Fellowship Report: December 2013

Dr Matthew Inglis, Loughborough University

1 Summary

As discussed in earlier reports, the overall goal of the Fellowship is to determine the extent to which studying mathematics develops general reasoning skills. Following the award of additional funding from various other sources I have been able to appoint additional research assistants, which has allowed me to expand the scope of work substantially. Here I briefly summarise progress to far.

- My main empirical work from Years 1-3 was published in July in *PLOS ONE*. This is available online at `http://dx.doi.org/10.1371/journal.pone.0069399`.

- Two follow-up studies, focused on the development of undergraduate reasoning and reasoning development in the Cypriot curriculum are now complete, and have findings consistent with those published in the *PLOS ONE* paper. I give details of these studies below.

- A strand of work which focuses on how reasoning skills can be improved in the mathematics curriculum—using the Comparative Judgement method (CJ)—has expanded in scope substantially this year. Ian Jones and I were awarded £11,409 from AQA (the examination board) to apply the method to study how standards of A Level Mathematics examinations have changed over time. I report the findings of this work below. In addition, the Nuffield Foundation awarded Ian Jones, Camilla Gilmore and I a large grant (£132,069) to develop the general method further. This work commenced in October and will run for the remaining two years of my Fellowship.

- Following Nina Attridge's successful PhD completion this year, I obtained £55,000 from Loughborough University to recruit a replacement. Sara Humphries, a mathematics and psychology graduate, began working with me this year and is currently focusing on the relationship between post-compulsory study and matrix reasoning skills.

Overall, I am very happy with the progress of the Fellowship. Once again, I am extremely grateful to the Worshipful Company of Actuaries for its continued generous support. In the sections below I report on the various strands of the Fellowship. Note that I am assuming that readers are familiar with the contents of earlier reports.

2 Longitudinal Reasoning Studies

2.1 PLOS ONE Paper

Since my last report, the main empirical work I have been conducting in Years 1-3 has been published in the high impact journal *PLOS ONE*. The article is freely available online at `http://dx.doi.org/10.1371/journal.pone.0069399`. To date, the article has been viewed over 1000 times, which is roughly twice the average figure for psychology articles in the first six months following publication.

Liverymen will recall that this paper reports a relationship between the study of A Level mathematics and an improvement in conditional reasoning skills. I have presented this work on several occasions, at the University of Oxford, the University of Nottingham, the Technische Universität München, the Advisory Committee on Mathematics Education, and at a SIAS Lecture at Staple Inn.

2.2 Cypriot Study

One important policy issue which is currently being discussed is whether all students should study mathematics until the age of 18. Both major political parties are now committed to this proposal. In view of these discussions I was keen to investigate whether the association with reasoning development and the study of A Level mathematics would also be present for a programme of study which contains less mathematical content. To this end I set up a replication of the *PLOS ONE* study in Cyprus. The Cypriot context is interesting for at least two reasons. First, all Cypriots are compelled to study mathematics until the age of 18, but they must choose between a 'high intensity' and a 'low intensity' course. The high intensity course is similar to A Level mathematics, although contains marginally less content. The low intensity course is similar to what has been proposed for students in England and Wales, should mathematics become compulsory to 18. Second, the Cypriot curriculum contains substantially more deductive geometry than the English curriculum, which the kind of content most often associated with logical reasoning by proponents of the Theory of Formal Discipline.

I recruited 150 Cypriot students to take part in the study, which ran over two years. They were asked to complete a translated version of the same instrument used in the *PLOS ONE* study, which measured students' conditional reasoning behaviour. The results are shown in Figure 1.

