourcing for Adult Education. *APLIS,* 25(4), pp. 166-173.

Gredler, M. E., 2004. Games and Simulations and Their Relationships to Learning. In: *Handbook of research on educational communications and technology 2.* Mahwah NJ: Lawrence Erlbaum Associates Publishers, pp. 571-581.

Herbert, S. & Pierce, R., 2004. Gifted and lifted higher: And exploration of the development of higher order thinking skills of gifted students played strategy games. *TalentEd,* 22(1).

Klein, J. D. & Freitag, E., 1991. Effects on Using an Instructional Game on Motivation and Performance. *Journal of Educational Research,* 84(5), pp. 303-309.

Ofsted, 2012. *Mathematics: made to measure. Messages from inspection evidence (Report No. 110159).* Manchester: Office for Standards in Education, Children's Services and Skills.

Oldfield, B. J., 1991a. Games in the Learning of Mathematics Part 2: Games to Stimulate Mathematical Discussion. *Mathematics in School ,* 20(2), pp. 7-9.

Oldfield, B. J., 1991b. Games in the Learning of Mathematics Part 3: Games for Developing Strategies. *Mathematics in School ,* 20(3), pp. 16-18.

Silva, J. N., 2011. On Mathematical Games. *Journal of the British Society for the History of Mathematics,* 26(2), pp. 81-104.

Stein, M. K., Grover, B. W. & Henningsen, M., 1996. Building Student Capacity for Mathematical Thinking and Reasoning: An Analysis of Mathematical Tasks use in Reform Classrooms. *American Educational Research Journal,* 33(2), pp. 455-488.

CASE STUDY

The Role of a Full-time Mathematics Support Tutor

Peter Mulligan, Department of Mathematics and Statistics, Maynooth University, Ireland. Email: peter.mulligan@mu.ie
Ciarán Mac an Bhaird, Department of Mathematics and Statistics, Maynooth University, Ireland. Email: ciaran.macanbhaird@mu.ie

Abstract

In September 2016, Maynooth University appointed a Mathematics Support Centre (MSC) University Tutor on a three-year contract, a unique position for an institution on the island of Ireland. In this paper, we briefly explore the important role of MSCs and their tutors. We describe the University Tutor's activities for the 2016-17 academic year and we discuss the benefits of establishing such a position. We also advocate for the establishment of more contract positions.

Keywords: Tutor, Full-time, Mathematics Support, Higher Education, student retention, STEM.

1. Introduction

The issue of student retention in science, technology, engineering and mathematics (STEM) and student mathematical difficulties in the transition from second to further or higher education have been widely publicised for many years (Lawson et al., 2012; HEA, 2017). In an effort to counteract this attrition at a local level, Maynooth University opened the Mathematics Support Centre (MSC) in September 2007. Since then, the MSC, which is based in the Department of Mathematics and Statistics has grown steadily, from 2,493 student visits in 2007-8 to 16,624 in 2016-17, and is one of busiest MSCs in Ireland or the UK. The range of services it provides has also grown, from 18 hours of drop-in per week in 2007-8, to now offering 27 drop-in hours per week, on-demand workshops, online courses and resources, and a drop-in service for local secondary school students (from ages 12 to 18 approximately). As a consequence of this rapid growth, staffing levels and demands on staff have also increased.

In Maynooth University, time and time again, either through feedback provided at the end of each MSC visit or through an annual anonymous survey of their MSC experience, students single out the MSC tutors for praise. This is not unique to Maynooth University, as identified by the Irish Mathematics Learning Support Network (IMLSN) report on a large multi-institutional student evaluation of Mathematics Learning Support (MLS) in Ireland (IMLSN, 2014). This report highlighted the key role that quality tutors play in students' experiences of MLS (IMLSN, 2014: p.30). The importance of tutors in MLS was also the central theme of the IMLSN's 2016 annual conference (Pfeiffer et al., 2016). However, for several reasons, including the precarious nature of the employment that many MSCs have to offer, it is of little surprise that tutors typically move on to other careers after working a few years in MLS. Furthermore, the majority of MSC tutors, at least in Maynooth University, work only a few hours each week, and these factors can negatively impact on the consistency of service that is provided to students.

This recurring problem was brought to the attention of Maynooth University's senior management through the MSC's annual reports and emphasised at various meetings. In 2016, Maynooth University took a very positive first step towards addressing this issue by creating the position of "MSC University Tutor".

In this paper, we begin by giving a brief description of the tutors who work in the MSC and their employment situation. The majority of the paper focusses on the first year of the Maynooth University MSC University Tutor (the first author), and we close with a discussion on the need for full-time MLS tutors.

2. The Tutors

When the MSC opened in 2007-8, the MSC Director (the second author) and one other tutor covered all 18 drop-in hours, along with trying to maintain records of student attendance and registration. As a result of the high number of attendees, additional funding for more staff was requested and received. In 2015-16 (the year before the appointment of the MSC University Tutor), all MSC teaching involved the Director and 22 tutors, who each covered between 2 and 16 hours per week. With the exception of the MSC Director, all other MSC staff, including an administrator, were employed on an hourly basis.

2.1 Who are they?

In Maynooth University, MSC tutors are often sourced from mathematics or statistics postgraduates and to a lesser extent, from final year undergraduate students. Due to the demanding nature of postgraduate or indeed undergraduate courses in mathematics and statistics, a tutor may only be available to work a few hours per week in the MSC. On occasion, the MSC has recruited graduate staff who may not have studied in Maynooth University at all. It is also common for some tutors, on completion of their studies, to commit to working in the MSC (while also giving tutorials for the Department) for an academic year (or more), in order to save money while deciding their career options. Individuals in this situation, often cover considerably more MSC hours than tutors who are studying.

2.2 The current employment situation.

The casual, per hour employment, which seems to be commonplace for MLS tutor positions, especially in universities in Ireland, if not elsewhere (IMLSN, 2016), has benefits for both tutors and the employer.

- Freedom to choose hours (to some degree).
- Can work around classes if studying in the university.
- No/little commitments for either party.

However, there are also many downsides for both the tutor and the employer. The sourcing and retention of suitable tutors is a major problem for MSCs. Lack of suitable staff can inhibit an MSC's ability to evaluate, review and extend the range of services they provide. The high turnover of staff can also mean the loss of senior tutors and the crucial experience and key skills that they have developed working in an MSC. This can have a negative impact on the quality and consistency of service for students. The tutor problem was recognised in an audit of MLS provision on the island of Ireland in 2015 (IMLSN, 2016: p.12). One of their recommendations was

> *"Given the significant reliance on undergraduate and postgraduate students as tutors and the associated transience within MLS, institutions and the IMLSN should promote the role of a MLS tutor and explore the concept of longer-term contracts for tutors to ensure these positions are more secure."*

Maynooth University has been very lucky to have had 3-4 tutors (including the first author until his appointment as MSC University Tutor) who have worked on an ongoing basis in the MSC. However, without a proper contract, it is expected that such tutors will not stay working long term for the MSC. These tutors have no guaranteed or consistent income, no permanency/tenure (even though they may be working "full-time" hours indefinitely, no pay scale, sick pay or pension entitlements. Furthermore, these tutors can be on a low hourly rate of pay (compared to the hourly rate for a second level teacher) and often work very long hours to compensate for periods when there is no work available (holiday periods).

3. The MSC University Tutor

We will now provide some details of the MSC University Tutor role. We also briefly describe the (first author's) experiences of working in this position for the first year of the contract.

3.1 The Role.

The three-year contract contained a variety of duties to be completed during an academic year, primarily:

- Tutor 20 hours per week in the MSC.
- Lecture the preparatory week of mathematics at the start of the Certificate in Science programme.
- Lecture the Summer Mathematics Course for mature students.
- Various administration tasks as assigned by the MSC Director.

The first author successfully applied for this position through a competitive public process in the summer of 2016 and commenced work in September 2016.

3.2 The First Year

Even though the first author had been tutoring in the MSC for many years, he was unaware of the extensive work required to prepare the MSC for each new academic year. Prior to his appointment, these responsibilities fell largely to the MSC Director (the second author). The main item to attend to, at the beginning, was assisting the MSC Director in drafting the MSC tutor timetable. With approximately 120 tutor hours per week to fill, this is not a straightforward task. While the MSC does have a dedicated space, which allows some degree of flexibility, the MSC timetable is heavily reliant on the university academic timetable to determine the best slots for the drop-in service. Also many of the tutors need the academic timetable to determine when their classes are. Furthermore, there is considerable negotiation with academic departments over the availability of some of the more experienced tutors. The countless e-mails and phone calls to source tutors, fill every slot on the MSC timetable, and accommodate subsequent changes, took three weeks from the first to the final draft. During this period there were, of course, more routine but important jobs, such as replenishing handouts and various forms in the MSC, putting up posters, organising MSC lecture announcements, payroll for the tutors, updating the various MSC website pages and, where appropriate, signing up tutors to Moodle, the University's Virtual Learning Environment (VLE).

At the start of the second week, the focus quickly turned to teaching the preparatory week of mathematics for the Certificate in Science students. This Certificate in Science programme is aimed at mature students wishing to return to education. This course starts a week earlier than the standard university lectures in order to give the students a head start. Before securing the MSC University Tutor position, the first author had lectured the full mathematics component of this course, and therefore was familiar with its content.

Preliminary discussions with the MSC Director, prior to the opening of the MSC, highlighted that he was not content with the existing level of tutor training. It was decided to review the extensive tutoring training resources already available in both the UK (Croft and Grove, 2011, 2016) and Ireland (Fitzmaurice et al., 2016) and tailor these resources to meet the needs of the MSC in Maynooth University. A half day training session was held in late September and a follow up session was held later in the first semester. One of the main outcomes of these sessions was that tutors wanted to know about all feedback received from the students which mentioned them by name. As a result, on a weekly basis throughout the academic year, student comments were responded to by the first author and comments which mentioned specific tutors were relayed (anonymously) to tutors by the MSC Director. In previous years, due to time constraints, feedback was communicated in a general sense to the entire tutor team and the MSC Director only dealt with individual tutors on the rare occasion that there were negative comments.

By early October, the semester was well under way, the MSC was fully operational and weekly tasks were reasonably similar. While most of the time was taken up with teaching in the MSC, there were many other duties such as organising workshops (the MSC runs four different types of workshop per week), updating the MSC Facebook and Twitter accounts (social media is used to help publicise workshops and other MSC events), keeping track of tutor hours for payroll (monthly payment documents for each tutor), monitoring handout levels (the MSC have almost 100 handouts, some developed locally, most of which are available on the MSC website), as well as any other ad hoc tasks that occurred.

The rest of the academic year was reasonably routine, with the exception of organising drop-in sessions for midterm and exam study weeks. Towards the end of the academic year in May, the MSC launched a large paper based student survey of MSC services. The MSC issues a similar survey every year. During the month of June, this data was collated, coded and analysed using SPSS. This valuable information is used to review and evaluate the services that the MSC provides. This data was also used to assist in the writing of an annual report, which is circulated internally in Maynooth University to the relevant offices, departments and senior management.

Every year, during the last three weeks of July, Maynooth University Access Office provides a refresher course in mathematics for mature students that wish to study in Maynooth University. These students would typically have studied mathematics to the end of secondary school in Ireland (there are exceptions) but it may have been quite some time since they have studied any mathematics. While most of these students have ambitions to study a science based degree, many would also apply for courses such as business or accountancy. For the past 6 years the first author has taught this course and from 2016-17 it is part of the duties of the MSC University Tutor. As this is a refresher course, the pace is quite fast and intensive. To give the students a break, tours of the National Science Museum and the University's Russell Library (where some beautiful old mathematics and science books are housed) were arranged. Guest speakers lectured on topics such as astrophysics and the history of mathematics.

MLS provision is guided in Maynooth University, and elsewhere, by evidence based research. It is essential that we constantly review, and update our approaches, techniques and resources to maintain best practice in meeting the diverse and ever-changing needs of our students. During the year, the MSC University tutor, with the assistance of the second author, completed a paper (Mulligan and Mac an Bhaird, 2017) on the first author's experiences of motivating mature students on the Certificate in Science programme. Further research is under way investigating the influence of the MSC on student attitudes towards mathematics, and the impact of the second level drop-in services. In January 2017, the first author was elected as the "Tutor Representative" on the committee of the IMLSN. Getting involved with this committee provides a platform to make contributions to MLS nationally and internationally. It also facilitates the networking of MLS practitioners and researchers throughout the country.

4. Conclusion

The appointment of a full-time MSC tutor has been an undoubted success at Maynooth University, and this is especially evident as preparations are underway (September 2017) for the next academic year. In previous years, the MSC Director would have organised and carried out most of the tasks himself and trained a new MSC tutor to take over some MSC set-up activities. Now, for the first time, no such duties need to be assigned and the MSC Director and MSC University Tutor can spend their time more efficiently by reviewing and updating existing practices. Furthermore, the MSC University Tutor can use his experience to offer insights to the Director on the day-to-day running and the strategic direction of the MSC.

