SAFETY IN NUMBERS? NUMBER CRUNCHING AND MAKING SENSE

Tim Rowland

University of Cambridge

Quantitative and qualitative research methods offer us different kinds of insights. Whilst statistical analyses might tell us 'that', only by getting 'inside' the data can we begin to understand 'why'. In the session at the BSRLM conference I presented some 'thats' which had emerged from statistical analysis of the relationship between subject knowledge variables and teaching performance variables, in the context of primary initial teacher education. One purpose was to invite conjectures about why these relationships might hold (or not).

The purpose of this session was to offer some statistical findings from an ongoing research project, to offer possible interpretations for comment, and to invite additional and/or alternative interpretations. By 'interpretation', I mean some kind of unpacking of the context in which the data were collected so as to try to make sense of distributions, correlations and the like. Of course, context is multi-layered and multi-dimensional. In this case, it includes the mathematics component of primary initial teacher education and the observation of these students ('trainees') as they taught mathematics on their school-based placements. Participants were encouraged to draw on their own experience to inform their reaction to the data and the analysis presented. No attempt is made in this account to attribute hunches, impressions or information input to particular individuals. *All such contributions are shown in indented text later in this paper*.

BACKGROUND

Circular 4/98 (DfEE, 1998) sets out what is considered to be the "knowledge and understanding of mathematics that trainees need in order to underpin effective teaching of mathematics at primary [elementary] level", and charges Initial Teacher Training ITT 'providers' with the audit and remediation of students' mathematics subject knowledge. This paper draws on data from the 1998-99 audit with one cohort of students following the one-year PGCE. A 90-minute written assessment consisting of 16 test items in mathematics was administered some four months into the course.

The course includes two extended placements in schools in the latter parts of the second and third terms. During these placements, each student works under the joint supervision of a school-based mentor and a university tutor. The two supervisors agreed on assessments of the student's performance in teaching mathematics towards the end of each placement, against the standards of Circular 4/98. More comprehensive accounts of the goals and methods of the research are given in Rowland, Barber, Martyn and Heal (2000, 2001).

SUBJECT KNOWLEDGE AND CLASSROOM PERFORMANCE

The level of each student's subject knowledge (based on the audit) was categorised as low, medium or high, corresponding to the need for significant remedial support, modest support (or self-remediation), or none. In addition, the assessments of the students' teaching of mathematics were made on a three-point scale weak/capable/strong. These assessments were made (a) on both first and second placements, and (b) with respect to both 'pre-active' (related to planning and self-evaluation) and 'interactive' (related to the management of the lesson in progress) aspects of mathematics teaching (following Bennett and Turner-Bisset, 1993). Tables 1 to 4 below show the four 3 by 3 contingency tables, for Placement 1 (N=167) and Placement 2 (N=164), together with expected frequencies (in parentheses) based on the null hypothesis that audit performance and teaching performance are independent.

		TEACHING PRACTICE PERFORMANCE		
SUBJECT KNOWLEDGE AUDIT		Strong	Capable	Weak
	High	**17 (12.4)**	**16 (17.3)**	**1 (4.3)**
	Middle	**31 (29.2)**	**38 (40.7)**	**11 (10.1)**
	Low	**13 (19.4)**	**31 (27.0)**	**9 (6.7)**

Table 1: Placement 1, pre-active

		TEACHING PRACTICE PERFORMANCE		
SUBJECT KNOWLEDGE AUDIT		1 (strong)	2 (capable)	3 (weak)
	A (high)	**18 (12.4)**	**12 (16.1)**	**4 (5.5)**
	B (middle)	**32 (29.2)**	**37 (37.8)**	**11 (12.9)**
	C (low)	**11 (19.4)**	**30 (25.1)**	**12 (8.6)**

Table 2: Placement 1, interactive

		TEACHING PRACTICE PERFORMANCE		
SUBJECT KNOWLEDGE AUDIT		Strong	Capable	Weak
	High	**12 (8.1)**	**18 (14.1)**	**4 (11.8)**
	Middle	**20 (18.5)**	**33 (32.3)**	**25 (27.1)**
	Low	**7 (12.4)**	**17 (21.6)**	**28 (18.1)**