There are several remarks worth making about these results. First, the students who opted to study high intensity mathematics started out with a similar profile of

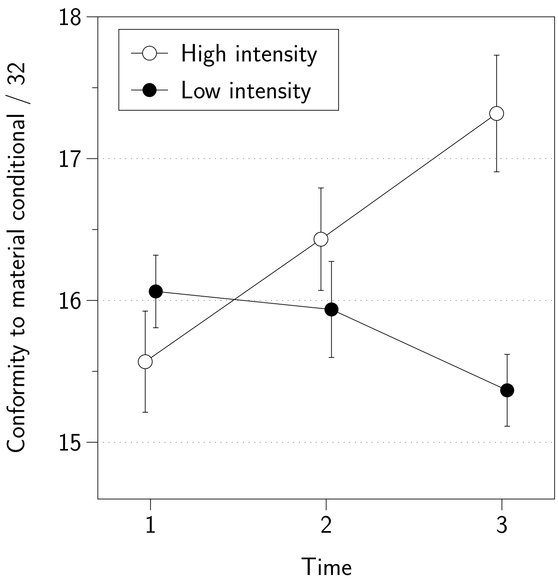


Figure 1: Reasoning development in the Cypriot study. Error bars show ±1 SE of the mean.

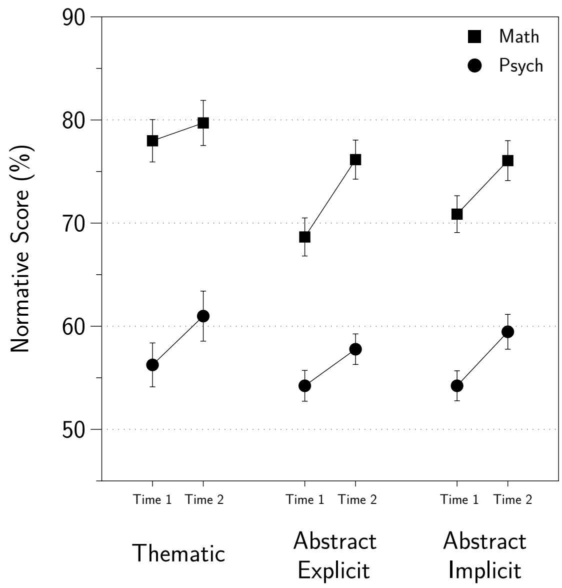


Figure 2: Reasoning development among undergraduate mathematics and psychology students on three different measures of conditional inference.

conditional reasoning behaviour to those who chose to study low intensity mathematics. Second, they improved their scores substantially over the two year period. Finally, the low intensity students' behaviour did not significantly change over the two years. The time by group interaction effect (the key test of whether the two groups showed different developmental patterns) was highly significant, $p = .008$ with a reasonable effect size, $\eta_p^2 = .045$.

I think these results are useful for two reasons. First, they replicate the key findings from the *PLOS ONE* study, bringing some degree of reassurance that they were not a false positive. Second, they suggest that studying a small amount of mathematics is unlikely to develop reasoning skills in the same way that studying a large amount of mathematics seems to. This is important, because the current drive to require all students to study (a small amount of) mathematics until 18 is based, in part at least, upon the belief that this will develop their 'thinking skills'. It may be that this view is misplaced, and that significant exposure to mathematical ideas is needed if this development is to take place. Of course, there are other legitimate rationales for advocating compulsory mathematics until 18. I believe that this is a sensible policy, but it seems that it may not necessarily increase 'thinking skills' (at least, 'thinking skills' as indexed by the conditional inference task) in quite the way that one might have hoped.

2.3 Undergraduate Study

As discussed in my last report, I also been investigating whether the *PLOS ONE* findings generalise to undergraduate education. The results of this study are shown in Figure 2. Here the comparison group were psychology students rather than English Literature students (frustratingly, the Department of English at Loughborough declined to participate in the study). Participants were given a slightly different measure of conditional inference, which permitted splitting scores up into three different types (I won't go into details, as behaviour was similar on each).

There are several findings worth highlighting:

- The mathematics students performed better than the psychology students on the conditional inference task on entry to university (and this difference remained significant when controlling for differences in intelligence and thinking disposition).
- Both groups improved significantly across their year of study.
- Both groups improved to a similar extent on all three types of conditional inference problem.

These results are interesting, as they suggest that (a) reasoning develop continues from A Level into undergraduate study, but also that (b) mathematics may not be the only subject which develops such skills, as claimed by some proponents of the Theory of Formal Discipline.

3 Comparative Judgement

As noted in my last report, Ian Jones and I have been working on developing a novel method of assessing mathematics. The Comparative Judgement (CJ) approach allows for the assessment of complex conceptual ideas by presenting pairs of scripts to experts and simply asking them which is the better.