Having a full-time MLS tutor, who is a familiar face in the MSC, can also offer considerable benefit for students. For most tutors, the most enjoyable part of their role is the time spent assisting and interacting with students in the MSC. It may appear, to those not involved in MLS, that tutors simply help students when they get stuck. However, the role of MSC tutor is much more besides, and just because someone is good at mathematics or statistics, does not mean they will be a good MSC tutor. In addition to having the mathematical and statistical content knowledge, MSC tutors need the people skills required to engage and interact with a diverse group of students with a wide range of needs. Full-time MLS tutors will have a clear advantage over other MLS tutors in such situations, with the added benefit of having an opportunity to develop good working relationships with students. It is well documented that students often have a fear of mathematics, or are afraid to ask for help (Grehan et al., 2016). In the MSC, students often need a tutor who has both a sympathetic ear and the skills to help them cope with their mathematical difficulties. Student evaluations of MLS frequently highlight the relaxed and friendly atmosphere of the MSC, where students feel that they can ask any question, and experienced MSC tutors are key to facilitating this social environment (Waldock et al., 2016). Without doubt, effective tutoring in a MLS environment requires a unique set of skills and tutors with such skills, are an invaluable asset to any MSC.

The advantages to the MSC, the MSC student and the MSC tutor, of having a full-time contract for an MSC tutor, have already been discussed in this paper. While it may appear on the surface that MLS provision costs an institution money through salary, equipment etc., MLS, if properly supported, can have a clear financial benefit to the institution. There is evidence, in Maynooth University (Mac an Bhaird et al., 2009; Berry et al., 2015) and elsewhere (Symonds et al., 2007), that strongly suggests that MLS positively impacts on student retention and progression. For example, in student surveys conducted at the end of 2016-17, 26 Maynooth University students said that they had considered dropping out of university due to mathematical difficulties but cited the MSC as influencing their decision to remain. Eleven second level students stated that they intended to study at Maynooth University based on their experiences at the second level drop-in. The financial benefit to Maynooth University of these 37 students exceeds, by far, the financial cost of MLS provision for that academic year.

We argue that having more full-time MSC tutors is one approach for an institution to publicly endorse and promote their dedication to the student learning experience. It would recognise the positive contribution that experienced tutors can have and as outlined in this paper, due to an improved quality of service for students, this could have an added positive impact on student retention and progression.

Over the next two years, in addition to the work outlined in this paper, the MSC University tutor intends to investigate and partake in different forms of continuous professional development. For example, they will sit in on certain modules to extend their mathematical and statistical background. Being a valued member of university staff, with the opportunity to upskill, allows the MSC tutor to take ownership of their work and develop a professional identity. As a result of the success of the appointment of the MSC University Tutor, it is hoped that senior management of Maynooth University will consider appointing further full-time MSC tutors and that other institutions may follow their lead.

5. References

Berry, E., Mac an Bhaird, C. and O'Shea, A., 2015. Investigating relationships between the usage of Mathematics Learning Support and performance of at-risk students. *Teaching Mathematics and its Applications*, 34(4), pp. 194 - 204.

Croft, T. and Grove, M., 2016. Mathematics and Statistics Support Centres: Resources for training postgraduates and others who work in them. *MSOR Connections*, 14 (3), pp. 3-13.

Croft, A. and Grove, M. eds., 2011. Tutoring in a mathematics support centre: a guide for postgraduate students. Birmingham: National HE STEM Programme.

Cronin, A., Cole, J., Clancy, M., Breen, C., Ó Sé, D. (Eds) Cole, J., Cronin, A., O'Sullivan, C., Mac an Bhaird, C., 2016. *An audit of Mathematics Learning Support provision on the island of Ireland in 2015.* ISBN: 978-1-910963-07-4 Available via http://www.teachingandlearning.ie/wp-content/uploads/2017/03/MLS-Ireland-survey-2015-2.pdf (last accessed 16 September 2017)

Frawley, D., Pigott, V. and Carroll, D., 2017. *A study of progression in Higher Education: 2013/14 to 2014/15.* Higher Education Authority (HEA), Available via http://hea.ie/assets/uploads/2017/06/A-STUDY-OF-PROGRESSION-IN-IRISH-HIGHER-EDUCATION.pdf (last accessed 21 September 2017)

Grehan, G., Mac an Bhaird, C. and O'Shea, A., 2016. Investigating students' levels of engagement with mathematics: critical events, motivations, and influences on behaviour. *International Journal of Mathematical Education in Science and Technology,* 47(1), pp. 1-28.

Lawson D., Croft, T. and Waller, D., 2012. Mathematics support past, present and future. *Innovation Practice and Research in Engineering Education*. Available via http://cede.lboro.ac.uk/ee2012/papers/ee2012_submission_179_gp.pdf (last accessed 11 September 2017)

Mac an Bhaird, C., Morgan, T. and O'Shea, A., 2009. The Impact of the Mathematics Support Centre on the Grades of First Year Students at the National University of Ireland Maynooth. *Teaching Mathematics and its Applications*, 28 (3), pp. 117 – 122.

MSC, 2018. *Mathematics Support Centre.* [online] Available at: http://supportcentre.maths.nuim.ie/topiclinks [Accessed 7 February 2018].

Mulligan, P. and Mac an Bhaird, C., 2017. Motivating Mature Students of Mathematics. *MSOR Connections,* 15(3), pp. 36-43

O'Brien, C., 2017. More than 6,000 students drop out of college in first year. *The Irish Times.* 27th April 2017, Available via https://www.irishtimes.com/news/education/more-than-6-000-students-drop-out-of-college-in-first-year-1.3062362 (last accessed 1 October 2017)

O. Fitzmaurice, Cronin, A., Ní Fhloinn, E., O'Sullivan, C. and Walsh, R., 2016. Preparing Tutors for Mathematics Learning Support. *MSOR Connections*, 14 (3), pp. 14-21.

O'Sullivan, C., Mac an Bhaird, C., Fitzmaurice, O. and Ní Fhloinn, E., 2014. *An Irish Mathematics Learning Support Network (IMLSN) Report on Student Evaluation of Mathematics Learning Support: Insights from a Large Scale Multi-institutional Survey.* Available via http://eprints.maynoothuniversity.ie/6890/1/CMAB_IMLSNFinalReport.pdf (last accessed 8 October 2017)

Pfeiffer, K., Cronin, A. and Mac an Bhaird, C., 2016. The key role of tutors in mathematics learning support – a report of the 10th annual IMLSN workshop. *MSOR Connections*, 15 (1), pp. 39-46.

Symonds, R. Lawson, D. and Robinson, C., 2007. The effectiveness of support for students with non-traditional mathematics backgrounds, *Teaching Mathematics and its Applications*, 26 (3), pp. 134–144.

Waldock, J., Rowlett, P., Cornock, C., Robinson, M. and Bartholomew, H., 2016. The role of informal learning spaces in enhancing student engagement with mathematical sciences, *International Journal of Mathematical Education in Science and Technology*, 48(4), pp. 587-602.

CASE STUDY

Teaching statistical appreciation in quantitative methods

Peter Mitchell, Department of Animal and Plant Sciences and Department for Lifelong Learning, University of Sheffield, Sheffield, U.K. P.L.Mitchell@Sheffield.ac.UK

Abstract

Statistical appreciation is defined here as the knowledge about statistical tests, how they are chosen, the procedure and interpretation of the results, but without the calculations of the test statistic. This was taught in modules on research skills to students taking part-time degrees at the University of Sheffield. Details of the content, teaching methods and assessment are given here, with stress on the correct understanding of P-values and interpretation of statistical significance. Given that more people need to understand the results and interpretation of statistical tests than to do the calculations, statistical appreciation is of general value, especially to research supervisors. It also provides a firm base for further learning and training in statistics.

Keywords: statistical appreciation, quantitative methods, statistical tests, P-value, statistical significance.

1. Introduction

For over ten years I have taught quantitative methods to students taking part-time degrees in a range of subjects, in a module initially called Developing Research Project Skills and later called Research Methods. Students in the Department for Lifelong Learning (DLL) at the University of Sheffield took this module at Level 2 (usually the third and fourth years of a part-time degree) in preparation for research carried out at Level 3. As is common in service teaching of statistics, the DLL had in mind initially that students would acquire all the knowledge and skills required to carry out whatever statistical analysis might be required in their research. In practice, it was soon realized that this ambition would have to be severely trimmed, for two reasons. First, only 16 hours of contact time were available for quantitative methods, the rest of the module being qualitative. Secondly, the starting points in mathematical ability of these students were varied but mostly low so basic numeracy had to be refreshed, and only a moderate rate of progress could be expected.

After some introductory material on numeracy, use of tables and graphs, and an outline of the research process, I concentrated on descriptive statistics and simple inferences using standard errors and confidence intervals. A firm grasp of these concepts would be useful in all fields of study. This left about four hours of contact time for statistical tests which some students might use in their research, and many would encounter in their reading in their subject. Clearly, there was not time to offer training in the use of a statistical package on the computer, nor to provide full information on t-tests, chi-squared tests, regression, and so on, nor to practise carrying out each test on data and interpreting the results. I decided that the best that could be done was to teach statistical appreciation. These particular part-time degrees ceased enrolling several years ago, with the teach-out period almost finished, so the purpose of this case study is to record what was done and to share more widely my experience of teaching the ideas behind statistical appreciation.

2. Definition

By statistical appreciation, I mean the knowledge *about* statistical tests, including the ability to recognize them and follow the procedure and interpretation, without being able to carry out the actual calculations. Appreciation here is used as in art or music appreciation: many people enjoy art and learning about art without being able to draw or sculpt, and cannot sing or play an instrument but listen to music attentively and acquire much knowledge about it. For a long time, I used the phrase statistical awareness but this has been used by others with much wider meaning, e.g. Davies et al. (2012). Their view includes competence with calculations for statistical tests. On reflection, appreciation rather than awareness is more suitable, not least for the parallel with art or music appreciation.

I stumbled on statistical appreciation as a matter of expediency but now see that it has value in its own right. Calculation of test statistics is a stumbling block for many students. It is onerous by calculator, except for the smallest sets of data, and nowadays requires training in the use of a computer package. My impression is that students are often repelled by this experience unless there is sufficient time and progress can be gradual and linked with statistical understanding. Paul Wilson (personal communication, Sigma Network meeting, July 2016) pointed out that more people need to know about statistical tests, to understand the results and interpretation, than are actually required to make the calculations from the data. Of course, being able to do the calculations, even if only from the menus of a statistical package, is useful but not at the expense of a sound understanding of the procedure of statistical testing and how to interpret the results.

3. Prerequisites

Statistical testing makes use of probability, and there is much talk of P=0.05 or a one in twenty chance or 5% probability, etc. Consequently, an important prerequisite for statistical appreciation is the ability to understand probability on a scale from 0 to 1 and alternative means of expression, with the ability to convert from one to the other. Descriptive statistics were introduced with small sets of data obtained by students themselves and included the once-in-a-lifetime calculation of standard deviation (SD) by hand and calculator using the standard formula. This was checked using entry of data in statistical mode on a calculator, which brings in discussion of n or n–1 as a divisor, and hence samples and populations. Further calculations for standard errors (SE) and confidence intervals followed, including a brief presentation of the Normal distribution, and the use of a value of Student's t obtained from statistical tables.

4. Content of statistical appreciation

Allowing for complete beginners, some barely mastering descriptive statistics and confidence interval, I wrote a handout of 7500 words entitled Overview of Statistical Tests (Table 1). This included how to choose a statistical test, the procedure that all tests have in common, interpretation of probability associated with the test statistic (P-value) as a means of determining significance, use of statistical tables, statistical and practical significance, and effect size versus hypothesis testing. Outlines of five tests were given, plus descriptive statistics and simple inference (not a test as such but using the same mathematical foundation). The outline of correlation is given in Table 2 as an example.

Students were warned that the handout was a rather abstract explanation of statistical testing and that they should not expect to understand it immediately. It would need to be read several times, with cross-references followed when necessary. Once they had gained some understanding, during this module, they would be able to return to the handout when needing to refresh their memory because they had encountered a statistical test in their reading or because a particular test was required in their own research. This material would provide the basic information about the test except for how to calculate the test statistic. If this was wanted in their own research, then further help from textbooks, a statistician, colleagues or the help facility on a statistical package should be taken up. This handout is available from the MSOR Connections website (Mitchell, 2018).