Table 3: Placement 2, pre-active

		TEACHING PRACTICE PERFORMANCE		
SUBJECT KNOWLEDGE AUDIT		1 (strong)	2 (capable)	3 (weak)
	A (high)	**13 (8.5)**	**19 (18.2)**	**2 (7.3)**
	B (middle)	**21 (19.5)**	**42 (41.9)**	**15 (16.6)**
	C (low)	**7 (13.0)**	**27 (27.9)**	**18 (11.1)**

Table 4: Placement 2, interactive

The association between audit score and teaching performance is significant ($p<0.05$) for three of the four analyses, the exception being the pre-active dimension of the first placement.

Taken together, these results support our earlier findings with the 1997-98 cohort (Rowland *et al*., 2000) and point to the positive effect of strong subject knowledge in both the planning and the 'delivery' of elementary mathematics teaching.

Why should pre-active/placement 1 be an exception? For the moment, we conjecture that the assessment of pre-active aspects favours not only clarity about mathematics teaching and learning, but also a certain kind of bureaucratic competence that acts as some kind of 'leveller' in the very first exposure to work in schools.

The students seem to be 'worse' on the second placement than the first. For example, the number judged as 'strong' falls from about 60 to about 40. It seems improbable that their performance deteriorates over the year, and a more likely explanation might be the rising expectations of

> tutors and mentors. Thus, whilst assessments are supposedly made against 'objective' criteria (the Circular 4/98 standards), in practice some prediction of potential seems to come into play in the assessment of the first placement. Arguably this does not invalidate the goodness-of-fit analysis which depends only on the hypothesis that the same *proportions* of weak/capable/strong teaching will be found among the students with low/medium/high subject knowledge.
>
> The association between subject knowledge and teaching performance need not be causal. Maybe the students successful in both are just 'good' students. Perhaps there is a cycle in which commitment and motivation are, in turn, fuelled by success. A combination of good teaching practice experiences and success on the audit could have increased both, whilst, conversely, poor teaching practice experiences and weak results on the audit might be expected to have the opposite effect. However, a downward spiral due to a poor audit score does not resonate with our knowledge of these and other primary PGCE trainees. There is stiff competition to gain a place on the course; most students are resilient, well-motivated and goal-oriented.

INTRODUCING ADDITIONAL VARIABLES

A more recent, unpublished independent analysis of the 1998-99 data (Proctor, 2001) incorporates hitherto unexamined variables in an attempt to 'predict' the mathematics teaching competence of these trainees. These additional variables are:

1. The gender (male/female) of the trainee.
2. Their chosen age specialism (Early Years 3-8 or Middle Years 7-11)
3. A subject knowledge audit self-assessment. For each audit item, trainees were asked to indicate their degree of confidence on each item in advance of tackling it, on a 5-point scale from 0 ("Terrifying, I can't really think about it") to 4 ("I could show someone else how to do this").

Variables S, A and TP then denote respectively the sum of the *self-assessment scores* for the 16 items, the sum of the actual *audit scores* for the 16 items and the sum of the four *teaching practice grades* (pre-active and interactive for the two placements). S lies in the range 0 to 64, A in the range 0 to 32 and TP from 4 to 16. It should be noted that a low TP score is 'good'. Additional dichotomous variables G, Y denote respectively Gender (M=0, F=1) and Course/phase (MY=0, EY=1).

> How do we know that students made the confidence self-assessments *before* trying each question? (Answer: we don't). A self-assessment *after* answering a question might be very different. Which of the two is the more informative?
>
> Summation of the four teaching grades is expedient for this analysis, but what does it mean? Two students scoring 10 on this scale might look very different. On the other hand, the same is true of the S and A scores.

Proctor investigates whether TP can usefully be 'predicted' from some or all the variables S, A, G, Y, for the whole cohort and for various subgroups.

