Using multivariate data as a focus for multiple curriculum perspectives at secondary level.

Ridgway, Jim
University of Durham, School of Education
Leazes Road
Durham DH1 1TA, United Kingdom
E-mail: jim.ridgway@durham.ac.uk

Nicholson, James
University of Durham, School of Education
Leazes Road
Durham DH1 1TA, United Kingdom
E-mail: j.r.nicholson@durham.ac.uk

McCusker, Sean
University of Durham, School of Education
Leazes Road
Durham DH1 1TA, United Kingdom
E-mail: sean.mccusker@durham.ac.uk

1 Introduction:

Multivariate data are commonplace in children's lives and they routinely encounter a degree of complexity which goes far beyond anything they meet within the school curriculum. Pupils engage in debate about the relative merits of football teams, pop bands, teachers, computer games, mp3 players and the like where the comparisons are not like for like exactly but decisions still have to be made about what to buy, or arguments put forward to justify why a team or band is liked.

Historically, subject areas which deal with complex contexts, such as citizenship or geography, have tended to deal qualitatively with the different factors in any particular situation and not introduce the data because of the difficulties in handling multiple variables, both conceptually and practically.

Technology offers the opportunity for pupils to engage with multiple variables in contexts they are familiar with. The visualisations help them build mental models of the semi-quantitative relationships between 3, 4 or 5 variables - so they have a grasp of which factors make substantial differences, and some feeling for the relative size of the effects of different factors or variables. Schield (2005) rightly observes that when very large data sets can now be routinely analysed, statistically significant differences abound, and the notion of effect size is a much neglected concept.

We have argued elsewhere (Nicholson *et al.*, 2006, Ridgway *et al.*, 2007, in press a) that the current curriculum in the UK provides an inadequate preparation for full participation in an adult world where evidence-based policy is a key government objective, and where the commercial sector is increasingly dependent on making sense of all the information available in a very fast moving and competitive environment. There is evidence that young students can make sense of multivariate data, given appropriate support - which in the case of the very young students reported by du Feu (2005) was the use of lego blocks to build a tactile graph, and in the case of the 14 - 15 year old students reported by Ridgway *et al.* (2006) was the use of computer-based interfaces.

If students are to understand their world, and to make informed decisions about their own lives, and

in their capacity as informed citizens, they need to be able to reason with complex and multivariate evidence. Statistics and statisticians have a great deal to contribute to help develop this understanding, but students in school are exposed to very few valuable and empowering ideas. Rather, they are obliged to master pre-computer statistical techniques applicable to a very narrow range of applications, to the exclusion of reasoning with evidence.

Here, we describe work on curriculum development that engages students in issues relevant to their own well being, that requires engagement with multivariate data taken from large scale surveys. Qualitative and semi-quantitative reasoning often provides a good basis for action [quantitative analyses via statistics packages would often challenge graduate students]. Activities are designed to encourage students to acquire some heuristics that are useful when reasoning with complex data. These heuristics include:

- Critique the quality of the data
- Describe and explore phenomena before you try to explain things
- Focus on effect size not significance level
- Check that the effect size is a lot bigger than the likely error of measurement
- Identify variables that have the strongest effects
- Look at absolute levels – are they big enough to be worth worrying about?
- Look for non-linear relationships
    - Explore the effects over different values of each variable
- Look for changes over time
- Look for interactions, and think about 'data surfaces'
- Think about possible confounding variables
- Disaggregate data, and see if the patterns of relationships stay the same as in the aggregated data
- Look for the 'dog that didn't bark' – were there things you expected to see, but didn't?
- Be cautious of claims about causality – especially in observational data

Many of these heuristics can only be acquired by working with multivariate data – such as looking for interactions. Here, we show tasks and describe activities, and illustrate student awareness of powerful heuristics by describing some classroom events.

2 The Northern Ireland curriculum at Key Stage 3 (ages 11 - 14).

Northern Ireland Council for the Curriculum, Examinations and Assessment (CCEA) are currently introducing a new curriculum for Key Stage 3 (pupils aged 11 - 14). This is the start of secondary school in the UK, and for many pupils it represents a major upheaval in the structure of their school day. In primary school, pupils will often stay in the one classroom for most of the week apart from time spent on physical education, art and music, and would commonly have a single teacher for all other subjects. On transfer to secondary school, which is usually much bigger as well, they will often move classroom between 5 and 8 times in the day, studying upwards of 12 subjects, mostly with a different teacher for each subject.