This year we have applied this method in several ways. First, we completed a project designed to test whether this approach could be used to assess GCSE mathematics. We have written up this as a manuscript which is currently under review, and a discussion of this approach was published in *The Actuary* magazine. We have been developing this approach in two main directions.

3.1 A Level Standards Over Time

In a collaboration with OfQual (the examinations regulator) and AQA (one of the three main English examination boards), we have been using CJ to assess changes to A Level standards over time. In recent years there has been concern about a 'dumbing down' of standards, but this is hard to assess accurately. Although the pass rate in A Level mathematics has increased substantially, this could be down to better quality teaching, better prepared candidates (some argue that, because of improvements in health care and nutrition, we should expect educational achievement to improve over time) or lower standards.

Existing approaches to comparing examination standards over time typically involve comparing performances with a standardised test. For example, Coventry University has asked every incoming mathematics undergraduate to take the same diagnostic test since 1991. When analysing these data, Professor Duncan Lawson suggested that a 1991 grade N student exhibited the same levels of competence as a 1997 grade C student or a 2001 grade B student. However, such analyses can be criticised because the referent test did not take account of changes in the A Level syllabus.

Ian Jones and I were awarded £11,409 from AQA to apply the Comparative Judgement technique discussed in my last report and in my talk at Common Hall 2013 to this problem. Because our approach relies upon expert holistic judgements about a candidates overall mathematical competence, rather than assessing their knowledge of specific mathematical topics, we believe that the method is less vulnerable to the criticism which can be levelled at the Coventry findings.

We began by obtaining historical examination papers from the OfQual archive. Unfortunately, this archive is extremely patchy: remarkably no examination scripts from the 1970s and 1980s have survived various government reorganisations. Original scripts from the 1960s do exist, but only a very small number which were awarded a restricted range of grades (A, B and E). We ended up with scripts from 66 candidates, each of whom was awarded an A, B or E grade, in 1964, 1968, 1996 or 2012.

In order to prevent judges becoming aware of our research hypothesis by detecting historic paper or handwriting, we hired a mathematics student to rewrite all answers in neutral handwriting. Example scripts are given in Figures 3 and 4.

We then asked judges (mathematics PhD students) to assess responses to each question using the CJ approach. Each judge was presented with two responses to two (different) questions and asked simply to indicate which candidate they felt had demonstrated the better mathematical understanding. Judges were not told that they were looking at scripts from different years, only that the project was designed to compare the standards of different mathematics syllabi. After all judgements had been completed judges were asked to complete a questionnaire with a number of free text responses. Critically, they were asked to speculate on the purpose of the study. None hypothesised that we were interested in standards over time. This is important because we did not want media perceptions about falling standards to influence judges' behaviour.