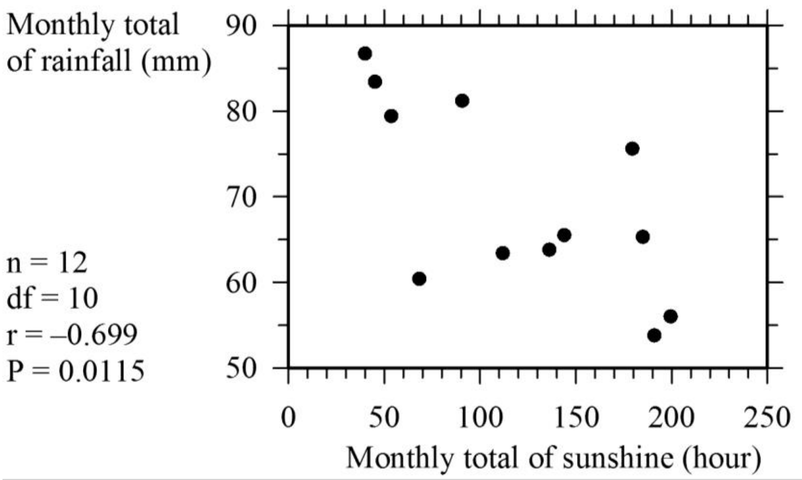
Table 1. Coverage of statistical appreciation in the handout Overview of Statistical Tests.

Heading	**Summary**
What statistical tests do	Distinguish interesting and uninteresting variation in data, i.e. signal from noise.
Scientific philosophy	The null hypothesis (in words only).
How statistical tests work	Choice depends on type of question and data. Standard key terms: test statistic, degrees of freedom, probability associated with the calculated test statistic, statistical significance. Significance as researcher's interpretation of the probability (P-value); standard thresholds (P=0.05, 0.01, 0.001) and words (significant, very significant, highly significant).
Degrees of freedom	Method of indicating how many independent pieces of information there are in the data.
Correct understanding of P-values	Definition of P-value; recognition that thresholds 0.05, 0.01, 0.001 are arbitrary; a way of converting continuous probability to categories of significance. Use of statistical tables and resources on the internet.
Worked example of using a statistical table	Working across columns to find range of probability for the value of the test statistic.
Proof and probability in statistical testing	You *cannot* prove anything with statistics. You *can* compute the probability of obtaining a set of results like this (or more extreme) if the null hypothesis was true.
Procedure	Consider question being asked, data to be obtained, which statistical test needed, how assumptions will be satisfied; gather data; obtain test statistic; interpret the probability for statistical significance; conclude about the difference or association of the initial question.
Choosing a test	Tests classified by purpose of test. 1. Describe the data and make simple inferences: descriptive statistics and confidence interval. 2. Look for association: classification of attributes (categories) needs chi-squared test; measurable attributes on a scatter graph needs correlation. 3. Compare means: two means needs t-test; two or more means needs analysis of variance. 4. Predict one quantity from another: regression.
A worked example	Mean weight of voles in two populations, using t-test. Key terms highlighted when they occurred in the text. Data contrived to provide a result that was not significant (P=0.126). Discussion of what not significant means, how larger samples would have produced a significant difference (variability held constant).
The conclusion of a statistical test	Clear statement of results always required; examples shown.
What "not significant" means	Say "a difference (or association) could not be detected"; dependence on variability and number of degrees of freedom.
Statistical and practical significance	Statistical significance as "is it likely to be true?"; practical significance as "is it worth taking account of, basing decisions on?".
Effect size versus hypothesis testing	Complementary but effect size provides direction of difference and its confidence interval; links to practical significance.
Outlines of statistical tests	One page each on descriptive statistics and simple inferences, chi-squared test, correlation, t-test, analysis of variance, and regression. Standard headings: purpose, data, null hypothesis, test statistic, degrees of freedom, assumptions, example (with statement of results), variations and elaborations, equivalent non-parametric test.

Table 2. The outline for correlation, from the handout Overview of Statistical Tests.

Pearson's Correlation

1. **Purpose.** To look for association (linear correlation) between two measured attributes.
2. **Data.** Two measurements from each case or item or subject, that can be plotted on a scatter graph (does not matter for the statistics which measurement is x). *Always* plot the graph, even if only a sketch.
3. **Null hypothesis.** No (straight-line) association between the two measurements; the points on the graph are a random cloud or occur in a horizontal or vertical line.
4. **Test statistic.** The correlation coefficient, r, measures the strength of association, between 0 for no association to –1 for perfect negative association, to +1 for perfect positive association. Test of significance is usually based on Student's t-test but tables of critical values of r are available for direct assessment.
5. **Degrees of freedom.** For sample of n cases, df = n–2. Note that there are n *cases* (or items or subjects), each case with two numbers.
6. **Assumptions.** Data from a random and independent sample of the population, where there is an underlying straight-line relationship between the measurements. The measurements are distributed bivariate Normally.
7. **Example.** Scatter graph of sunshine and rainfall recorded at Sheffield (monthly means for the period 1981–2010), and the correlation between them.



   Statement of results. There is a moderately strong correlation (r = –0.699) between monthly totals of rainfall and sunshine; the correlation is significant (10 degrees of freedom, P = 0.0115). It is a negative correlation, i.e. high sunshine tends to be associated with low rainfall.
8. **Variations and elaborations.** None: correlation is a simple and fairly crude technique. If there is a statistically significant correlation then it may prompt further investigation and gathering of data, for example to predict one measurement from the other using regression.
9. **Equivalent non-parametric test.** There are two methods of correlation when the data are ordinal: Spearman's and Kendall's rank correlations.

5. Teaching and assessment

The learning outcomes are given in Table 3. After a brief explanation of statistical appreciation, i.e. that they would learn *about* statistical tests without any calculations, I found that it was best to plunge in with an accessible and intriguing example. The question was whether men or women spoke more words during the day, and Mehl *et al*. (2007) had collected suitable data and performed a t-test. I presented the summary data (means and SDs) and the result of the t-test. In this case the result was not significant (P=0.50) despite a small difference in the mean numbers of words spoken, thus leading to an explanation about statistical significance and its interpretation. Having seen one statistical test in action, I then mentioned briefly the other tests and when they were used, and discussed whether statistics could prove anything. A second example test followed, chi-squared, applied to data about whether women who were handed a baby or a parcel held it to the left or right side of the body (Campbell 1989, p. 130). This too engaged students, who wanted to know more than I anticipated initially about calculation of expected values in the computations for the chi-squared test. After these two examples I presented the procedure for statistical testing, pointing out where the key terms (null hypothesis, test statistic, degrees of freedom, probability associated with the test statistic, statistical significance) had occurred in the previous examples. Two tasks in pairs or trios were undertaken: interpreting probabilities as statistical significance, and recognizing the key terms in statements of results from statistical tests.

Table 3. Learning outcomes for the two sessions (4 hours contact time) on statistical appreciation.

Students will
(a) be able to identify which test to use from the research question and type of data;
(b) recognize that there is a procedure in common for statistical tests;
(c) realize that statistics cannot prove anything but can quantify the uncertainty;
(d) be able to find statistical significance from the probability of a statistical test;
(e) be able to identify key terms used in statistical tests;
(f) be able to state what has been found from a statistical test;
(g) realize that assumptions must be satisfied for calculated probabilities to be reliable; and
(h) be able to calculate predicted values from a regression equation.

The second session started with a review of the procedure for statistical tests. Then examples of correlation, analysis of variance and regression were presented. Regression required much more time than the other tests. Few of these students were familiar with the equation for a straight line and interpreting the coefficients as a slope and an intercept. This had to be introduced for beginners, leading to the ability to calculate a predicted value (y) from any input value (x) supplied (used for work in class: each student computed a predicted value from the regression equation given a value for x, to assemble a set of results for the class). The final topic was the importance of assumptions in statistical testing, pointing out that these were stated in the outline of each test given in the handout (see Table 2 for an example). Omitted in class for want of time were statistical and practical significance, and effect size versus hypothesis testing; students were referred to the handout as a starting point, in case these topics occurred in their reading or research project.

Assessment was by coursework to avoid needless anxiety and unrealistic time constraints as in a test or examination. The single assignment of six questions to be completed individually over a period of three or four weeks comprised a mix of questions requiring numerical and narrative answers. Narrative answers were often merely a word or two, sometimes a sentence of explanation to test understanding of a concept. Some questions were scenarios or sets of data where the responses sought were which test would be suitable, what initial analysis of the data could be undertaken, but no computation of test statistic was required. Other questions presented the results of a test and asked for interpretation or simple calculations from the results (e.g. predicted values from a regression), or requested examination of the assumptions for the test. Assessment by coursework allowed students to work at whatever pace suited them and to consult any resources such as handouts and class notes, textbooks, and the internet. Despite this freedom, a wide spread of marks was always encountered.

6. Discussion

My aim was to produce students who could work out which test they needed for a research question and data, could pick out the key terms of the procedure, could state what a P-value meant and interpret it for statistical significance, and could answer the research question from the results of the test. If they acquired this knowledge then they could appreciate statistical tests in the literature, and do everything for their own use of a test except calculate the test statistic (which is arguably a trivial ability compared with the knowledge above). Statistical appreciation should provide a firm starting point for further learning on how to carry out the test and for training on a statistical package.

There are two implications for support services for students (not taking mathematics or statistics degrees) who require statistics as part of a project, and also for the service teaching of statistics.

1. If students have acquired statistical appreciation (i.e. correctly identified the test required, obtained suitable data, and can interpret the P-value and answer the research question), should it become routine for the support service to provide the test statistic?

2. Should statistical appreciation, as a minimum, be expected of any research supervisor? It requires nothing more than basic numeracy, and no training on statistical packages. Supervisors could then discuss with students use of statistical tests in the literature and their own project, and obtain further help, or plan further learning, for calculation of the test statistic when required.

Is there a role for the **sigma** Network in promoting statistical appreciation? There are no doubt weaknesses and imbalances in my coverage so I welcome further discussion, and am happy to share teaching materials.

7. Acknowledgements

Dr Verity Brack first recruited me to teach quantitative methods and was an inspiring colleague. I am grateful to the students who formed successive interesting classes. Dr Paul Wilson's talk "Statistics first, software later" at the **sigma** Network meeting, 1st July 2016, encouraged me to think further about statistical appreciation. Comments from a reviewer improved the manuscript.

8. References

Campbell, R.C., 1989. *Statistics for Biologists (third edition)*. Cambridge: Cambridge University Press.

Davies, N., Marriott, J. and Martignetti, D., 2012. A statistical awareness curriculum for STEM employees. *MSOR Connections*, 12, pp. 12–16.

Mehl, M.R., Vazire, S., Ramirez-Esparza, N., Slatcher, R.B. and Pennebaker, J.W., 2007. Are women really more talkative than men? *Science*, 317, p. 82.

Mitchell, P., 2018. Overview of the Statistical Tests, *MSOR Connections* 16(1). Available from: https://journals.gre.ac.uk/index.php/msor/index

CASE STUDY

A Comparison of Nursing Maths Support Approaches

John Little, Department for the Enhancement of Learning, Teaching and Assessment, Robert Gordon University, Aberdeen, United Kingdom. Email: j.little@rgu.ac.uk

Abstract

This article compares interventions for improvement of nursing and midwifery students' numeracy and drug calculations skills respectively; numeracy test / re-test with maths support versus a numeracy workshop and a calculations drop-in. In the first intervention supported students (n=11) increased their mean score from 19 out of 40 to 30 out of 40 and their peers (n=165) increased from 27 out of 40 to 30 out of 40, both statistically significant increases. In the second intervention the highest scoring group of students made use of both the workshop and the drop-in. Whilst there was no statistically significant difference in mean scores across the four identified groups of students (those with no maths support (n=205), those with foundation numeracy support only (n=18), those with calculations drop-in support only (n=11) and those (n=6) with both foundation numeracy and calculations drop-in support) there were no low scoring outliers amongst the students who made use of any of the support offered. Although there is stronger evidence for the first intervention having impact, the second intervention is more aligned to drug calculations skills development. However, the latter would need a larger scale sample to confirm efficacy.

Keywords: dosage, calculations, numeracy, nursing.

1. Introduction

The Nursing and Midwifery Council (NMC), the responsible professional body in the UK, seeks to focus on numeracy skills development, implied through its code (NMC, 2015a) and, more explicitly, through the standards that underpin the code (NMC, 2015b), albeit there is no national standard drug calculations test.

At the author's institution one to one, small group and large group maths support outside of main teaching programmes has been available to nursing students since 2003 and over the last ten years a commercial e-assessment package has been adopted as the means to deliver drug calculations proficiency. Without a doubt this has brought great benefits, not least by indirect promotion of a self-test strategy via online practice and formative assessments that aligns well to the known successful study and learning strategy of frequent testing and spread effort (Hartwig and Dunlosky, 2012). This author has also previously noted the benefits of multiple online formative test attempts (Little, 2006).

Note, in this article p values are mostly reported to the three decimal places presented in SPSS outputs with an assumed acceptance level of 0.05, the exception being p values obtained for transformed data.

2. Session 2012-2013 Intervention 1: Re-test of Foundation Numeracy Skills with Interim Maths Support

In session 2012-2013 students who didn't achieve full marks on their formative foundation numeracy (FN) assessment were asked to make a second attempt at the assessment, with the option of one to one or small group maths support (MS) appointments as an intervention in the interim. Note, the following figures only include students who completed both first and second attempts at the intended times.

