# REASONING WITH DATA OR MATHEMATICAL STATISTICS? IS THE UK MOVING IN THE RIGHT DIRECTION?

James Nicholson
SMART Centre, Durham University, Leazes Road, Durham, DH1 1TA, UK
j.r.nicholson@durham.ac.uk

*Reasoning with data has been recognised as an important life skill for a number of years, but the statistics curriculum at school level has done little or nothing to prepare our young people to meet the challenges of working with large multivariate datasets. New curricula proposed for England seem to place an increasing emphasis on statistics as a branch of mathematics, and less on its role of making sense of the complex evidence in the world around us. The paper outlines the proposed curriculum developments, and discusses the proposed developments in the context of developments in the wider statistics community in respect of the use of hypothesis testing. It compares the direction of curriculum change with that of New Zealand and discuss the challenges facing the successful implementation of the new curriculum.*

## BACKGROUND

There are a number of contextual factors to be considered when a curriculum review is instigated: whether there have been any significant changes in the world which are relevant to the review; what the current curriculum is; what the bureaucratic mechanisms are for implementing change; how well the current curriculum meets its objectives; whether the objectives for the new curriculum should remain as they are for the current curriculum or if changes elsewhere suggest that the objectives need to be revised; whether there are any examples of good practice developed elsewhere which could be adopted or adapted in the review of this curriculum.

Access to real data is becoming easier; datasets are typically larger than they used to be; datasets typically contain a number of related variables. It seems inevitable that these trends will continue and indeed accelerate.

In the United Kingdom, the school curriculum is determined at national level. Wales and Northern Ireland no longer follow exactly what England does, and Scotland have had a separate system for some time, so this paper will refer to the proposed new curriculum for mathematics and statistics in England. Compulsory education in the UK has been extended to age 18. Students normally take GCSE qualifications (General Certificate of Secondary Education) at age 16 in a range of academic or vocational subjects – often in 8 to 12 subjects. The academically capable will take GCE qualifications (General Certificate of Education) at age 18 in a much smaller number of subjects – typically 3 or 4 though now only a small number take 4 subjects. The government have introduced funding mechanisms and other drivers designed to encourage all students aged 16 – 18 still to be studying something with a strong quantitative component. This may be retaking a GCSE qualification in mathematics, which is a required qualification for many jobs, taking GCE level mathematics or a new qualification being introduced under the collective label of Core Maths, designed to prepare students for further study and careers which don't require such a mathematical focus as GCE Mathematics provides but where mathematical knowledge and its application are nonetheless important, such as geography, business and the social sciences.

Nicholson, Ridgway & McCusker (2006) and Ridgway, Nicholson & McCusker (2007) argued that reasoning with data was already pervasive in society; that evidence-based (or evidence-informed) policy requires an understanding of the data relating to complex social issues in which there are many factors at play; that the statistics curriculum in the United Kingdom at the secondary level did not prepare our young people adequately. The assessment of the curriculum is focused on procedural competencies of standard techniques with little emphasis on modelling or interpretation.

The value of hypothesis testing as a research tool is being called into question. Gorard (2014) not only criticises the way researchers apply it in inappropriate contexts, and often misrepresent the conclusions of significance tests, but he argues that the theoretical basis for hypothesis tests is fundamentally flawed and that even when researchers follow the technique's requirements, it is worthless. Trafimow & Marks (2015) announced in an editorial that the journal

*Basic and Applied Social Psychology* would no longer accept articles using the null hypothesis significance testing procedure, on the basis that the procedure is invalid.

New Zealand undertook a fundamental review of their mathematics curriculum for schools which resulted in them renaming the subject area (across both primary and secondary levels) as Mathematics and Statistics. Central to the revision was the introduction of multivariate data at an early age becoming more sophisticated through primary level and into secondary level, and the use of technology to support teaching and learning.