CCEA (2002) noted research evidence that relatively slow progress was being made by many pupils as they made the transition into Key Stage 3. Pupils felt the curriculum repeated too much that they had already covered, that they were insufficiently challenged, and what they did in school was not relevant to them now, or in the future (Lord and Jones, 2000). The situation has not changed much in the interim (see Lord and Jones, 2006). CCEA also state (2002):

In addition, patterns of employment are changing and, consequently, the needs of employers. Feedback from employers' organisations indicates concern about levels of literacy and numeracy in the workforce, but there is also a feeling that those leaving schools and colleges should be better equipped for the world of work. This means better able to work with other people, better self-management and better problem solving.

These are skills that schools help to develop in young people, not just as preparation for working, but preparation for the whole of life.

The other important change in society is the development of new technologies. These are having an ever increasing effect in the classroom and need to be reflected in the curriculum. Information is easy to find now and knowledge for its own sake is less important than it once was. It is more important that we know how to make use of information, how to evaluate its usefulness, to sort out patterns and to solve problems with it.

The revised Northern Ireland curriculum encourages a move towards more collaborative work between traditional subject areas, a greater relevance of the curriculum to learning for life and work, to developing pupils as individuals, as contributors to society, and as contributors to the economy and the environment.

A common theme across these laudable aims is to introduce a more realistic view of the complexity of problems in the real world. The ability to manage information and handle data across the curriculum from real world contexts is one of the important skills that the revised curriculum seeks to develop.

3 The Durham - CCEA curriculum project

CCEA have commissioned us to produce some exemplars of data-rich resources to support these changes, and in particular to act as a focus for different curriculum perspectives.

The Northern Ireland Commissioner for Children and Young People [http://www.niccy.org/] has identified *risk-taking behaviour in the young* and *children and poverty* as two of its current priorities. Both these priorities offer opportunities for using the visualisation tools we have developed to show the likely consequences of different types of behaviour, and to display comparative data between Northern Ireland and other places. Pensions and savings have been areas the educational establishment have wanted to engage youngsters with for some time, and have so far had limited success. We believe that the interactive materials we are developing in this project have the potential to engage the pupils with some of the big ideas in these areas.

To illustrate the materials, and the general approach, we can only offer some static screen shots with commentary, but interactive versions of these, and other, materials can be found at www.dur.ac.uk/smart.centre/CCEA

3.1 Alcohol use by 11-15 year olds in the UK

This is now a major social issue in the UK, and schools are expected to deal with aspects of it in a number of curriculum areas. There are health related aspects in science, as well as aspects which are central to citizenship and to personal development. Apart from curriculum areas which have to deal with it, there are others which could usefully contribute a perspective to it: for example in mathematics, lessons could focus on the stories which the data tell, what other data it would have been useful to have access to in order to understand more fully what is going on, the need to be cautious about attributing causal reasons to patterns seen in observational studies; in geography, lessons could focus on exploring differences between different groups and different cultures.

Our materials provide a number of datasets about alcohol use in the 11 to 15 year old age group. Before they start to view them students are asked questions about what they expect the data to show, e.g. what proportions of 11-15 year olds are drinking alcohol in the UK, whether they think there is any difference between boys and girls, and any difference between ages. Figure 1 shows the data interface they can access to get answers to these questions for children in 2004. By dragging the pointer at the bottom across, they can see the corresponding data for 1990 up to 2004 to allow investigation of any trends over time. The 3 variable labels here (sex, age, and year) are interchangeable in position by a simple drag and drop action, which allows the user to control the priority of the variables in their display e.g. if sex and year are interchanged, the user will see only male or only female data in the comparative bar chart, but will see the time series shown for each age.