Her report begins with some exploratory data analysis. Calculation of correlation coefficients shows significant correlation between the three pairs of variables S, A, TP, although scatter graphs strongly suggest that neither A nor S alone is going to be a useful predictor of TP. That for A/TP is shown in Figure 1.

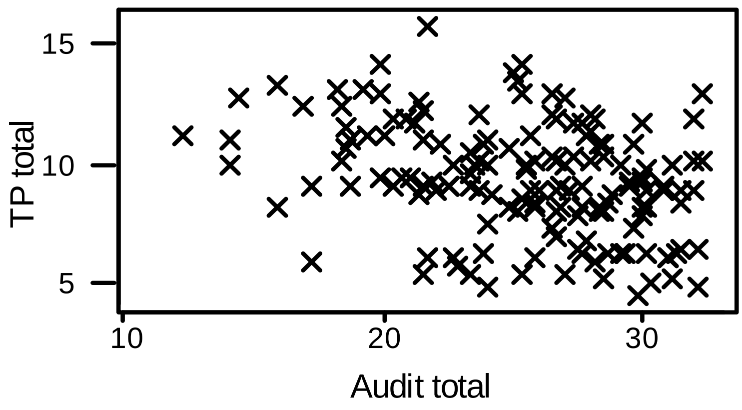

Figure1: Scattergraph showing audit total and TP total for all trainees.

> At least one person perceived a top-left to bottom-right trend in the forest of crosses! This was rather convincing because they had been expecting (if anything) a positive correlation until it was pointed out that a low TP score is 'good'.
>
> International studies (e.g. Foxman, 1992) indicate that confidence is a poor predictor of mathematical performance, although these compare school students in one country with those in another.

A box and whisker plot comparison of *audit scores* over gender (female F, male M) and course (early years EY, middle years MY) indicates that there is little difference between the four groups F/EY, F/MY, M/MY.[1]

> This seemed fairly unremarkable, with no a priori expectation that any of these three groups would have stonger/weaker subject knowledge relative to the others.

A comparison of *self-assessment scores* (Figure 2) for the same groups, however, suggests a higher level of confidence among the MY students irrespective of gender (omitting the M/EY pair).

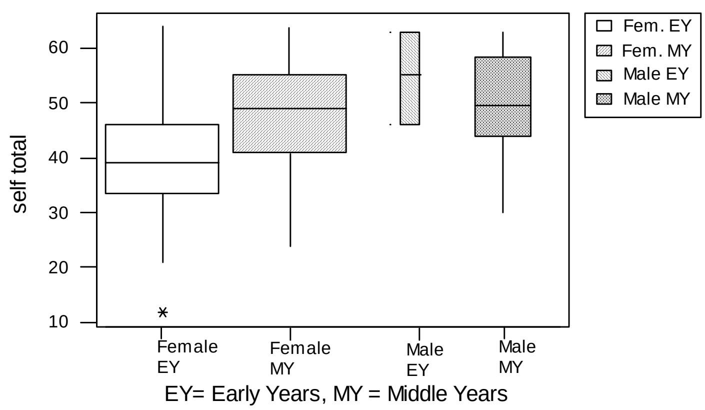

Figure 2: Comparison of self-audit totals by course/gender group

---

[1] The M/EY scores are high relative to the other groups, but the group is very small (2 individuals) and this group is excluded in subsequent analysis.

> This was not thought to be surprising. Mathematics confidence is perceived to be an issue with many Early Years trainees. Perhaps some trainees opt for EY partly because they doubt their ability to handle mathematics in KS2 (or Year 5/6). If this is the case, should it be a matter for concern? Arguably it should, given that the EY students' actual knowledge (as measured by the audit) was no different from that off the cohort as a whole.
>
> There is a suggestion that male students have an exaggerated confidence in their mathematical capability. How well will they respond to diagnostic feedback from the audit?

A similar comparison of *Teaching Practice scores* points to the superiority of F/MY as a group. The mean TP score of M/MY is the highest (i.e. the worst) and this group contains a long 'tail' of weak students as regards mathematics teaching performance.