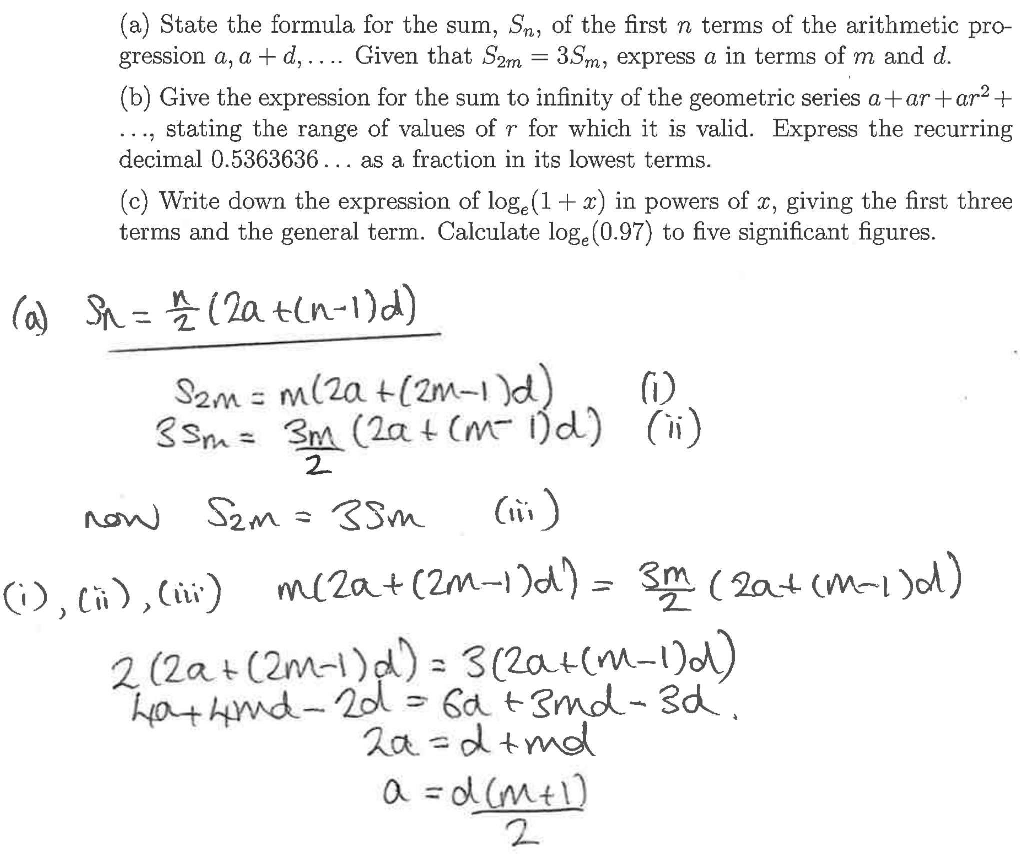
The CJ procedure results in a standardised parameter for each question, which we used to create mean scores for each candidate (this was necessary because the examination papers had different numbers of questions). Finally, we used these scores as the dependent measure in a regression involving two predictors: year and grade awarded.

The results are shown in Figure 5. According to our analysis, standards have fallen over time, in the sense that our judges rated B-grade candidates from 2012 as having similar levels of mathematical understanding (or, more precisely, had *demonstrated* similar levels of mathematical understanding) to E-grade candidates from the 1960s. As you would expect from looking at the graph, in the regression analysis there was a significant effect of year, $B = -.025, p < .001$, meaning that for every year that has passed since the 1960s, a candidate with the same grade has had (demonstrated) mathematical understanding approximately 0.025 standard deviations lower than an equivalent candidate from the year before (about 1 standard deviation over a 40 year period).

So our findings suggest that, as Lawson suggested, A Level standards have declined somewhat over time. However, we found evidence for a much slower rate of decline than earlier researchers (recall that Lawson suggested that a 1991 N grade student was equivalent to a 2001 B grade student). It is also possible that this decline has been stabilised: our mean parameters for A and E grade candidates in 2012 were higher than their 1996 counterparts. Of course, the evidence for our conclusions would be much more compelling if we had been able to include examination scripts from the 'lost decades' in the 70s and 80s. We understand that OfQual now has a considerably more systematic archive policy, and this will aid similar studies in the future.[1]

This work is currently being written up for publication and has not yet been peer reviewed.

[1] Although, bizarrely, it transpires that storing examination papers is not straightforward because apparently the student holds the copyright to the script, not the examination board. Because of this, it took a non-trivial amount of negotiation before OfQual were comfortable with our strategy of copying out the answers to the scripts.


(a) State the formula for the sum, S_n, of the first n terms of the arithmetic progression $a, a+d, \ldots$. Given that $S_{2m} = 3S_m$, express a in terms of m and d.

(b) Give the expression for the sum to infinity of the geometric series $a + ar + ar^2 + \ldots$, stating the range of values of r for which it is valid. Express the recurring decimal 0.5363636... as a fraction in its lowest terms.

(c) Write down the expression of $\log_e(1+x)$ in powers of x, giving the first three terms and the general term. Calculate $\log_e(0.97)$ to five significant figures.