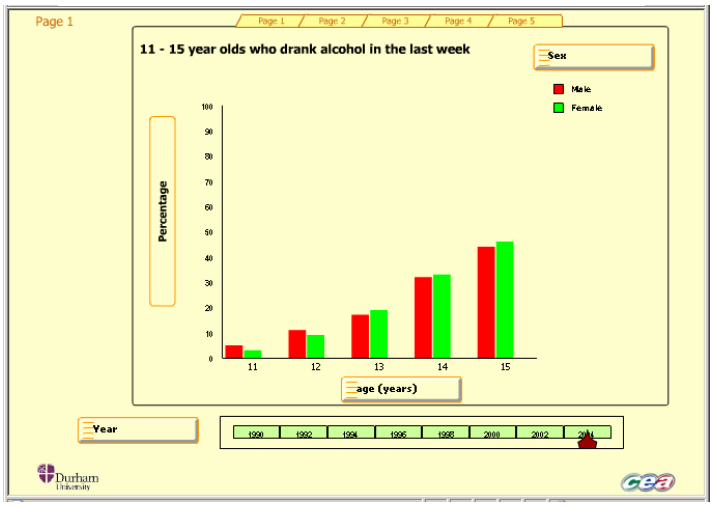

Figure 1: Rates of alcohol use among young people in the UK.

This data set provides some information about alcohol consumption in the young, for example that around 50% of 15 year old females had had an alcoholic drink in the previous week in the 2004 survey (the bar on the extreme right of figure 1), but it is a long way from the whole story - so other data sets allow the pupil to explore how often young people drink, how much and what they drink. One important aspect of the materials is that they open with newspaper headlines about the scale of the problem e.g. **Curb on teenage drinkers and smokers** in *The Daily Telegraph, September 14, 2006*, but the pupils are then being asked to interpret the data for themselves, and indeed whether the headlines are reasonable summaries of the data or if they sensationalise a small aspect of an otherwise unremarkable situation. Figure 2 shows the average weekly consumption of alcohol by young people who drink in the UK, so for example, in 2004, the average weekly consumption of 15 year old males who drank alcohol was 12 units, or 6 pints of beer. Because the interfaces allow the visualisation of relationships between variables, pupils naturally address issues relating to effect size in describing the relationships.

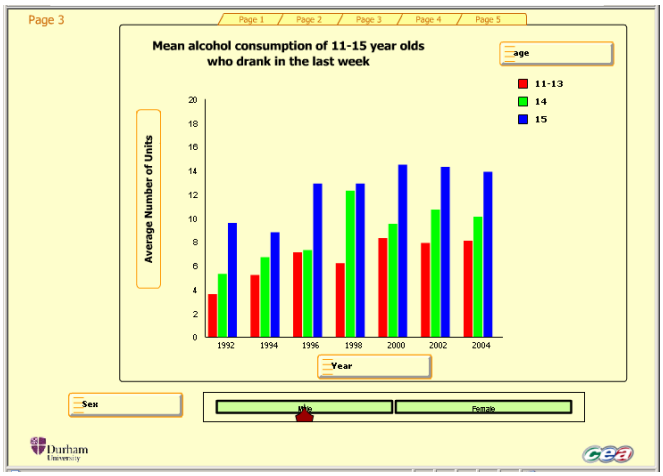

Figure 2 Average weekly alcohol consumption by young people in the UK.

Our brief is to produce data-rich resources which link with other curriculum materials, so these data explorations will be used before various curriculum areas consider alcohol use from their own perspective.

3.2 Poverty in Northern Ireland and other parts of the UK

Here we have started with the UNICEF (1999) activity on the rights of the child. Once the class

have worked though this, and made their own decisions about what a child should be entitled to, they can compare their answers to the results of a national survey and explore any differences. This links to exploring the deprivation indices data (Noble *et al.*, 2001) for local government districts (LGDs) in Northern Ireland.

There are 7 indices, and one composite index of these 7, as well as a child poverty index which focuses only on the aspects most directly affecting a child in Northern Ireland.

The interface shown in figures 3a and 3b allows the user to choose which of these indices to show as the variables in a scatter graph. All the LGDs are listed down the side, and up to 3 LGDs can be selected either by their position on the scatter graph, or by their name in the list. When selected, both of the references to an area is highlighted in the same colour. When the user changes variables on the axes, the same LGDs remain selected so the user can explore whether areas which do badly by one measure of poverty also do badly by others. While some of the indices show strong correlation, as in 3a, others do not, as in 3b, but 2 of the 3 highlighted areas are down amongst the most deprived areas on all 4 of the measures represented in these two graphs.