Of 205 students, 29 are excluded due to the absence of one or more tests scores. Second re-test data for 6 of those 29 excluded students is available (all of these having appeared to miss their original second attempt) but only one of those six scored full marks and only one took advantage of maths support. Of the remaining 176 cases a group of 11 made use of maths support.

This test / re-test approach arguably aimed to exploit the learning effect of test repetition which Hartwig and Dunlosky (2012) allude to. Shapiro Wilk normality tests indicate that only the maths support students' data is suitable for assessing the change in test score using a paired samples t test (n=11, p=0.401), the other group (n=165, p<0.001) ostensibly necessitating use of the non-parametric alternative, the Wilcoxon Signed Rank comparison of medians test or, preferably, data transformation. In the latter case, for which Shapiro-Wilk is less sensitive due to the large group size (n>30), the relevant histogram indicates an obvious positive skew for the non-maths support data, and indicators of data that probably doesn't belong to a normal distribution (see Appendix for histogram). That said, it is known that the paired samples t test can be robust to departures from normality (Zumbo and Jennings, 2002) for larger samples (n>30) and effect size > 0.36 or, more commonly, effect size > 0.5 and so results from both tests and for the transformed non-maths support paired data are reported here. Assuming that mixed between-within ANOVA is similarly robust, results for the latter are also reported.

According to the SPSS help file (IBM Support, 2018) the most appropriate transformation for positively skewed data that includes values close to but not predominantly zero is a logarithmic transformation. However, this presents a difficulty in that it cannot be directly used with negative values i.e. a number of non-maths support students actually saw their scores fall after re-test. One option then is to add a constant to the paired difference data and then apply a logarithmic transformation. This in turn presents the problem of different results for different constants added. Notwithstanding the latter issue the t statistic can then be obtained from the mean of the transformed differences divided by the standard error of the transformed differences (sample standard deviation divided by the square root of the sample size n=165) with n-1 degrees of freedom (Shier, 2004) and a t table consulted to obtain the relevant critical value.

A paired samples t test of supported students' scores indicates a statistically significant increase in mean score from 19.09 to 30.36 (t = -5.214, n = 11, p < 0.001) and the corresponding Wilcoxon Signed Rank Test result is also significant (p = 0.003). A paired samples t test of the peer group also indicates a statistically significant increase in mean score from 26.65 to 30.35 (t = -10.140, n = 165, p < 0.001) and the corresponding Wilcoxon Signed Rank Test result is also significant (p < 0.001). The effect size, being calculated from $t^2/(t^2+(n-1))$ (Pallant, 2010), yields a large effect size of 0.731 and a medium effect size of 0.385 in each respective case, the latter exceeding the 0.36 robustness threshold suggested by Zumbo and Jennings (2002).

For completeness, however, a paired samples t test of non-maths support students' log transformed paired difference data was also performed (adding on a constant of 10 to eliminate negative and zero differences) and this yields t = 93.12 with t_{crit} for 164 degrees of freedom = 1.97, this indicating a statistically significant increase in mean score. Changing that constant to the smallest workable whole number yields t = 68.44 which, whilst a noticeably smaller value, still far exceeds the critical value. In either case p is < 0.00001 so the choice of constant doesn't affect the conclusion or, rather, since there is agreement between the paired samples t test, the Wilcoxon test and the test of the transformed data it is reasonable to conclude that there was a real increase in mean score for the non-maths support students.

Note, the change in sign from the negative t values noted earlier simply arises from the order that the paired differences are taken in. One would assume that entering the relevant assessment attempt score variables, mark 1 and mark 2, in that same order in the paired t test SPSS dialogue box that mark1 might be subtracted from mark2 but clearly SPSS interprets the left-hand cell as being the 'more recent' data. The transformed data t values were obtained using MS Excel with the paired differences obtained as mark2-mark1.

A mixed between-within anova yields a significant interaction between time and MS (Wilks Lambda $p < 0.001$), implying that caution must be exercised in interpretation of main effects. Time (re-test opportunity) is a significant main effect (Wilks Lambda $p < 0.001$) with a medium effect size (partial eta squared = 0.36) but the between subjects' effect for MS is not statistically significant ($p = 0.074$), indicating that the maths support intervention was not so important in increasing test scores as the re-test opportunity.

Based on a paired difference standard deviation of 4.68 derived from the present data set, Minitab power and sample size calculator indicates that a change in mean score of greater than 5.0 can be detected with a statistical power of 0.85 with 10 subjects. Much as Wright (2008) observes, it might be reasonably argued that an increase in mean score of 1.0 is clinically important as this may influence a future clinical decision on a calculation but then a much larger sample (n=199 for both test and re-test group) would be needed.

The results that have been obtained ostensibly represent the ideal outcome from a maths support perspective; the weaker group improved to a level comparable to their peers. The scatter graph in Figure 1 provides a visual illustration and further insight into the performance of the maths support group. Most students lie in the upper triangular part of the graph, which means the second attempts are better than the first attempts for most students. However, the students with maths supports got more improvement in their scores i.e. they are generally more to the top and left of the scatter graph than their peers. In fact, all 11 of the maths support students increased their scores whereas 26 out of 165 non-maths support students either scored more poorly or the same as in their first attempt.

However, there are arguably two groups of supported students – those who responded very well to the support (scoring >= 30) and those who responded less well (scoring < 30). This may be because the weaker students were harder to reach in the relatively short two-week period between assessment attempts and hence the latent need for more one to one or small group maths support.

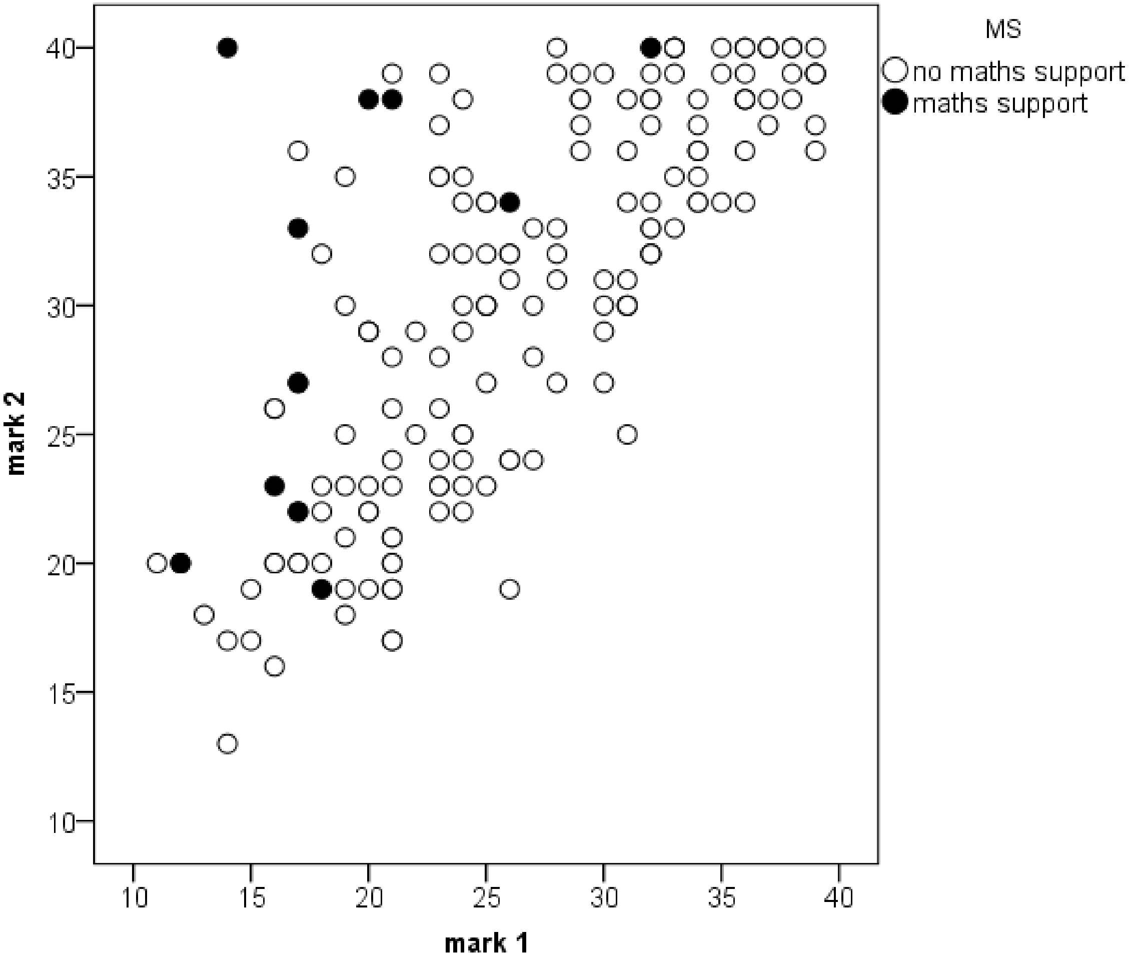


Figure 1. Scattergraph of 2nd Attempt (mark 2) vs 1st Attempt (mark 1) Clustered by MS

3. Session 2015-2016 Intervention 2: Foundation Numeracy Workshop and Drug Calculations Drop-in

In session 2015-2016 a foundation numeracy (FN) workshop intervention was employed to help students improve their numeracy skills. Additionally, a student-led large group drop-in session of solids, liquids and injections (SLI) calculations was arranged prior to the corresponding formative assessment.

Whilst attendance at workshops and uptake of the maths support appointment service was limited, the student led drug calculations drop-in was well attended and there is some weak evidence that both of these interventions helped students but especially so where both help opportunities were used i.e. there may have been a cumulative benefit.

Table 1. Range and Mean SLI Drug Calculation Scores for Four Different MS Groups: 1 (those who didn't use maths support), 2 (those who attended the FN workshop), 3 (those who attended the SLI drop-in) and 4 (those who attended both the FN workshop and the SLI drop-in)

Maths Support Group	N	Min score	Max score	Mean score	Std. Deviation
1 No maths support	205	17	90	81.76	12.349
2 FN maths support only	18	37	90	80.06	12.379
3 SLI maths support only	11	65	90	80.82	7.859
4 SLI and FN maths support	6	83	90	87.17	2.714

While the widely differing sample sizes mean that these groups' results are not directly comparable, Table 1 indicates that the mean SLI mark was highest for the small group of students who attended both the FN workshop and the SLI drop in. The only check against this group being biased towards self-motivated high achievers is anecdotal. That is, through discussion of question solutions with the students it became clear that all of the drop-in students exhibited some amount of difficulty with drug calculations during the session; the hope being, of course, that those difficulties were resolved by the session.

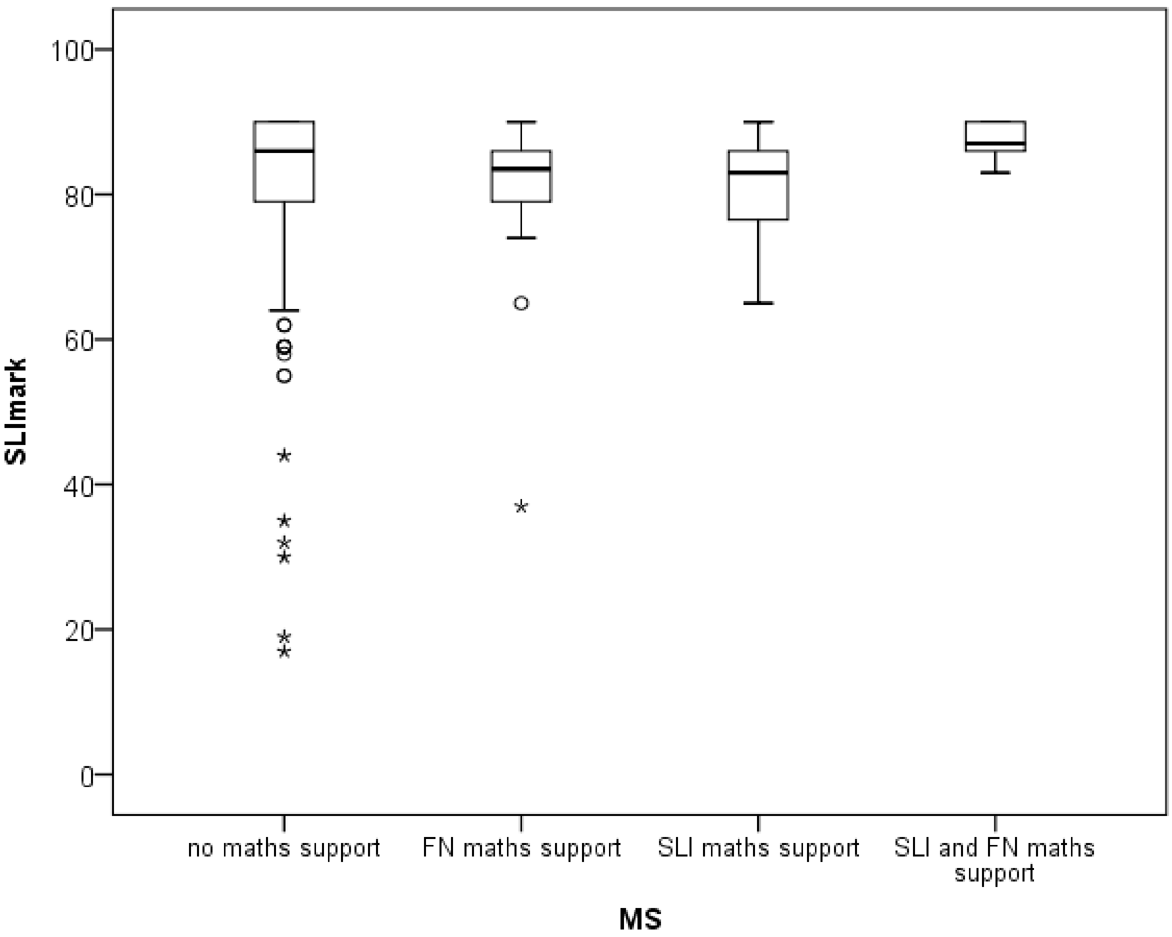


Figure 2. Boxplots of SLI Drug Calculation Marks for Different MS Groups: those who didn't use maths support (no maths support), those who attended the FN workshop, those who attended the SLI drop in and those who attended both the FN workshop and the SLI drop-in