(a) $S_n = \frac{n}{2}(2a+(n-1)d)$

$S_{2m} = m(2a+(2m-1)d)$ (i)

$3S_m = \frac{3m}{2}(2a+(m-1)d)$ (ii)

now $S_{2m} = 3S_m$ (iii)

(i), (ii), (iii) $m(2a+(2m-1)d) = \frac{3m}{2}(2a+(m-1)d)$

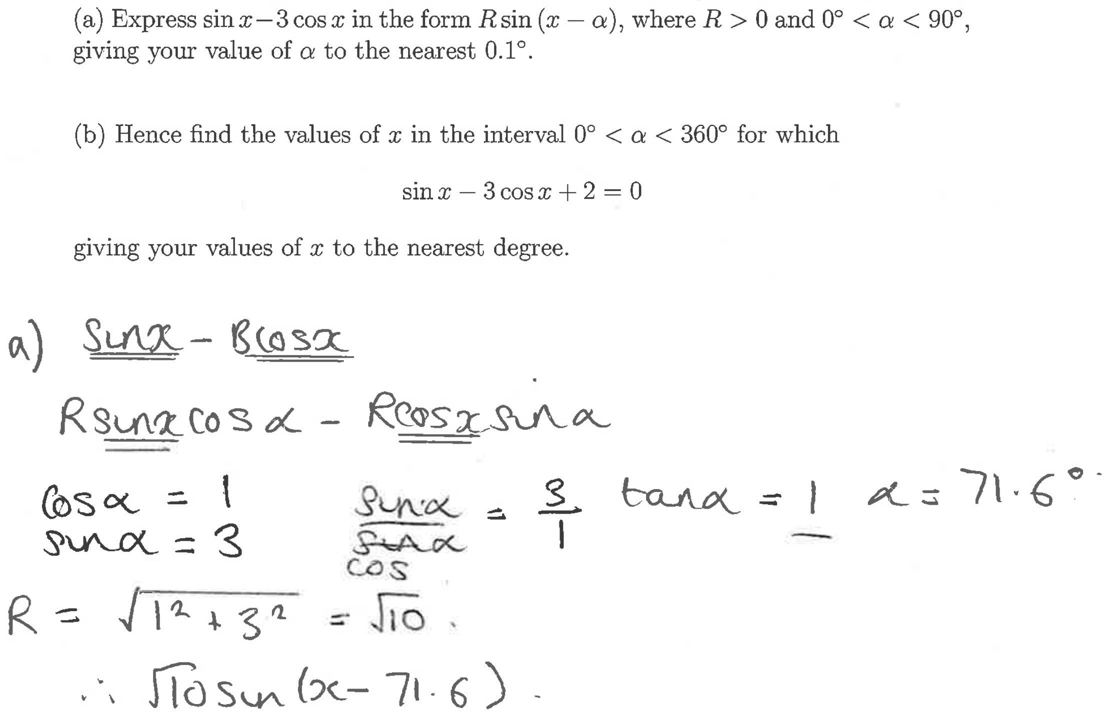
$2(2a+(2m-1)d) = 3(2a+(m-1)d)$

$4a+4md-2d = 6a+3md-3d$.

$2a = d+md$

$a = \frac{d(m+1)}{2}$

Figure 3: A response to a 1968 examination from an A grade candidate.


(a) Express $\sin x - 3\cos x$ in the form $R\sin(x-\alpha)$, where $R > 0$ and $0° < \alpha < 90°$, giving your value of α to the nearest 0.1°.

(b) Hence find the values of x in the interval $0° < \alpha < 360°$ for which

$$\sin x - 3\cos x + 2 = 0$$

giving your values of x to the nearest degree.

a) $\sin x - 3\cos x$

$R\sin x\cos\alpha - R\cos x\sin\alpha$

$\cos\alpha = 1$ $\quad \frac{\sin\alpha}{\cos\alpha} = \frac{3}{1}$ $\quad \tan\alpha = 1$ $\quad \alpha = 71.6°$

$\sin\alpha = 3$

$R = \sqrt{1^2+3^2} = \sqrt{10}$.

$\therefore \sqrt{10}\sin(x-71.6)$.

Figure 4: A response to a 2012 examination from an A grade candidate (crossing out was in the original).