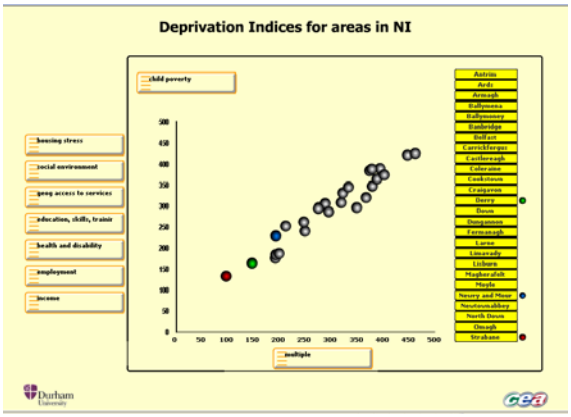

Figure 3a Deprivation indices interface

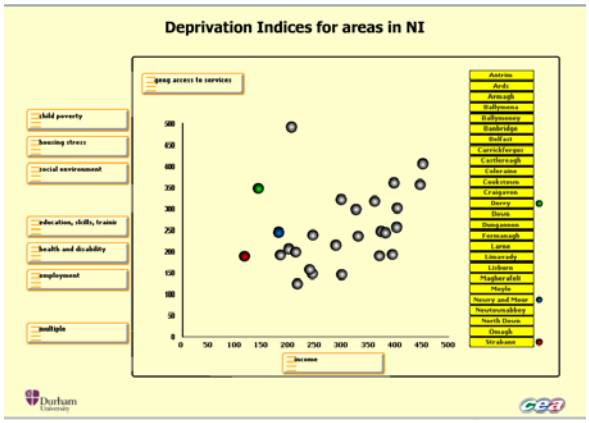

Figure 3b Deprivation indices interface

Further data interfaces allow poverty data in Northern Ireland to be compared to other areas in the UK, and can be extended to comparisons with other regions in the world.

3.3 Pensions and savings

The initial activities draw out the rationale for larger pension payments to people who start drawing their pension at an older age, exploring why women seem to be discriminated against by getting a lower pension for the same pension pot at the same as than a man would, and why smokers appear to be rewarded for their habit by getting a larger payout in pensions.

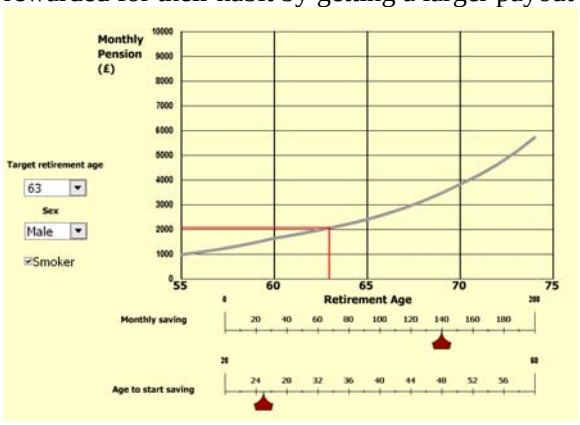

Figure 4: Pensions and savings activity

In figure 4 the user states whether they are male or female, smoker or not, and chooses a target age for their retirement. The graph displays the monthly pension available for different retirement ages depending on the monthly savings and the age at which savings started. The red line indicates the pension payable at their target age for retiring.

Associated activities will link life styles in retirement with monthly pension being paid, and life styles during working life with savings strategies to illustrate the reality and consequences of different behaviour.

4 Developing curriculum in partnership with teachers

The project described above has the curriculum development generated in Durham, and then trialled in schools. We are currently working with the Geographical Association, on a project funded by the Wellcome Trust relating to understanding risk using disease as a context, which uses a different model of curriculum development. Pairs of teachers in mathematics and geography from 4 schools are using our interfaces to develop cross-curricular materials i.e. more of the curriculum development is being done by the teachers, with support from Durham in tailoring data interfaces and simulations to suit the context chosen, and the 4 schools have each developed quite different approaches, and each is working with a different disease. The resources and a fuller report are available at www.dur.ac.uk/smart.centre/disease

Danum Technology College are looking at skin cancer: they wanted something immediately relevant to pupils so they are looking explicitly at risk factors and effective interventions – the aim is to not 'preach' but allow students to discover the important messages for themselves. Aston Comprehensive School are looking at whether parents should give their children the MMR vaccine because their local area has had a big debate about this recently – the pupils are collecting data about vaccination patterns from current students and older generations and will look at local, national and global vaccination patterns and the aim is to produce a magazine article summarizing their findings. St Ivo School is holding a mock World Health Organisation summit on combating tuberculosis, where groups of pupils will represent different countries and their interests. Wycombe High School, a Mathematics and Computing Specialist School, are planning to involve almost all their departments with mathematics, geography, citizenship, biology, ICT and history playing major roles and engaging students in analysis, decision making and interpretation in real contexts – including looking at the potential spread of malaria in Europe as one of the effects of global warming.