In Figure 2 the horizontal lines within each box represent medians, the ends of each box are the upper and lower quartiles and the whiskers are drawn at the maximum and minimum value (excluding outliers defined as lying beyond 1.5 box lengths above or below the box with ordinary outliers denoted by circles and extreme outliers, more than 3 box lengths away from the box, denoted by asterisks) for each group (Pallant, 2010). There is a notable absence of low scoring outliers amongst students who attended the SLI drop-in. This supports Wright's (2008) finding that specialist skills intervention raised the minimum score on a numeracy test compared to a control or standard group.

Referring again to Table 1 it is difficult to compare group scores because the widely varied group sizes and potential associated heteroscedasticity (unequal variance across groups) presents a problem for comparing means using one-way ANOVA. Surprisingly, however, there is no violation of the homogeneity of variance assumption as Levene's test yields p = 0.342. However, normality tests indicate that the two largest groups do not belong to a normal distribution.

Assuming ANOVA is robust to departures from normality, the one-way ANOVA yields p = 0.653. For exponential transformed data the result is p = 0.686 so there is some agreement that there is no difference in group means.

For two of the comparison groups though the kurtosis to kurtosis standard error ratio (Gilchrist and Samuels, 2014) is greater than 2 so the robust test of equality of means must be consulted instead, of which the Welch test is deemed the more appropriate in this case.

The Welch test yields p = 0.006 and this contradicts the Kruskall-Wallis test result (p=0.187) and the earlier noted ANOVA results. An innocuous looking note in the SPSS output observes that group sizes are unequal and type 1 error levels are not guaranteed and, indeed, there is also disagreement between the robust tests of equality of means with the Brown-Forsythe test yielding p=0.391. It would seem reasonable to say then that there is some doubt over what conclusion can be drawn from the data and test results or it can be said that the Welch result either indicates a difference in means or represents a type 1 error.

Minitab power and sample size calculator indicates that larger numbers (n=143) are needed in each comparison group in order, for example, to be able to detect a difference in means of 5.0, assuming a test score standard deviation of 12 based on the present data set, and the figures for the non-parametric alternative are inevitably going to be higher. This would also cast doubt on any firm conclusion being drawn with smaller and unequal groups.

In one sense, this data is suggestive of ineffective interventions but if the aim is to ensure parity of performance across groups then there is perhaps scope to say that the second intervention was useful to students.

4. Conclusions

Students re-sitting a basic numeracy test improved their performance but more so, and to the level of their peers, if they also engaged in one to one or small group maths support.

Foundation numeracy workshops and drug calculations drop-in sessions seem to have a cumulative beneficial effect.

A drug calculations drop-in prior to formative assessment largely eliminated low scores for drop-in students in comparison to their peers.

E-assessment has broadly been a leap forward for students and promises more in the future (Sabin et al, 2013) but isn't a panacea; there is still scope for a maths support tutor to add value.

Whilst there is stronger evidence for the first intervention having impact the second intervention is possibly more worthy of pursuit, since it is more aligned to calculations skills development. However, the latter would need a larger scale sample to confirm efficacy.

Logarithmic data transformation, indeed any of the recommended transformations, is problematic for paired differences that include zero but would still seem to yield a result that can lend support to conclusions based on non-parametric test results.

5. Acknowledgements

The author acknowledges the excellent, friendly and pragmatic assistance of Kate Goodhand, Marina Ritchie and colleagues in the School of Nursing and Midwifery.

6. Appendix

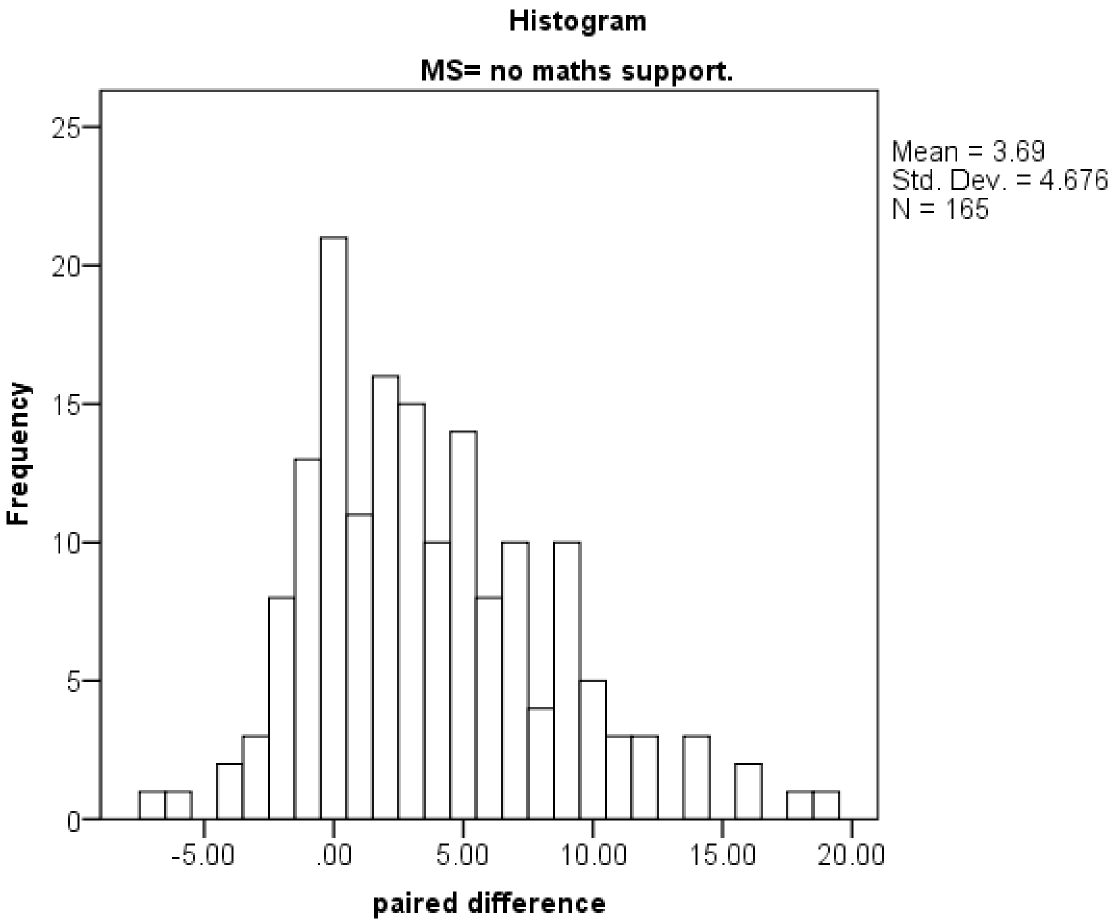


Figure 3. Histogram of Assessment Marks Paired Differences

7. References

Dunlosky, J. and Hartwig, M., 2012. Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychon Bull Rev* 19:126–134. Available at: http://dx.doi.org/10.3758/s13423-011-0181-y [Accessed 5 January 2017]

Gilchrist, M. and Samuels, P., 2014. *One Way ANOVA – Additional Material*. [pdf] Loughborough and Coventry: statstutor. Available at: http://www.statstutor.ac.uk/resources/uploaded/anovaadditional2.pdf [Accessed 8 January 2018]

IBM Support, 2018. *Transforming Variable to Normality for Parametric Statistics*. IBM: New York. Available at: http://www-01.ibm.com/support/docview.wss?uid=swg21479677 [Accessed 8 January 2018]

Little, J., 2006. Online support for medicine dosage calculations. *British Journal of Nursing* 5(21), pp. 1192-95. Available at: http://dx.doi.org/10.12968/bjon.2006.15.21.22380 [Accessed 27 May 2016]

Nursing and Midwifery Council, 2015a. *The NMC Code of Professional Conduct: Standards for Conduct, Performance and Ethics*. NMC: London. Available at: https://www.nmc.org.uk/standards/code/ [Accessed 27 May 2016]

Nursing and Midwifery Council, 2015b. *Standards for medicines management*. [pdf] NMC: London. Available at: https://www.nmc.org.uk/globalassets/sitedocuments/standards/nmc-standards-for-medicines-management.pdf [Accessed 1 July 2016]

Pallant, J., 2010. *SPSS* Survival *Manual 3rd Ed*. McGraw Hill / Open University Press: Maidenhead.

Sabin, M. et al., 2013. Safety in numbers 5: Evaluation of computer-based authentic assessment and high fidelity simulated OSCE environments as a framework for articulating a point of registration medication dosage calculation benchmark. *Nurse Education Today* 13(2): 55-65. Available at: http://dx.doi.org/10.1016/j.nepr.2012.10.009 [Accessed 5 January 2017]

Shier, R., 2004. *Statistics: Paired t tests*. [pdf] Loughborough and Coventry: statstutor. Available at: http://www.statstutor.ac.uk/resources/uploaded/paired-t-test.pdf [Accessed 8 January 2018]

Wright, K., 2008. Can effective teaching and learning strategies help student nurses to retain drug calculation skills? *Nurse Education Today*. 28(7): 856–864. Available at: http://doi:10.1016/j.nedt.2008.01.002 [Accessed 2 June 2016]

Zumbo, B. D. and Jennings, M. J., 2002. The Robustness of validity and efficiency of the related samples t-Test in the presence of outliers, *Psicológica*, 23(2), pp. 415-450.

Statistics Teaching Note H

Peter Mitchell, University of Sheffield 21st February 2018

Overview of Statistical Tests

1. **What statistical tests do.** Given the variability in any set of data, statistical tests allow us to distinguish interesting variation from uninteresting, background, random variation; if you like, to tell the signal apart from the noise. Interesting variation arises from differences between groups of items in the set of data, or from associations in the data. Differences between groups spring from questions such as: do left-handed cricketers score more runs than right-handed ones, or how do the stone axe-heads of Wales, the Lake District, East Anglia and Kent differ in size? Associations occur when we ask questions about whether colour of hair is related to colour of eyes (including do redheads really tend to have green eyes?), or if the number of mink living on a river is correlated with the number of water voles, and so on. The background variation arises from miscellaneous causes, not relevant to the question of interest.

   In the set of data, interesting and background variation are mixed up in the actual values of the observations. To answer the question about differences or associations we need a method that can separate the two sources of variation and give us a way of assessing whether the difference or association is likely to be genuine or just a chance effect in the data. In principle, statistics is simple: it is about measuring variation, attributing causes, and calculating the probabilities of obtaining particular sets of results by chance.
2. **A dose of scientific philosophy.** In terms of formal logic and scientific method, we set up a **null hypothesis** that there is a specified difference between the groups, or a specified association. The specified difference is often zero difference or zero association but it can be a non-zero difference or some particular association if required. We then attempt to falsify (i.e. to nullify) the hypothesis, using a statistical test to calculate how likely it is that we would get the results we have in the set of data *if* the null hypothesis were true. If it is not very likely then we proceed as if the null hypothesis has been falsified, i.e. we work on the basis that there *is* a difference between the groups, or there *is* an association. We cannot be perfectly sure, but we can be reasonably sure that the difference or association is "true" unless a rather unlikely event has occurred. The unlikely event is the occurrence of a set of data showing the specified difference or association when really there isn't one. We return to this point in §4.
3. **How statistical tests work.** We choose an appropriate statistical test, depending on the question asked and the type of data (see §5 and §6). We calculate the **test statistic** for this test from the data. Now we need to find the **degrees of freedom** (abbreviation df or d.f.); this depends on the number of observations or on the number of groups in the data (see Appendix A, §A1). Then we find out the **probability associated with the calculated test statistic** and its degrees of freedom. A computer package will provide the probability to three or four decimal places but for other calculations we look up the test statistic in tables of critical values to find a range of probability associated with it (see Appendix B). The probability is interpreted to provide the **statistical significance**, i.e. do we regard the difference or association of our initial question as statistically significant or not? The conventional levels of significance are shown in Figure 1.