PROPOSED NEW CURRICULUM IN THE UK
*The place of statistics within mathematics*

Since 2004 GCE Mathematics has been a modular qualification, with candidates taking 4 modules of pure mathematics and two applied modules – chosen from one or two branches out of decision mathematics, mechanics and statistics. There was also a half course (AS level) available where candidates took two pure and one applied module, and if this course was successfully completed, candidates could defer the award of the AS level and carry forward those results to be combined with 3 further modules for a full GCE qualification ('A-level'). The UK has a number of awarding bodies who offer GCE Mathematics qualifications. The pure mathematics modules are almost identical because the government stipulate the pure mathematics to be covered in the common core for both AS and A level. However, each awarding body could propose the content for applied modules, which had to be accredited by the government for them to be able to offer the qualification. The different awarding bodies developed different modules in decision mathematics, mechanics and statistics with the statistics modules showing the greatest amount of variation in content, and in topic order. Schools and students could then choose which awarding body's course to enter for and which applied modules to take.

Apart from changes in content, the new curriculum (Department for Education, 2014) has some important changes in structure. The courses will no longer be modular in nature: all of the assessment will be carried out in a single examination session. This means that if a student does well in the AS level, and wants to continue to a full A level they will have to undergo assessment on the full course with the AS qualification becoming redundant. The new courses at AS and A level will have 100% prescribed content common to all the examining bodies, and it has been decided that this will involve no decision mathematics but some mechanics and some statistics.

*Statistics content within the new GCE courses*

The curriculum guidance prescribes the use of technology should permeate the detailed course specifications to be brought forward by each awarding body, and also that they must require students to become familiar with one or more specific large data sets in advance of the final assessment. This sort of pre-release material has not been commonly used before so it will be interesting to see how it works in practice.

Porkess (2012) identified that hypothesis testing is carried out in a number of other curriculum subjects at school and recommended that hypothesis testing should be included in the first statistics module in any A level. The modular framework has been removed but this recommendation has been adopted in defining the compulsory content for the new GCE Mathematics course. The section 9.3.3 on Correlation and Regression in Porkess (2012) is perhaps worth reproducing in full for the insights it offers to the rationale for other choices:

> *The idea of correlation is used in many subjects, with rank correlation more common than product moment correlation. There is, however, considerable scope for misunderstandings, leading to invalid work and conclusions. It is important that best practice is established in mathematics. This includes the use of correlation coefficients as test statistics so that students learn how their interpretation is related to sample size.*
> *In mathematics syllabuses, correlation is usually presented alongside linear regression. This can lead to misunderstanding of the different types of variables involved.*

The new GCE Mathematics course includes hypothesis tests for correlation coefficients, but not the calculation of either rank or product moment correlation coefficients, and nothing on any form of regression. Pedagogically, this is perverse – students are not expected to learn, and be assessed on, how to calculate correlation coefficients where they should experience the effect of varying the

sample size among other factors, yet Porkess explicitly states the aim of including correlation coefficients as test statistics is that students learn how their interpretation is related to sample size. The choice of content seems to have its roots in the technical conditions or assumptions which are attached to the techniques of least-squares regression and product moment correlation, and a desire to prevent misunderstandings and the use of a technique when it is not technically appropriate, yet it does not provide a pedagogically coherent body of knowledge in which the conceptual basis for the technique can be developed adequately. Moreover, the course content makes no mention of the conditions under which it is or is not appropriate to calculate a correlation coefficient – presumably it is not needed as students are never required to calculate one, and therefore never have to make any decision whether the conditions are met. It is difficult to see how this course will do anything to increase the likelihood of students being prepared to deal with these techniques appropriately when they do have to use them. Moreover, in a course with very limited statistical content, it is hard to justify this choice as the most important for inclusion.

The detailed statistics content is presented under 5 headings for A level with the AS course covering about half of the material, including all of the first two sections. The section headings are:

- Statistical sampling
- Data presentation and interpretation
- Probability
- Statistical distributions
- Statistical hypothesis testing

There seem to be some important structural weaknesses in the course:

- Data presentation and interpretation – includes 'Understand informal interpretation of correlation' but not any calculation of correlation, yet candidates will be expected to 'Interpret a correlation coefficient as a test statistic and use it in a hypothesis test'.
- Probability – conditional probability is not included until the second part of the course, yet candidates at AS level have to 'conduct a statistical hypothesis test for the proportion in the binomial distribution and interpret the results in context'.
- Statistical distributions – the Binomial (only at AS) and the Normal also at A level are the only distributions required. There is no general work on random variables or on the calculation of the mean and variance of probability distributions, much less anything about how distributions combine – so the sampling distribution of the mean of a Normal distribution will have to simply be stated.
- Statistical hypothesis testing – for the Binomial and the Normal and interpreting a correlation coefficient as a test statistic and use it in a hypothesis test. This is demanded without any discussion of conditional probability, sampling distributions or calculation of the correlation coefficient. As with the more extended discussion of the pedagogical weakness of excluding any calculation of a statistic which is to be used in a hypothesis test, there are glaring pedagogical weaknesses in testing the mean of a Normal distribution with no estimation in the course.
- There is no mention of estimation in a statistical context – no confidence intervals, or the development of the concept of a sampling distribution on which the logic of hypothesis testing is predicated. Appreciation of the variability of the sample mean and its relationship to the variability of the underlying population and the sample size is fundamental to the conceptual understanding of testing for the mean of a Normal distribution.

*Statistics content within the new Core Maths courses*

All students in England are now expected to continue their full time education until the age of 18, and to continue with some mathematical study in the final two years, where until recently the school leaving age was 16 and those continuing at school could choose not to engage with any numerate discipline. Students who have not achieved a 'good pass' at GCSE will be expected to take a course leading to a similar level qualification, and those wanting to take a specialist mathematics qualification will take GCE Mathematics, described above. Others will be required to continue with one of a range of qualifications under the umbrella heading of *Core Maths qualifications* designed to ensure that they maintain some fluency in the mathematical skills they

have already developed, but also develop their capacity to apply these skills in contexts commonly met in adult life. While there are a number of courses available with different specific content, they have more coherent pedagogical rationale for the statistics content in each of them than the GCE Mathematics course does.

DOES THE DIRECTION OF TRAVEL OF THE UK CURRICULUM MATCH DEVELOPMENTS ELSEWHERE?
*Does it match developments in the wider statistical community?*

Hypothesis testing is at its most effective when sample sizes are moderate: if they are too small then the power of the test is so weak that they are ineffective in identifying shifts in a parameter value; if they are too large then the power is so great that almost any difference in parameter value will show up as significant, irrespective of how important the shift is – effect size is a much more than significance for making decisions. Moderate size data sets are becoming much less frequent with data logging facilities improving and issues regarding storage and retrieval becoming much less frequent. More importantly however, because of the widespread abuse of hypothesis testing by the research community in some disciplines, the procedure is being increasingly questioned as a valid instrument.

Gorard (2014) offers many valid criticisms of the way the technique is abused within educational research, and one can have a great deal of sympathy with the observations that he makes about the cavalier approach to the use of significant tests (and CIs etc.) in contexts where the requirements for their use are not met – almost always the requirements are not exactly met and in many cases not even plausibly closely met. One can also have a great deal of sympathy with the observations he makes about the misstatement of conclusions – there is a widespread sloppiness about it which encourages fallacious misinterpretation of probabilistic reasoning. However, he undermines the valid arguments he makes by overstating the case against the use of the technique by misstating aspects of the technique. Absolutely critical to the logic of hypothesis testing, and missing from his portrayal of it, is that you do not calculate the probability of that observation – you calculate the probability of seeing any observation which is at least as 'unusual' / rare / ..... in the distribution of observational outcomes if one assumes the hypothesis to be tested is true. In his article Gorard writes:

> *.... But the significance test is conducted post hoc. There is no way of telling what the remaining population is from the sample alone. To imagine otherwise, would be equivalent to deciding that rolling a 3 followed by a 4 with a die showed that the die was biased (since the probability of that result is only 1/36, which is much less than 5% of course)*

There is no 'distribution of outcomes' which allows a construction of a critical region comprising the 'most unusual 5% of observations' in this scenario. There are 36 equally likely outcomes under the null hypothesis (fair die) – and no alternative hypothesis that allows you to identify one of them which would suggest bias: contrary to what he says, the construction of a hypothesis test (the decision rule) is not done post hoc but should be done before the data is considered. It is unfortunate that the powerful arguments against the abuse of the techniques in educational research are diminished by the inclusion of this additional material.