5 Factors affecting student performance

The introduction of a compulsory extended statistics coursework project in the high-stakes qualifications for mathematics at age 16 in the UK in 2000 caused considerable angst for teachers and pupils. The expectation is that pupils should deal with complex data sets, yet this was outside their curriculum experience (Nicholson, 2003); the difficulties were compounded by the lack of suitable resources to build up student confidence in working with complex data, or any guidance for teachers as to the hierarchical structure of the skills in this area.

There was a paucity of materials available: textbooks contained almost no examples of data more complex than a linear bi-variate relationship, and rarely any guidance on appropriate language to use in describing relationships, or even on strategies to use to decide whether a difference [and there will always be some difference in observed data] is sufficient for it to be reasonably be described as a difference.

As these examples illustrate, expert teachers can provide very creative inputs into curriculum development. Ensuring that these materials are robust enough to allow their use in representative classrooms by representative teachers poses a major challenge however.

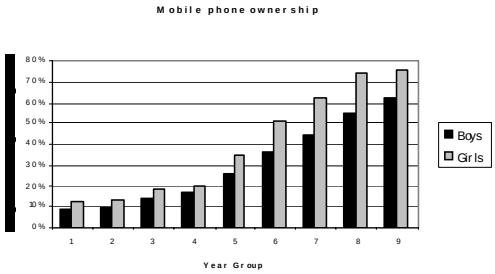

Figure 5a: Both variables show a difference

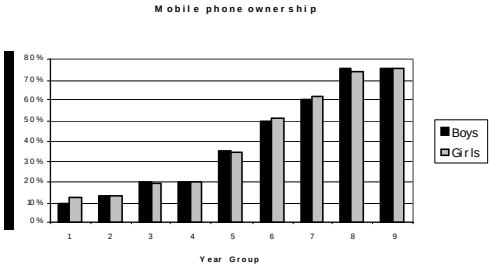

Figure 5b: One variable shows a difference

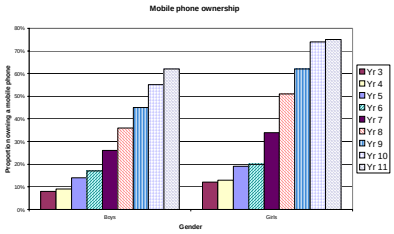

Figure 6: Both variables show a difference, data for boys and girls grouped separately.

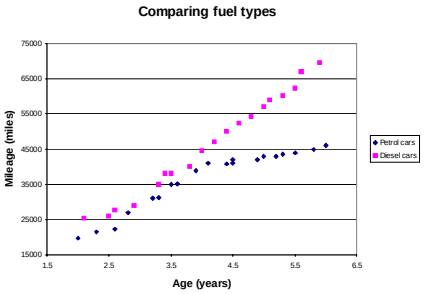

Figure 7a: Both variables show a difference

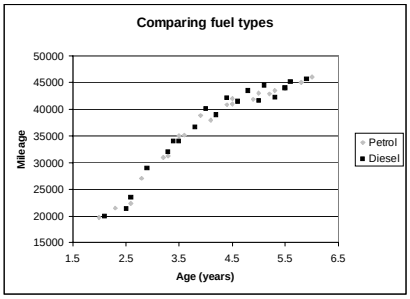

Figure 7b: One variable shows a difference

Initial analysis suggests that items such as these on reasoning with more complex data fit securely into the Watson-Callingham hierarchy of statistical literacy (see Ridgway *et al.*, 2006) and are at accessible levels of difficulty. It was noticeable that items where differences were substantial in both explanatory variables (sex and age) as in figure 5a; (age of car and fuel type) as in figure 7a were easier for students to deal with than very similar items where a substantial difference was present in only one of the explanatory variables, as in figures 5b and 7b. In figure 6 there are again differences in both variables, but in this case the information is grouped differently, so instead of seeing pairs of values for boys and girls of each age, and with only two, strong, colours used the graph shows population profiles for boys and girls, but now the graph requires 9 different colours or textures for the different ages. More work needs to be done to investigate whether there is a consistent pattern as to which grouping is easier to interpret, or if it might depend on the numbers of categories in each classification and / or the nature of the relationships. For example, it might be that if the populations have very different profiles that this is easier to describe from the representation used in figure 6, but when they are similar it is easier to compare the detail by having the representation as in figure 5a.