Threshold probabilities

		P = 0.05 (1 in 20)		**P = 0.01 (1 in 100)**		**P = 0.001 (1 in 1000)**	
Range of probability	P > 0.05		P < 0.05 (but > 0.01)		P < 0.01 (but > 0.001		P < 0.001
Commonly used phrase	Not significant		Significant		Very significant		Highly significant
Coded with asterisks	NS		*		**		***

Figure 1. Diagram to show how ranges of probability are coded to provide levels of significance. It is set out in the same way as tables of critical values for test statistics, reading columns from left to right with probability becoming smaller. If your probability is exactly on the threshold, then the interpretation is to the left. So $P \geq 0.05$ means not significant, $P < 0.05$ significant; $P \geq 0.01$ significant, $P < 0.01$ very significant, and so on.

4. **Proof and probability in statistical testing.** You *cannot* prove anything with statistics. You *can* compute the probability of obtaining a set of results like this (or more extreme) if the null hypothesis was true. If the probability is small, then you can proceed as if the unlikely event has not occurred, as if the result is genuine. We use certain thresholds of probability, arbitrary probabilities but well-accepted and conventional (see §12). The thresholds separate ranges of probability, as shown in Figure 1, and these ranges are interpreted as levels of statistical significance. See Appendix B for how to use the table of critical values for a test statistic to find the probability associated with your calculated value.

   It is important to realize that the probability associated with the test statistic is *not* the probability that any hypothesis is true (see §11 and Appendix A, §A2). It is the probability of obtaining this set of results or results more extreme than this *if* the null hypothesis *was* true. The true understanding is this. If the null hypothesis is true (i.e. there really is no difference or no association) then we would expect to obtain results like these (or more extreme), with this calculated probability. That is what "significant at P = some value" really means. There is more about P values and statistical significance in §11–§13, but for the moment we turn to how data are collected and used in statistical tests.

   If you obtain nothing else from this module, remember this, shouted here in a box.

> **You *cannot* prove anything with statistics. You *can* compute the probability of obtaining a set of results like this (or more extreme) if the null hypothesis was true. If the probability is small, then you can proceed as if the unlikely event has not occurred, as if the result is genuine.**

5. **Procedure.** Figure 2 summarizes the procedure as a flow chart. Notice the order: you must think of the question first, then find out what data will be required, then decide on which test will be appropriate, then consider the assumptions, and only then collect suitable data. (Well, that is the ideal, not always attained in practice.)

Decide on the question being asked.

↓

Decide on the data to be obtained.

↓

Decide which statistical test will be needed.

↓

Consider how the assumptions for the test will be satisfied.

↓

Gather the data.

↓

Arrange the data on paper in a meaningful way, i.e. classified into groups (for chi-squared, t-test or ANOVAR) or in pairs (correlation, regression).

↓

Decide whether to analyse
by hand and calculator **or** by computer package.

↓	↓
Find the formulae for the test.	Enter the data on the computer datasheet, arranged in columns.
↓	↓
Calculate the test statistic.	Specify the statistical test, usually from a menu.
↓	↓
Work out the degrees of freedom.	The package calculates the test statistic, and degrees of freedom.
↓	↓
Compare the calculated test statistic with the value in statistical tables.	View and print the test statistic and its associated probability.
↓	↓

Interpret the probability for statistical significance, and come to a conclusion about the difference or association of the initial question.

Figure 2. Flow chart of the procedure for statistical testing.

6. **Choosing a test.** Refer to Table 1 to choose a suitable test for your question and data. An outline of each test, including an example, is given on pages 11–16.

Table 1. Guide to statistical tests, arranged by their purpose.

Purpose	Statistical test
Describe the data and make simple inferences.	**Descriptive statistics and simple inferences.** Descriptive statistics include mean and standard deviation, and the frequency distribution of the observations. Simple inferences require calculation of standard error, looking up a value of Student's t, and calculation of confidence limits. Descriptive statistics and simple inferences are not statistical tests as such but do use Student's t which is used elsewhere as a test statistic. From the sample of data you can make inferences about the population that the sample represents, from which it was taken.
Look for association.	
Classification of attributes.	**Chi-squared test**, which looks for association between categories, cross-classified into groups, using the actual counts. The test statistic is chi-squared, χ^2.
Between two measured attributes.	**Correlation.** The two attributes must be measurable, continuous quantities that can be plotted on a scatter graph. The test statistic is the correlation coefficient, r.
Compare means.	
Only two groups.	**The t-test.** The means come from the two groups. The test statistic is t, Student's t.
Two or more groups.	**Analysis of variance** (abbreviated to ANOVAR or ANOVA). The means come from different groups of cases (or items or subjects) in the set of data. The test statistic is the variance ratio, F.
Predict one quantity from another.	**Regression.** The two quantities are measurable, continuous attributes, related by a straight line (in simple regression). We obtain the line of best fit to draw on a scatter graph with the equation for the line. Analysis of variance is used to test whether the regression is statistically significant (test statistic is F). You will also see the coefficient of determination, r^2, which indicates the proportion of variation explained by the regression.

7. **A worked example.** References to the procedure in Figure 2 are in italics, and technical terms in statistical testing (mentioned in §3) are in bold. This example is imaginary but plausible.

A researcher suspects that the voles on an island are larger than those on the mainland (*initial question*); they are known to be the same species, and adult males and females do not differ in weight. It will be necessary to catch some voles and weigh them (*data to be obtained*). Variation in vole weights is expected so a statistical test that compares the average weights of island voles and mainland voles is required. Reference to Table 1 shows that the t-test is appropriate for comparing the means of the two groups (*decide on suitable statistical test*) with the **null hypothesis** that the two groups do not differ in mean weight. The *assumptions* are considered for a t-test (p. 14) and used to write the protocol for fieldwork. Traps are set in widely scattered locations in vole habitat on the island and on the mainland. The trapped voles are weighed (*data gathered*) and released. Traps are moved after each successful operation to minimize the chance of catching the same vole again. The data and calculated quantities are given in Table 2 (*data on paper; calculations for test statistic and for degrees of freedom*).

Table 2. Weights of voles captured on an island and on the adjoining mainland, and calculations for the t-test.

	Weight of vole (g)	
	Island	Mainland
	118	119
	126	130
	130	120
	125	123
	120	121
	131	129
	122	114
	124	121
	121	118
		115
		117
n	9	11
mean	124.11	120.63
SD	4.4001	5.1239

Difference between the means 3.4747
Pooled variance = 23.191
SE of the difference = 2.1645
t = 1.605
Degrees of freedom = $n_{island} + n_{mainland} - 2 = 18$
Probability > 0.10 (from statistical table; in fact P = 0.126 from a computer package) so not significant

The 95% confidence limits for the *difference* (3.4747 g) are –1.0729 to 8.0223 g. Since the statistical test was not significant at $P \geq 0.05$, the limits include zero.

The researcher concludes that there is no significant difference between the weights of voles on the island and the mainland ($t = 1.61$, 18 degrees of freedom, $P = 0.126$; two-sample t-test, homogenous variances). (The sentence above includes the **statistical significance**, the value of the **test statistic**, the **degrees of freedom**, and the **probability associated with the calculated test statistic**. The interpretation is given first and then the evidence for it as the details of the statistical test in brackets. There are several versions of the t-test so for completeness the particular version, here the commonest, is named.)

Points to note. The researcher is sampling from two populations (it is thought) so in each location the traps are widely spaced to minimize capture of related voles which would not be independent observations (families of voles could be consistently smaller or larger than the mean). The procedure to avoid recapture was mentioned above, and this also reduces the chances of capturing a vole related to a previously trapped one. Consequently the observations are random and independent samples from the population, as far as is practicable. This is an important assumption for the t-test. Others are about homogeneity of variance (satisfied here because the standard deviations for the two groups are very close) and Normality (hard to tell with such small samples but each group has a central peak in its frequency distribution even if it is not much like the silhouette of a bell).

Note that we cannot say there is *no* difference in weight: clearly there is, with sampled island voles 3.47 g heavier on average. But we say that this is not significant statistically (our interpretation of the results of the t-test) which means that we will act as if there is no difference. We believe that this observed difference in weight in our data arose by chance alone. Moreover, we have calculated the probability that if there was in fact only one population of voles, spread across the mainland and the island, we could obtain two samples like this with this difference in weight or a more extreme difference. The probability is the one associated with the test statistic, i.e. $P = 0.126$. Expressed in another way, there is a better than one in ten chance (about one in eight) of obtaining samples like this, that have this difference in weight or a larger one, when there really isn't a difference.

8. **Statistical significance and practical significance.** Statistical significance can be paraphrased as "is it true?" or more correctly "is it likely to be true?". Practical significance is "does it matter?" or "is it worth taking account of, basing decisions on?". Practical significance is the more general term but you will see this kind of significance spoken of in a particular context, e.g. is it educationally significant, or clinically significant (in medical research), or archaeologically significant, or biologically significant?

Returning to the vole example above, it is possible to calculate how large the samples would need to be for a difference in weight of 4 g to be seen as statistically significant (i.e. with probability associated with the test statistic, t, less than 0.05—see Figure 1). It turns out that with two samples of 13 voles each, assuming the same variability as found before, then a difference of 4 g would be significant ($P < 0.05$). This is a common finding: larger samples may give a statistically significant result because the larger value of n affects the calculation of standard error and also, through the degrees of freedom, the critical value of the test statistic. If the samples of voles were 22 each from island and mainland then a 4 g difference could be detected with $P = 0.01$, i.e. it would have to be a hundred to one chance to get such samples if the difference in weight did not really exist.

However, is a difference in weight of 4 g of biological significance, given that it is about 3.3% of the average weight of voles? The weights of small mammals are variable during the day and over a season depending on food availability and temperature, and could easily change by 10% in an individual vole. In this context, the difference of 4 g, even if statistically significant, may have no biological significance from the point of view of assessing whether the voles were better fed on the island than the mainland.

On the other hand, we may be wondering if island voles are heavier than mainland ones, as often found, so we look for evidence that the island population has started to diverge from the mainland one through its isolation. In this case, a difference of 4 g could be biologically significant. A divergence has to start small and become larger, so a statistically significant difference could be seen as identifying the start of the divergence process.

There are no easy answers to these problems about statistical and practical significance. You need to be aware of the topic in case it turns up in your own research in the future. Another way of looking at this problem is the difference between hypothesis testing (to obtain statistical significance) and estimation of size of effect (to assess practical significance—see also §14). For example, in the vole example above, we can focus attention on the size of the difference, where we find that the 95% confidence interval is –1.07 to 8.02 g. These limits include zero so we have to conclude (with 95% limits) that we have not detected a significant size of effect, just as we concluded from the statistical test.

9. **The conclusion of a statistical test.** It is important to write a clear statement of results from a statistical test. This will include the calculated value of the test statistic, the degrees of freedom, the probability associated with the test statistic, and *your* interpretation of that probability as a statistical significance. (These are the terms highlighted in §3, the components of statistical testing.) If it is not evident from the test statistic named, or there are several versions of the test, state which statistical test was used. When you have an exact probability, from a computer package, then quote that in the statement of results. Otherwise, give the value with a < or > sign, as in Figure 1. In statistics, it is sufficient to say, for example, $P < 0.05$ without specifying also > 0.01 because it is understood that if P was < 0.01 you would have said so.

Example for a chi-squared test. Although the proportion of left-handed males (20%) is greater than for females (10%), this was not statistically significant ($\chi^2 = 2.550$, 1 d.f., $P > 0.05$). (The symbol, χ^2, is chi-squared, i.e. the Greek letter chi, squared.)

Example for a correlation. There is a significant correlation between monthly totals of rainfall and sunshine ($r = -0.699$, 10 degrees of freedom, $P = 0.0115$).

Example for a t-test. The mean interpupillary distance in males was 2.89 mm greater than in females and this was highly significant ($t = 5.22$, 141 d.f., $P < 0.001$; two-sample t-test, homogeneous variances). (Since there are several versions of the t-test, it is helpful to specify which one was used.)

10. **Parametric and non-parametric tests.** All the statistical tests given in Table 1, except chi-squared, are called parametric tests because the test statistic is calculated from the actual values of the data, the measured values. There are other tests, called non-parametric, which work on the ranks of the data, i.e. the place in order from smallest to largest, instead of the actual value. On the whole, parametric tests are to be preferred because they use all the information that you have. Non-parametric tests are required when the data really do not conform to certain assumptions necessary for parametric tests to provide correct results, i.e. an accurate value of the probability associated with the test statistic. In this module we have not included non-parametric tests because of the limited time available for teaching statistics. If non-parametric tests turn out to be needed for your research project then you will need to look them up in statistical textbooks, and seek help from a statistician if necessary.
11. **Correct understanding of P values.** It is hard to grasp the correct use of probability (P value) in statistical testing. The definition of P value is: "the probability of the observed data (or data showing a more extreme departure from the null hypothesis) when the null hypothesis is true" (Everitt (1998) Cambridge Dictionary of Statistics). See also Appendix A, §A2.