Trafimow & Marks (2015) have decided that the journal for which they act as editors will no longer accept articles using the null hypothesis significance testing procedure, on the basis that the procedure is invalid. Statslife (2015) (available at http://www.statslife.org.uk/news/2116-academic-journal-bans-p-value-significance-test ) reports on the editorial and invited comments from 5 respected figures in the statistical community (Peter Diggle, Stephen Senn, Andrew Gelman, Geoff Cummings and Robert Grant) which can be seen via a link in the report. Diggle, Senn and Grant all feel that a blanket ban goes too far and does not really address the underlying problem; Gelman and Cummings offer stronger support for the idea of a ban while observing that the real issue is that $p$ values given as the culmination of a research process suggest certainty, whereas confidence intervals provide a more appropriate window on the uncertainty associated with the results. Cumming refers to the 'dance of the $p$ values'. He has written a simulation which plots $p$ values of the difference between the means of samples drawn repeatedly from 2 well-defined populations. They show huge variability. The corresponding confidence intervals also

bounce but (of course) usually capture the difference in the population means. The confidence intervals give some insights into the p value bounce, while a single *p* value does not.

Given the lack of agreement amongst the wider statistical community about the validity and value of the hypothesis test, and the rapidly declining rate of occurrence of moderate sized data sets for which statistical hypothesis is particularly suited (if it is valid) it is reasonable to ask whether the UK should commit its best mathematical students to the study of it for a substantial period into the future.

*Does it match developments in the curriculae in other countries?*

Reviews by ACME (2011 a, b), Hodgen, Pepper, Sturman and Ruddock (2010), Vorderman, Budd, Dunne, Hart and Porkess (2011)) identify the UK as lagging well behind many other countries in terms of the number of 16 – 19 year old students who are studying some mathematics, and make recommendations that have been accepted by the current government that all students should study some mathematics, pre-university. The new 'Core Maths' qualifications target those mathematically competent students who have already achieved the grade C or above level at GCSE but who are not going to take GCE Mathematics at AS or A level. There are a number of different qualifications which will be available under this umbrella heading for schools and students to choose from, but all of them should mean that students in that target group maintain a level of fluency in mathematical skills through regular usage that previously they would not have encountered, and in most cases the courses will aim to develop some broader statistical literacy skills. These new courses and the requirement for all students to take some substantive quantitative course have the potential to improve the confidence and fluency in mathematics and statistics of our young people reaching the end of their school education.

The new curricula introduced in New Zealand (Ministry of Education, 2007) offer an encouraging example of multivariate data being a core part of the statistical reasoning from an early age. There, young children do not find it strange that boys and girls behave differently, that children of different ages behave differently, that things change over time etc. In this curriculum, children look at data presented in an accessible form and in a context for which they have the language that allows them to be able to talk about the data sensibly; consequently they are able to make sense of the stories in the data. A similar change has been introduced in South Africa when they revised their curriculum in 2005, but the context of implementing change in South Africa is much less easy to draw comparisons with the UK, because of the difficulties posed by the legacy of the apartheid regime on the capacity of native teachers to deliver a curriculum radically different from that which they experienced in their own education.

DISCUSSION

The various interpretations and misinterpretations of the nature of hypothesis testing, as demonstrated by the range of views referenced, is not enough in itself to drop it from the curriculum, or stop using it. It would be perverse that, just because many people get it mixed up or find it difficult, we decide to stop teaching it. Rather, the case is that there is less need for it, and as a statistical model, it has had its day. One could perhaps draw an analogy with the slide rule which was a very powerful tool at one stage but which is now obsolete.

However, there are much more powerful arguments against the inclusion of hypothesis testing in the new GCE curriculum. There are some serious weaknesses in the structure of the course: conditional probability underpins the theoretical framework of hypothesis testing; testing for the mean of the Normal distribution requires the sampling distribution of the statistic; testing for the value of the correlation coefficient must include some understanding of how this statistic is calculated and of the role that the sample size plays in the sampling distribution of the statistic. The omission of the logical foundations of hypothesis testing is not appropriate in a mathematical statistics course.