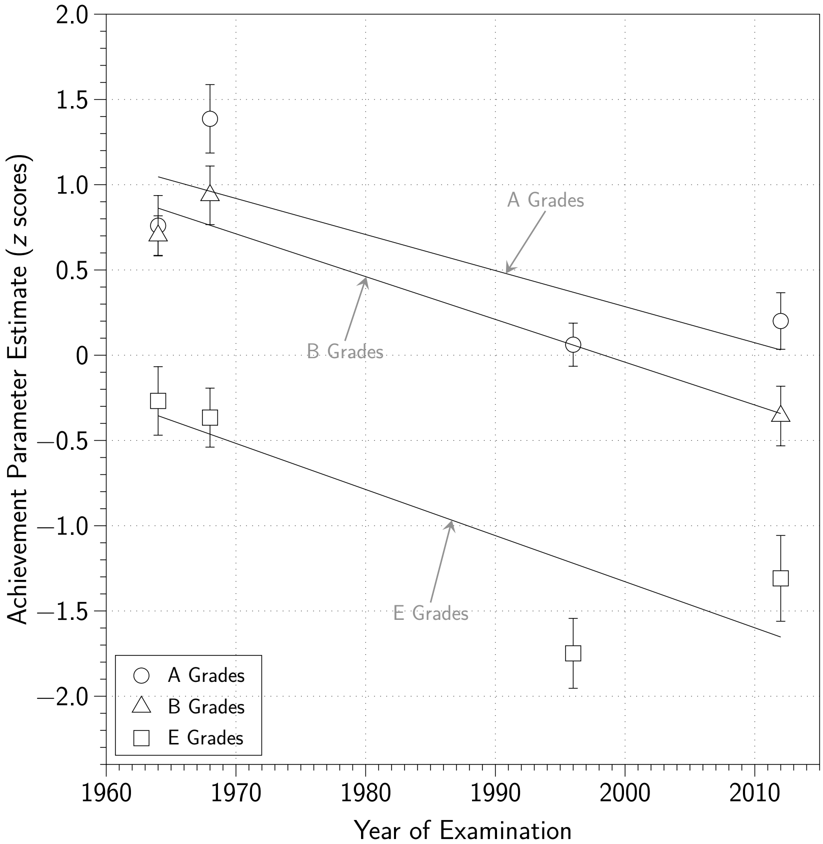


Figure 5: Standards over time in A Level mathematics examinations: an application of the Comparative Judgement method. Error bars show ± 1 SE of the mean.

3.2 The Measuring Conceptual Understanding Project

Together with Ian Jones and Camilla Gilmore (both colleagues at Loughborough who also have Royal Society Fellowships), I was awarded a £132,069 grant from the Nuffield Foundation to develop our work on comparative judgement as a technique for measuring conceptual understanding.

A major challenge for mathematics education research is how to measure pupils' conceptual understanding with acceptable validity and reliability. As an example of this problem, opinion is often divided on whether abstract or concrete representations[2] are better at supporting pupils' conceptual understanding, but to date no researchers have successfully developed a method of reliably measuring pupils' outcomes, so the issue is badly under-researched.

This study will develop a measure of conceptual understanding using the Comparative Judgement (CJ) approach, and demonstrate the application of CJ to this debate. It builds on our previous work and a recent pilot study demonstrating how CJ could be used to measure learners' understanding of mathematical concepts (discussed in my June 2013 Report).

The research will involve two teaching experiments at secondary level (Year 11), and at primary level (Year 5). In each experiment students will be randomly allocated into two groups and taught an unfamiliar mathematical concept. One group will be taught using abstract representations, the other using concrete representations. The researchers will then use CJ to establish whether the abstract or concrete representations led to a more sophisticated mathematical understanding of the target concept in each experiment.

The project aims to provide a major methodological contribution to transform how educational interventions are evaluated. It also hopes to provide robust evidence to inform the abstract vs. concrete debate, and influence how mathematics lessons are designed and taught.

The Nuffield Foundation funding, which runs for two years, is being used to employ a postdoctoral research associate to manage the project, and two teachers to develop and implement the teaching interventions. We have now appointed Dr. Marie-Josée Bisson, a recent PhD graduate from the University of Nottingham. She took up the post in October and has made great progress on the project so far. Full details of the project can be found at `http://mec.lboro.ac.uk/mcu/`.