    The probability associated with the test statistic provides a quantitative measure of the evidence about the null hypothesis. If the probability is small, we are in a position to not accept the null hypothesis because there is a lot of evidence (in the data) against it. "Not accept" is the recommended phrase over "reject" which is too definite. The question is: how small must the probability be? The practice has evolved of using P = 0.05, 0.01 and 0.001 as threshold or boundary values of probability (see Figure 1). Other people may refer to these values of probability as cut-off levels or cut-offs. By using these thresholds we are turning a quantitative measure into a qualitative one, i.e. categories, so that we can make a decision, one way or the other, about whether the result is to be regarded as genuine. (In a similar way, the quantitative score in goals of a football match is converted to a qualitative measure, categories, of win, lose or draw.) The categories or levels of statistical significance (Figure 1) are your interpretation of the P value. Whenever possible, provide the exact probability so that readers can apply different thresholds if they wish (see Appendix B, §B1).
12. **Why use P = 0.05 as the first threshold of statistical significance?** The value of P = 0.05 is entirely arbitrary but has become a well accepted convention. All that can be said to justify 0.05 is that its use over many years in research work in all fields has been successful, on the whole, in identifying interesting results. The one in twenty chance of being misled has not held back progress in research.

    Moreover, the use of P = 0.05, 0.01 and 0.001 thresholds was a necessity in the days before computers. It was not practicable to calculate the exact probability so statistical tables were used (see Appendix B) to find a range of probability for a particular calculated value of the test statistic with its degrees of freedom.

13. **What "not significant" means.** If your calculated probability is 0.05 or greater and you interpret this in the conventional way as not significant, what does this mean about your results? It is better to say that "a difference could not be detected" (a difference between means for example), or that "there was no detectable association" (in a chi-squared test), or "no detectable correlation". This is preferable to "there was no effect" (of experimental treatments or between groups in a survey or relationship between measurements or categories). There could be an effect but we simply failed to find it (as discussed in the vole example, §7). To use the well known adage: absence of evidence is not the same as evidence of absence.

    Not significant is a broad category. It could mean that:

    (a) there is really no effect; or

    (b) there is a small effect but it is lost in the background variability (noise); or

    (c) there is a large effect but not detectable in this set of data because there was also a large amount of background variability.

    This last case can be common in educational, medical or biological research where people and organisms are intrinsically so variable.

    Whenever possible give the exact probability (see Appendix B, §B1). The P values of 0.055 and 0.55 would each be interpreted as not significant but they have different implications. If $P = 0.055$ then you might think that this is a near miss. If you had more degrees of freedom (from larger samples or more replication) then you might have detected an effect. You might try the survey or experiment again with more degrees of freedom; or you would not be surprised if similar surveys or experiments *had* been successful in detecting an effect. In the case of $P = 0.55$ you would probably conclude that there was genuinely no effect, or such large amounts of background variation as to completely conceal an effect even with more degrees of freedom.

14. **Effect size versus hypothesis testing.** The procedure for statistical tests has been set out in terms of testing hypotheses (§2–4). It is also important to look at the size of the effect; effect size and hypothesis testing are complementary. This is easily done for the t-test or analysis of variance where we are comparing the means of treatments or groups. We calculate the difference between means and the confidence interval for the difference, as in Table 2, bottom right. The advantage of effect size is that we can see the direction of the effect (island voles heavier) and the size (by 3.5 g), plus make a decision on statistical significance based on the 95% confidence limits (–1.07 to 8.02). Since these limits include zero, we interpret the results as not significant. We assert, in the language of confidence limits, that the true value of the difference lies in the range –1.07 to 8.02 g unless a 1-in-20 chance has occurred. Since this range includes zero, and some values where the difference is in the reverse direction, we conclude that it is not significant.

    Setting out clearly the effect size and its confidence limits also leads to consideration of practical significance. Read §8 again because this is an important topic to understand.

15. **Outlines of statistical tests.** On the six pages following there is a one-page outline of descriptive statistics and simple inferences, chi-squared test, correlation, t-test, analysis of variance, and regression. A standard set of headings is used. The outline provides enough details to enable you to understand what the test does, when and how it is applied, and how the result is interpreted. This will help you to assess use of these tests in the research literature that you read. For your own research you may need further details which can be found in textbooks. To carry out the tests, the computer package MINITAB (available on the university network) is probably the easiest one for beginners to use.

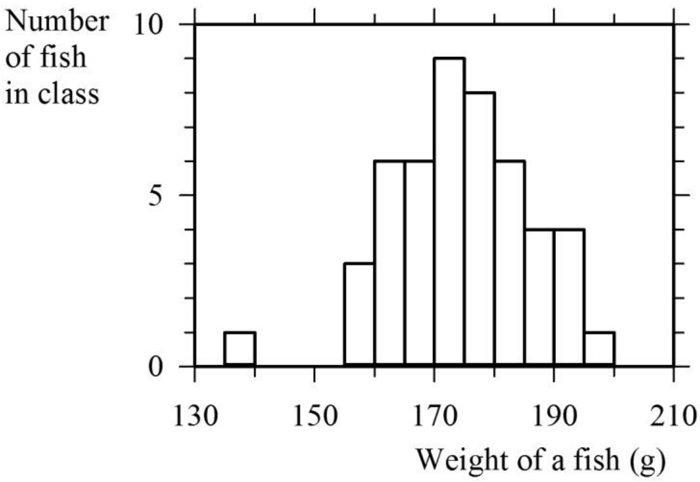
Descriptive Statistics and Simple Inferences

1. **Purpose.** To describe a set of data, and then to make inferences about the population from which the sample of data comes.
2. **Data.** Measurements for a number of cases or items or subjects; a single group.
3. **Null hypothesis.** Not applicable.
4. **Test statistic.** Not applicable, but Student's t is used for calculating confidence limits.
5. **Degrees of freedom.** For a sample of n cases, df = n–1.
6. **Assumptions.** Data from a random and independent sample of the population, if making inferences about the population. Measurements Normally distributed if using confidence limits of the observations (Normal distribution not necessary if using confidence limits of the mean and sample size larger than ten).
7. **Example.** Weights of a cohort of salmon caught on migration shown in a frequency distribution, and with descriptive statistics and simple inferential statistics. In the graph, classes are 135.0–139.9 g, 140.0–144.9 g, and so on.

Descriptive statistics

n = 48
Mean 174.0 g
Median 174.0 g
Modal class 170–175 g
Standard deviation 11.500 g
Range 136–196 g
Coefficient of variation
(SD/mean) 0.066092 or
6.61%

Frequency distribution



Simple inferential statistics
Standard error 1.6599 g
df = 47
Value of t 2.021
95% confidence limits 170.6–177.4 g (using t for 40 df, the closest value at hand)

Statement of results. The mean weight of salmon was 174.0 g (SD 11.500 g, n = 48; 95% confidence limits of the mean 170.6–177.4 g).

8. **Variations and elaborations.** Measures of skewness (asymmetry, especially in the tails) and kurtosis (thickness of the tails) can describe a distribution further in terms of departure from Normality. There are other distributions, e.g. Poisson, binomial, which can also be described by mean and standard deviation.
9. **Equivalent non-parametric test.** The median is a measure of central tendency that is obtained from the ranks so is non-parametric; quartiles and interquartile range, or use of centiles, are non-parametric measures of spread of a distribution.

The Chi-squared Test

1. **Purpose.** To look for association between categories, cross-classified into groups.
2. **Data.** Counts of occurrence in the groups, in a cross-classified table of c columns and r rows for the counts (not including marginal columns and rows for descriptors or totals).
3. **Null hypothesis.** No association between categories; cases or items or subjects occur in the groups independently of the categories.
4. **Test statistic.** Chi-squared, χ^2.
5. **Degrees of freedom.** For cross-classified table of c columns and r rows, df = (c – 1) × (r – 1).
6. **Assumptions.** Data from a random and independent sample of the population where factors affecting classification into categories acted uniformly (no identifiable sub-groups where the factors acted differently).
7. **Example.** Are handedness and sex associated? The data are numbers in each group classified by handedness and sex, from undergraduates taking APS240 in October 1993.

		Sex		
		Male	Female	Total
Handedness	Left	10	7	17
	Right	39	63	102
	Total	49	70	119

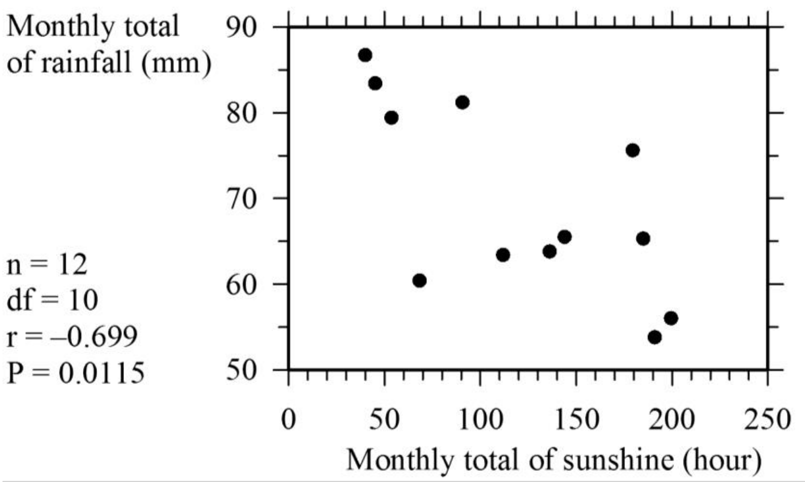
2 × 2 contingency table; df = 1
Chi-squared (χ^2) = 2.550
P = 0.110, NS

Statement of results. Although the percentage of females that is left-handed (10%) is lower than for males (20%), in a set of data of this size (119 individuals) handedness and sex are not associated, they occur independently (χ^2 = 2.550, 1 degree of freedom, P = 0.110).

8. **Variations and elaborations.** If counts are low, especially expected count (see details of method in textbooks) below 5, then beware of modifications (e.g. Yates's correction) to take account of this. The version of the test outlined above is the contingency test. Another version is the goodness-of-fit test where instead of a null hypothesis of no association there is a null hypothesis to test particular frequencies in each group. This hypothesis may come from theory or previous experience. The data are organized in a table with the actual counts in the groups as one column (or row) and the expected counts (from the hypothesis of the frequency in each group) as another column (or row). The degrees of freedom are the number of groups minus one.
9. **Equivalent non-parametric test.** Not applicable: the chi-squared test is a non-parametric test.

Correlation

1. **Purpose.** To look for association (linear correlation) between two measured attributes (Pearson's correlation).
2. **Data.** Two measurements from each case or item or subject, that can be plotted on a scatter graph (does not matter for the statistics which measurement is x). *Always* plot the graph, even if only a sketch.
3. **Null hypothesis.** No (straight-line) association between the two measurements; the points on the graph are a random cloud or occur in a horizontal or vertical line.
4. **Test statistic.** The correlation coefficient, r, measures the strength of association, between 0 for no association to –1 for perfect negative association, to +1 for perfect positive association. Test of significance is usually based on Student's t-test, and tables of critical values of r are available for direct assessment.
5. **Degrees of freedom.** For sample of n cases, df = n–2. Note that there are n *cases* (or items or subjects), each case with two numbers.
6. **Assumptions.** Data from a random and independent sample of the population, where there is an underlying straight-line relationship between the measurements. The measurements are distributed bivariate Normally.
7. **Example.** Scatter graph of sunshine and rainfall recorded at Sheffield (monthly means for the period 1981–2010), and the correlation between them.



   Statement of results. There is a moderately strong correlation (r = –0.699) between monthly totals of rainfall and sunshine; the correlation is significant (10 degrees of freedom, P = 0.0115). It is a negative correlation, i.e. high sunshine tends to be associated with low rainfall.

8. **Variations and elaborations.** None: correlation is a simple and fairly crude technique. If there is a statistically significant correlation then it may prompt further investigation and gathering of data, for example to predict one measurement from the other using regression.
9. **Equivalent non-parametric test.** There are two methods of correlation when the data are ordinal: Spearman's and Kendall's rank correlations.

The t-Test

1. **Purpose.** To compare the means of two groups.
2. **Data.** Measurements for the cases or items or subjects in the two groups.
3. **Null hypothesis.** No difference between the means of the two groups; the groups are samples drawn from the same population.
4. **Test statistic.** Student's t.
5. **Degrees of freedom.** For two groups of n_A and n_B cases, df = $n_A + n_B - 2$.
6. **Assumptions.** Data from random and independent samples of the populations, where the measurements have underlying Normal distributions which are equal in variability.
7. **Example.** Is the distance between the eyes different in males and females? The data are the interpupillary distances measured when setting up a binocular microscope by undergraduates taking APS116 in October 2005.