Moreover, the UK has had a shortage of qualified teachers in mathematics for some considerable time. Currently schools can choose the applied modules to be taught in GCE Mathematics at AS and A level, and only some teachers are comfortable teaching the full range of areas – pure, mechanics, statistics and decision mathematics. A completely defined GCE course which includes pure, mechanics and statistics is likely to place further strain on already stretched

resources, and it remains to be seen how well teachers with no experience of teaching statistics will deliver the compulsory statistics content.

Pullinger (2013) in his presidential address to the Royal Statistical Society argues that statistics are essential to good decision making, providing a basis on which to make tough decisions. He talks at length about the important role of statistics in helping us make sense of the world in which we live, and the last section before his call to action is the importance of statistical education and literacy. There appears to be little in the new GCE Mathematics course which will help further the goal of making sense of the world in which the students currently live, never mind the world in which they might expect to live for another 80 years with the attendant increases in data sophistication and data volumes. Indeed, the likely negative impact on statistical literacy and increased negative public perceptions of the value of statistics in the curriculum are predictable consequences of the proposed changes.

The example of New Zealand shows that it is possible to build an integrated course of mathematics and statistics running through the full school curriculum which introduces students naturally to the multivariate data which is pervasive in the world now, and reasoning with data is central to their development through the school curriculum. South Africa have followed a similar route in modernising their school curriculum opting to move away from a narrow focus on traditional statistical techniques towards a curriculum which encourages reasoning with data.

Regrettably, in the courses developed for the best mathematical students, the UK seem to have chosen to go towards mathematical statistics rather than reasoning with data, but in failing to provide the logical foundations of conditional probability and sampling distributions, they are failing even to provide a coherent course in mathematical statistics, and have developed something with profound pedagogical weaknesses. The Core Maths courses being developed for students not moving on to study the specialist mathematics qualification seem to offer a much more coherent and pedagogically sound framework, and is perhaps where other countries considering curriculum development might look rather than the GCE Mathematics course.

REFERENCES

ACME (2011a) *Mathematical needs: Mathematics in the workplace and in Higher Education.* http://www.acmeuk.org/media/7624/acme_theme_a_final%20%282%29.pdf

ACME (2011b) *Mathematical needs: The Mathematical needs of learners* http://www.acme-uk.org/media/7627/acme_theme_b_final.pdf

Department for Education (2014) GCE AS and A level subject content for mathematics. REF: DFE-00706-2014 downloaded 14 March 2015 from https://www.gov.uk/government/publications/gce-as-and-a-level-further-mathematics

Gorard, S. (2014) The widespread abuse of statistics by researchers : what is the problem and what is the ethical way forward?, *Psychology of education review.*, 38 (1). pp. 3-10.

Hodgen, J., Pepper, D., Sturman, L., Ruddock, G.(Nuffield Foundation) (2010). *Is the UK an outlier? An international comparison of upper secondary mathematics education.* www.nuffieldfoundation.org/uk-outlier-upper-secondarymaths-education

Ministry of Education (2007). *The New Zealand curriculum*. Wellington, New Zealand: Learning Media Limited.

Nicholson, J.R., Ridgway, J. and McCusker, S. (2006). Reasoning with data – time for a rethink? *Teaching Statistics  28 (1),* 2–9.

Porkess, R. (2012).*The Future of Statistics in our School and Colleges*. The Royal Statistical Society and the Actuarial Profession, London. ISBN 978-0-9571305-0-0.

Pullinger, J. (2013) Statistics making an impact. *J. R. Statistic. Soc. A, 176 (4)* 819 – 836.

Ridgway, J., Nicholson, J, and McCusker, S. (2007) Teaching statistics - despite its applications. *Teaching Statistics, 29(2),* 44-48.

Trafimow, D. & Marks, M. (2015) Editorial *Basic and Applied Social Psychology* 37, 1 – 2.

Vorderman, C., Budd, C., Dunne, R., Hart, M., and Porkess, R. (2011) *A world-class mathematics education for all our young people*. Downloaded from http://www.tsm-resources.com/pdf/VordermanMathsReport.pdf