It appears that students can deal with greater complexity in data if they have the appropriate tools and visualization support. It will take some time before we can gather evidence at to whether students better understand critical issues facing them and their world if they are better at understanding quantitative evidence, but we conjecture that this will be the case. There is still a lot of work to be done to understand how reasoning from evidence develops in the context of extensive ICT support, but there is initial evidence to suggest that these new interfaces offer new opportunities for students, and that they are accessible to a wide spectrum of academic ability. We have conducted some small scale trialling of this

hypothesis, with around 200 pupils aged 11 to 15 in Durham and Belfast, and the initial outcomes do show an interleaving of these items with those used by Watson & Callingham rather than a concentration at the upper end of the difficulty spectrum.

7 Discussion

Assessment is an arena in which curriculum development often suffers as the practicalities of formal testing regimes force classroom practice in ways which end up negating the central cognitive tasks. Employers emphasise the need for their future employees to be able to work in teams, yet almost all current assessment is on an individual, competitive basis. Ridgway *et al.* (in press b) discuss the opportunities for assessing collaborative contributions in using these new materials as a critical part of the development.

The relevance of the school curriculum to life after school is not an over-riding principle: there are many aspects of the curriculum whose value derives more from their aesthetic quality, or their capacity to develop modes of thinking than from a purely utilitarian perspective. However, where new curriculum developments have the potential to be more relevant as well engaging students who are not well motivated by their current curriculum diet, we believe they offer an important opportunity to be explored.

Further curriculum development is needed, and we need more research on the hierarchy of understanding reasoning from evidence, but the initial evidence is that embedded multivariate data can provide a focus for multiple curriculum perspectives which engage students and provide more realistic and relevant experiences than their current curriculum.

REFERENCES:

Chance, B. (2002). Components of Statistical Thinking and Implications for Instruction and Assessment, *Journal of Statistics Education*, vol 10(3).

du Feu, C. (2005) Bluebells and Bias, Stitchwort and Statistics. *Teaching Statistics*, 27(2), 34-36

Lord, P. and Harland, J. (2000). *Pupils' Experiences and Perspectives of the National Curriculum: Research Review* [online]. Available: http://www.qca.org.uk/254_1956.html

Lord, P. and Jones, M. (2006). *Pupils' Experiences and Perspectives of the National Curriculum and Assessment: Final Report of the Research Review*. Slough: NFER.

Nicholson, J.R. (2003). GCSE statistics coursework, and dealing with multi-dimensional data. *Mathematics in Schools, 32(3),* 8-14.

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

Northern Ireland Council for the Curriculum, Examinations and Assessment (CCEA), (2002) A New Approach to Curriculum and Assessment 11-16

Ridgway, J., Nicholson, J.R., & McCusker, S. (2007) Teaching Statistics – Despite its Applications. *Teaching Statistics*, 29(2), 44-48.

Ridgway, J., Nicholson, J.R., & McCusker, S. (in press a) Reasoning with evidence – new opportunities for assessment, curriculum and public understanding.

Ridgway, J., Nicholson, J.R., & McCusker, S. (in press b) Embedding statistical assessment within cross-curricular materials.

Schield, M. (2005) Statistical Literacy Curriculum Design in *Burrill, G., & Camden, M. (Eds.). (2005) Curricular Development in Statistics Education: International Association for Statistical Education 2004 Roundtable. Voorburg, the Netherlands: International Statistical Institute*

Noble, M., Smith, G., Wright, G., Dibben, C. and Lloyd, M. (2001) *The Northern Ireland Multiple Deprivation Measure 2001*. Belfast: Northern Ireland Statistics and Research Agency.

UNICEF (1999) Talking rights, taking responsibility: A Speaking and Listening Resource for Secondary English and Citizenship. UK Committee for UNICEF, UK.

Watson, J. M. & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal,* 2(2), 3–46.