[2] For example, when introducing differentiation some argue that it is preferable to link the ideas to students' intuitive understanding of real world concepts such as velocity and acceleration. Others argue that this concrete context hinders students' learning by focusing their attention on details which are irrelevant to the mathematics being studied. Instead they propose it is better to introduce students to such ideas using entirely abstract representations (i.e. find the tangent to a curve at a given point).

4 Plans for the Year Ahead

I have several major goals for the year ahead:

- Between February and April I will be visiting the Graduate School of Education and the Department of Mathematics at Rutgers University in New Jersey. My colleagues Keith Weber and Juan Pablo Mejía-Ramos are working on similar issues there, and the visit will provide an opportunity to exchange ideas and collaborate on future studies.

- My PhD student Sara Humphries and I have begun a large scale study investigating the relationship between A Level study choices and reasoning development on a different measure of 'general thinking skills': matrix reasoning. On the face of it, this measure is less clearly related to mathematical study, but has a much greater relationship with real world outcomes than conditional inference scores. We have collected data from approximately 900 sixth formers so far, and will return at the end of their first year of study to determine their development over the year.

- I would like to conduct a large scale cross-sectional study looking at the differences between mathematicians and non-mathematicians on a wide variety of reasoning tasks. Although I now have good evidence for developmental effects on certain tasks, I do not have a good sense of the generality of these findings (in terms of reasoning types). My current thinking is that a web-based study will allow for a sufficiently large sample to test cross-sectional differences on a variety of tasks.

- A further goal for the year is to continue to disseminate recent findings, and explore with publishers the possibility of writing a book length summary of the work.

5 Dissemination

Over the past six months I have spoken at several conferences, and published a number of papers. Specifically:

Papers/Reports:

- Inglis, M. & Gilmore, C. (in press). Indexing the Approximate Number System. *Acta Psychologica*.

- Alcock, L., Attridge, N., Kenny, S. & Inglis, M. (in press). Achievement and Behaviour in Undergraduate Mathematics: Personality is a Better Predictor than Gender. *Research in Mathematics Education*.

- Weber, K., Inglis, M. & Mejía-Ramos, J. P. (in press). How mathematicians obtain conviction: Implications for mathematics instruction and research on epistemic cognition. *Educational Psychologist*.

- Duah, F., Croft, T., & Inglis, M. (in press). Can peer assisted learning be effective in undergraduate mathematics? *International Journal of Mathematical Education in Science and Technology*.
- Gilmore, C., Attridge, N., De Smedt, B., & Inglis, M. (2014). Measuring the Approximate Number System in children: Exploring the relationships among different tasks. *Learning and Individual Differences, 29,* 50-58.
- Hodds, M., Alcock, L., & Inglis, M. (2014). Self-explanation training improves proof comprehension. *Journal for Research in Mathematics Education, 45*, 98-137.
- Attridge, N. & Inglis, M. (2013). Advanced mathematical study and the development of conditional reasoning skills. *PLOS ONE, 8*, e69399.
- Inglis, M. & Gilmore, C. (2013). Sampling from the mental number line: How are approximate number system representations formed? *Cognition, 129,* 63-69.
- Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., Simms, V., & Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLOS ONE, 8*, e67374.
- Jones, I., Inglis, M., Gilmore, C, & Hodgen, J. (2013). Measuring conceptual understanding: The case of fractions. In A. M. Lindmeier & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 113-120). Kiel, Germany.
- Inglis, M., & Mejia-Ramos, J. P. (2013). How persuaded are you? A typology of responses. In A. Aberdein & I. Dove (Eds.), *The Argument of Mathematics*, (pp. 101-118). Springer: Dordrecht.

Talks: This year I have given talks about my work at the University of Oxford, University of Cambridge, Christian-Albrechts-Universität zu Kiel, University of Nottingham, Sheffield Hallam University, Staple Inn Actuarial Society, London Mathematical Society, and the Technische Universität München.

Copies of published papers are available from my website: `http://www-staff.lboro.ac.uk/~mamji`

Matthew Inglis
10th December 2013