> The drive to recruit men into primary school teaching for their qualities as role models for disaffected (and other) boys may have its downside. Complicating factors include the identity crisis experienced by some male primary trainees.

REGRESSION ANALYSIS

Proctor's study uses ordinary least squares (OLS) regression analysis in order to predict the TP score from A and S, together with the dichotomous variables G and Y. This results in the following model for the whole cohort:

$$TP = 16.7 - 0.166A - 0.0376S - 1.178G + 0.365Y$$

The overall significance of the model is high, and the greater contribution of the audit score A to this difference is much greater than that of the confidence score S However, the p-value of the course variable Y is found to be 0.35, indicating that there is little justification in including it in the model. The OLS model then becomes:

$$TP = 16.95 - 0.164A - 0.044S - 1.67G$$

> The contribution of the audit score is much greater than that of the self-audit.
>
> As a group, females are likely to perform slightly better than males.

An OLS model to predict the sum of the two scores for the second teaching placement only – the one that ultimately matters most in the PGCE course assessment – gives constants roughly half of those shown immediately above, as might be expected, although the contribution of the audit score and gender is even greater.

Perhaps the most startling result comes from a comparison of separate OLS models for the Early Years (excluding the two males) and Middle Years students.

For the Middle Years trainees: $TP = 16.8 - 0.233A - 0.0045S - 1.82G$

For the Early Years trainees: $TP = 14.5 - 0.0646A - 0.0818S$

For the MY group, the predictive effect of the self-assessment score is negligible compared with that of the audit score and the gender effect, whereas for the EY trainees confidence is the more powerful predictor. (This may not be self-evident since the coefficients of A, S in the EY model are in the ratio 3:4 approximately. However, it must be remembered that the range of values of S is twice that of A.

Again, the confidence of Early Years students seems to be significant. As a group, they show less confidence than the MY students. For these students, confidence is a relatively powerful predictor of classroom performance. Is it possible to engender greater confidence in these trainees, whose actual knowledge is much the same as their MY colleagues?

Finally, it is important to recognise the 'flatness' of these models model. For the whole cohort, for example, the difference between the highest and the lowest (in practice) audit scores accounts for less than one grade on each of the four teaching assessments. These models are statistically significant, but not 'substantive'.

So there is something else going on! The variables which this study has chosen to focus on are important, but only a part – perhaps a small part - of the total story.

Acknowledgements

Sincerest thanks to those who attended the Southampton session. I acknowledge that the space available here makes it possible to record only a small part of their contribution. I am also grateful to Elaine Proctor for permission to include figures from her dissertation. Patti Barber, Sarah Martyn and Caroline Heal were co-researchers in the project from which the data are taken.

REFERENCES

Bennett, N. and Turner-Bisset, R. (1993) 'Knowledge bases and teaching performance'. In Bennett, N. and Carré, C. (eds.) *Learning to Teach*, pp. 149-164. London: Routledge.

DfEE (1998) *Teaching: High Status, High Standards: Circular 4/98*. London: HMSO.

Foxman, D. (1992) Learning Mathematics and Science. The second International Assessment of Educational Progress in England. Slough: NFER.

Proctor, E. (2001) 'Relationships between mathematics knowledge, mathematics confidence and teaching ability'. Unpublished dissertation for the MSc in Applied Statistics. Sheffield: Sheffield Hallam University.

Rowland, T., Martyn, S., Barber, P. and Heal, C. (2000) 'Primary teacher trainees' mathematics subject knowledge and classroom performance'. In Rowland, T. and Morgan, C. (eds) *Research in Mathematics Education Volume 2: Papers of the British Society for Research into Learning Mathematics*, pp. 3-18. London: BSRLM..

Rowland, T., Martyn, S., Barber, P. and Heal, C. (2001) 'Investigating the mathematics subject matter knowledge of pre-service elementary school teachers'. In M. van den Heuvel-Panhuizen (ed.) *Proceedings of the 23rd Conference of the International Group for the Psychology of Mathematics Education*, Volume 4 pp. 121-128. Utrecht, The Netherlands: Freudenthal Institute, Utrecht University.