   Summary of data.

	Males	**Females**
Number	53	90
Mean	62.36 mm	59.47 mm
Standard deviation	3.470 mm	3.032 mm

   $n_A = 53$, $n_B = 90$; df = 141
   Student's t = 5.22
   P = 0.00000063, ***

   Statement of results. The difference in interpupillary distance is statistically highly significant (t = 5.22, 141 degrees of freedom, P = 0.00000063; two-sample t-test, homogeneous variances). On average the interpupillary distance is 4.6% smaller in females than in males. The difference is 2.89 mm with 95% confidence limits of 1.79 to 3.99 mm.
8. **Variations and elaborations.** For marked unequal variability of the two groups, especially with small groups or unequal size groups, then use a t-test designed for this, known as separate variances t-test, Behrens–Fisher test, Welch test or Satterthwaite-adjusted t-test, These tests are best carried out with a computer package.

   The confidence limits for the difference between the means can also be calculated, to examine practical significance (and put more emphasis on the size of effects than on hypothesis testing). If the confidence limits include zero then the difference will not be statistically significant (at the same probability, e.g. P = 0.05 for 95% confidence interval).
9. **Equivalent non-parametric test.** The non-parametric test that is analogous to the t-test is called the Mann–Whitney test or U-test or Wilcoxon test or Mann–Whitney–Wilcoxon test.

Analysis of Variance (ANOVAR)

1. **Purpose.** To compare the means of two or more groups.
2. **Data.** Measurements for the cases or items or subjects in the groups.
3. **Null hypothesis.** No difference between the means of the groups; the groups are samples drawn from the same population.
4. **Test statistic.** Variance ratio, F.
5. **Degrees of freedom.** The variance ratio has two degrees of freedom, first for the hypothesis being tested and second for the residual variation, spoken, for example, as "5 on 26 degrees of freedom". The first df is the number of groups minus one in the initial analysis, or the number of means being compared minus one in further analyses. The residual df is calculated from the ANOVAR table; in the simplest example of k groups each with n cases, residual df = k(n – 1).
6. **Assumptions.** Data from random and independent samples of the populations, where the measurements have underlying Normal distributions which are equal in variability (technically, the residuals (observation minus treatment mean) must come from one Normal distribution). Differences between group means arise from a small amount added or subtracted, not so large as to be multiplied or divided by a (mathematical) factor.
7. **Example.** Results from an experiment on growing potatoes with different amounts of fertilizer (imaginary data).

Table of treatment means

Fertilizer (kg/ha)	0	50	100	150	200
Yield (t/ha)	20.5	29.5	37.0	39.5	39.0

Table of Analysis of Variance

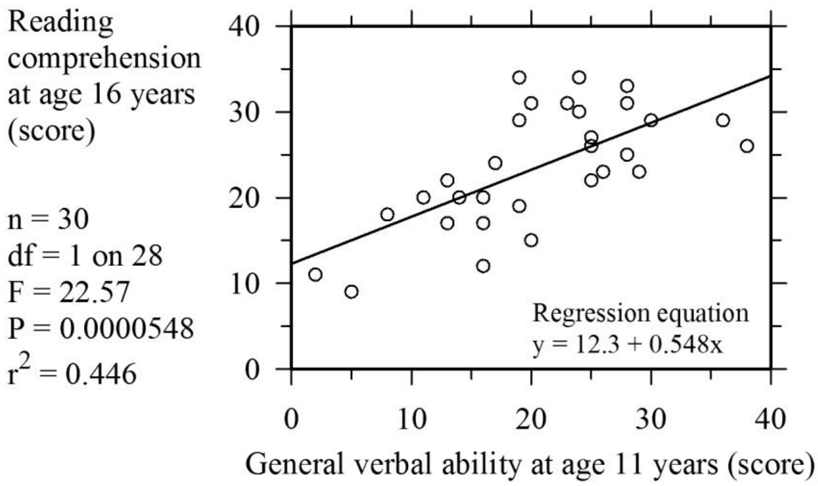
Source	Degrees of freedom	Sum-of-squares	Mean square	Value of F	Probability
Treatments	4	1050.80	262.70	5.53	0.0061 **
Residual	15	713.00	47.53		
Total	19	1763.80			

Statement of results. There is a very significant effect of fertilizer on yield: yield increases with increasing amount of fertilizer applied, especially at lower rates of application (F = 5.53, 4 on 15 degrees of freedom, P = 0.0061).

8. **Variations and elaborations.** Analysis of variance is a very general technique which has many variants for different sorts of experiments and different situations. The aim is always to calculate the background, uninteresting variation correctly and compare it with the variation for which there is an explanation (experimental treatments or groups identified in a survey).
9. **Equivalent non-parametric test.** The non-parametric tests that are analogous to ANOVAR are the Kruskal–Wallis test and the Friedmann test (each for particular versions of ANOVAR).

Regression

1. **Purpose.** To predict one measured attribute from another by fitting a line to the data and using the equation of the line.
2. **Data.** Two measurements from each case or item or subject, that can be plotted on a scatter graph with the measurement to be predicted (response or dependent variable) as y. *Always* plot the graph, even if only a sketch.
3. **Null hypothesis.** The slope of the fitted line is zero, i.e. the predicted y value will be constant whatever the value of x.
4. **Test statistic.** Variance ratio, F, for the statistical significance of the regression. The coefficient of determination, r^2, indicates the proportion of variation explained by the regression.
5. **Degrees of freedom.** The variance ratio has two degrees of freedom, first for the regression, always one df, and second for the residual variation, spoken, for example, as "1 on 26 degrees of freedom". The residual df for a regression is n–2, where n is the number of *cases* (or items or subjects) with paired values (x and y).
6. **Assumptions.** Data from a random and independent sample of the population, where there is an underlying straight-line relationship between the measurements (in the ranges of x and y used; extrapolation beyond these ranges is unreliable). The x values known without error so that all the variability in the point on the graph is in the y value. The variability of the y values is homogeneous (the same all along the x-axis, between smallest and largest value of x) and Normally distributed.
7. **Example.** Regression of score of reading comprehension when aged 16 years on score of general verbal ability when aged 11 years, for 30 children drawn at random from the National Child Development Study (children born in 1958).



Statement of results. The regression of reading comprehension (age 16) on verbal ability (age 11) is highly significant (F = 22.57, 1 on 28 degrees of freedom, P = 0.0000548). The reading comprehension (age 16) can be predicted from the equation $y = 12.3 + 0.548x$ where y is the score for reading comprehension (age 16) and x is the score for general verbal ability (age 11).

8. **Variations and elaborations.** This is simple regression for fitting a straight line. The method can be elaborated for curves of all sorts, and for more than one x value being used to predict y (multiple regression).
9. **Equivalent non-parametric test.** There are non-parametric methods for regression but no simple equivalent for the straight-line regression used here.

Appendix A
Background information

A1. **Degrees of freedom.** Even statisticians have trouble with this term. Brian Everitt (The Cambridge Dictionary of Statistics, 1998) starts his technical definition: "an elusive concept that occurs throughout statistics"! The degrees of freedom indicates how many independent pieces of information there are in the data, or how many opportunities there are for the data to vary. Loosely speaking, the number of degrees of freedom is a method of saying how large the set of data is, in a way that is relevant to the particular statistical test. We are calculating the probability of getting a certain result (our result or a more extreme one) if the null hypothesis is true. Clearly this probability will depend in part on the size of the set of data, in other words, on how many possibilities there are of obtaining this result. The larger the set of data, the more possibilities there are.

A2. **Frequentist and Bayesian statistics.** The classical statistics that we are using is called frequentist because it uses the overall frequency of events as the probability of a single event occurring. That is why the explanations have phrases such as "if we sampled repeatedly" or "if we performed this experiment many times" or "in the long run". It leads to the results of statistic tests being couched in what seems convoluted terms of the probability of this set of data (or one more extreme) occurring if a particular hypothesis was true. That is why it is not possible to assign a probability to any particular hypothesis, only to the data given the null hypothesis.

An alternative approach is to take the set of data as given and find the probability of the hypothesis. This is the province of Bayesian statistics which is slowly becoming more widely used in many research fields. You need to be aware of the Bayesian approach because it may turn up in research that you are reading, or you may find that you need to learn about it and use it yourself. Bayesian statistics requires a deeper knowledge of probability and mathematical logic.

The main reason for knowing of the difference between frequentist and Bayesian statistics is that it helps to understand the exact meaning of the probability produced in statistical testing (§4). Frequentist (classical) statistics provides the probability of obtaining data like the results (or more extreme) given the null hypothesis (usually that there is no difference between means, or no association). In contrast, Bayesian statistics provides the probability of the hypothesis given the data found.

Appendix B
Using statistical tables

B1. **From probability to significance.** Computer packages analysing data provide an exact figure for the probability associated with the calculated test statistic. You can interpret it as a level of statistical significance using Figure 1. You need to think carefully about the range in which your value of probability falls ($P \geq 0.05$, or $P < 0.05$ but ≥ 0.01, and so on). Give the exact probability figure when writing the conclusion of the statistical test, whenever you can, because it is part of the evidence for your conclusion (see §13). Rounding off to three significant figures is a suitable measure of precision for these probabilities.

If you calculate the test statistic by hand and calculator then you can use statistical tables to find the range of probability associated with the value—see below. Once you have the range of probability, it is easy to interpret this as statistical significance using Figure 1. Alternatively, there are resources on websites which will calculate the exact probability given your value of the test statistic and one or two other inputs, typically degrees of freedom.

There are many of these resources and currently I prefer the one at http://danielsoper.com/statscalc3 (look for "Probability (p-Values)" and then for the relevant test statistic). Be sceptical: always check one of these resources with a couple of entries where you know the answer from a table of critical values. For instance, suppose you use Daniel Soper's p-Value Calculator for a Student t-Test. Try entering 9 degrees of freedom and $t = 2.262$ (from a t-table, column for $P = 0.05$). Two answers are produced, one-tailed and two-tailed, and the two-tailed value of probability is 0.05001285 which is close to the 0.05 that you are expecting. (We have not used one-tailed statistics in this module—consult statistical textbooks if you wish to find out more). Now we have confidence in the resource, and realize that we will need to use the two-tailed value. For this reason—to check web-resources—knowing how to use statistical tables is still a valuable skill, so read below.

B2. **How to use a table of critical values.** There is a separate table of critical values (as they are called) for each test statistic, often found at the back of statistical textbooks. Beware of statistical tables obtained from websites because they may not be the common ones that you need but more specialized tables. It is safer to use those at the back of statistical textbooks. In particular, you need the table for two-tailed t-tests, not one-tailed, for work in this module.

Start with the column (sometimes a separate table) for $P = 0.05$. Enter the table using the number of degrees of freedom for the test statistic. (Sometimes df is denoted by ν, the Greek letter nu, lower case. It looks remarkably like a letter v but is actually nu; see here in larger size: ν (nu) versus **v**.) If the calculated value is larger than the tabulated value (for a given P value and degrees of freedom) then the result is declared statistically significant, at the stated probability. If significant with the $P = 0.05$ column, you can proceed to the next column for $P = 0.01$ and then to $P = 0.001$, to find the lowest probability for your calculated test statistic. This probably sounds obscure but once you have used statistical tables a few times you will get the hang of it.

B3. **Worked example of using a statistical table.** Suppose you have a correlation coefficient, r, of 0.756 with 12 degrees of freedom (from the 14 observations, since df = n–2 for correlation). A portion of the statistical table for the correlation coefficient is shown in Table B1. Find the row for 12 degrees of freedom. The critical value tabulated for $P = 0.05$ is 0.532 and the calculated value (0.756) is much larger than this so we have significance at least at $P < 0.05$. We go further, in this table literally further along the row for 12 degrees of freedom: the critical value for $P = 0.01$ is 0.661, again a hit; but at $P = 0.001$ the table has 0.780 and the calculated value does not exceed this. So the probability associated with this value of r (12 df) is $P < 0.01$ (but not $P < 0.001$) so you can claim that it is very significant, and code as ** (from Figure 1).

Table B1. Part of the table of critical values for the correlation coefficient, r. Extracted from Parker, R.E. (1979) Introductory Statistics for Biology, London (Arnold).

Degrees of freedom	**Probability** 0.05	0.01	0.001
1	0.99692	0.999877	0.99999877
2	0.950	0.990	0.999
…			
…			
11	0.553	0.684	0.801
12	0.532	0.661	0.780
13	0.514	0.641	0.760
…			
…			
90	0.205	0.267	0.338
100	0.195	0.254	0.321

P.L. Mitchell, Dept. of Animal and Plant Sciences, and Mathematics and Statistics Help (MASH), University of Sheffield.
21st February 2018. P.L.Mitchell@Sheffield.ac.UK