



THE
JUBILEE CENTRE
FOR CHARACTER & VALUES

UNIVERSITY OF
BIRMINGHAM

Assessing Interventions Designed to Improve Understanding of Virtues

Tom Harrison and Ian Davison

These are unpublished conference papers for the 'Can Virtue Be Measured?', held by the Jubilee Centre for Character and Values at Oriol College, Thursday 9th – Saturday 11th January 2014. These papers are works in progress and should not be cited without author's prior permission.

**School of Education
University of Birmingham
Edgbaston
Birmingham**

Assessing Interventions Designed to Improve

Understanding of Virtues

Ian Davison, Centre for Research in Medical and Dental Education, and Tom Harrison, Jubilee Centre for Character and Values, University of Birmingham

Abstract

This paper draws upon our experiences working on two projects funded by the John Templeton Foundation: a) Pilot and Feasibility Cluster Randomised Controlled Trial into the Effectiveness of Character Sessions (My Character Project); and b) Quasi-Experimental Trial of the Knightly Virtues Programme. We discuss aspects of experimental design including randomisation, clustering of data and outcome measures.

The My Character project ran from September 2013 until July 2014 and involved 27 classes from 6 secondary schools. The pre- and post-tests were identical, consisting of a) 'I Believe' Questionnaire which asks children to rate their agreement to 40 statements related to the 8 virtues addressed during the project; and, b) a monetary choice questionnaire to assess children's propensity to delay gratification. Results will be presented regarding the feasibility of running such a trial.

The Knightly Virtues Programme seeks to engage Year 5 and 6 children with 4 classic tales: Gareth and Lynette from Arthurian legends, El Cid, Don Quixote, and the Merchant of Venice. Virtues are highlighted, discussed and related to the children's own lives. The programme trial ran from September to December 2013. Six experienced primary school teachers acted as a consultation group and helped to develop and mark the outcome measure, which was structured around the Key Stage 2 English Reading Test but marked for understanding and application of virtue concepts as well as English comprehension.

In summary, running educational trials is easier than clinical trials but it's still very difficult to recruit and retain schools, teachers and children. Before embarking on such a trial, you need: a) an interesting, high quality programme that appears to improve children's understanding of virtues; b) excellent relationships with schools for them to adhere to the trial protocol and, c) robust, piloted outcome measures.

Introduction

As trials are the gold standard to obtain evidence about 'what works', one perspective is that teachers should welcome the opportunity to participate in experimental trials to test new educational interventions. This would help ensure that only educational resources that have been proven to have a positive impact make it to the classroom. Rigorous trials should be the basis for more sustainable, consistent and successful educational policies. Recently there has been a call for more trials in educational research (Goldacre, 2013) and it has been argued that 'without subjecting curriculum innovations to a RCT then potentially harmful educational initiatives could be visited upon the nation's children' (Torgerson and Torgerson, 2001, p316). Elsewhere, these authors have given numerous examples of RCTs across the social sciences and suggest that it's easier to conduct such trials in education than medicine as it's easier to recruit participants (Torgerson and Torgerson, 2008).

Despite this, traditionally, there has been resistance to conducting trials in education (Oakley et al., 2005) and perhaps as a result few schools are encouraged to participate in them. The objections to RCTs in education are often made on ethical grounds i.e. disadvantaging one group with respect to another, that contamination between arms of the trial are high, and that you cannot blind participants to the educational intervention (Hutchison and Styles, 2010). In addition, it is claimed that the government insistence on randomised trials is controlling the educational research agenda in the USA, leading to bad science, managerialism and reducing the role of expert opinion; this claim is that there is little evidence that the trials approach works as the "complexity and the messiness of practice-in-context cannot be fantasized away" (Lather, 2004, p768). However, our view is that Lather's critique is of an overly positivist approach that discounts qualitative research and philosophical considerations.

Medicine has a strong history of conducting Randomised Control Trials (RCTs) to assess the effectiveness of treatments, particularly drugs. Rigorous methods have been developed by the CONSORT (Consolidated Standards of Reporting Trials) group (Moher et al., 2010) and all medicines have to undergo three phases of clinical trials to be licensed in the UK (MHRA, 2008).

Applying the standard design of RCTs to understanding of virtues involves assessing children's performance at baseline with a pre-test, randomly allocating them into the intervention and control groups; then re-assessing the children's performance after the intervention with a post-test that is the same or very similar to the pre-test. As RCTs control for all other factors, only an intervention that genuinely improves children's understanding of virtues should produce a positive, significant result. Quasi-Experimental Trials are similar but do not have the random allocation. This paper considers our endeavours to address two key issues in assessing the success or otherwise of interventions designed to improve understanding of virtues in schools: i) the development of pre- and post-tests that reliably assess efficacy of interventions designed to develop character and virtues, and ii) the practical aspects of running trials in schools.

The Knightly Virtues: developing pre- and post-tests designed to assess understanding of character language

Young people in Britain today lack a knowledge and understanding of character and virtue language; this issue has been neglected in schools in Britain (Arthur, 2010). The overarching aim of the Knightly Virtues programme was to use classic narratives and reflective learning to improve Primary school pupils' understanding and use of virtue language concepts. To date, nearly 4,000 young people across Britain have experienced the programme which uses narratives that have been adapted to focus on eight virtue concepts: humility, honesty, love, service, courage, justice, self-discipline and gratitude. Here we report on the development of a measure designed to assess the impact of the programme. A consultation group, consisting of experienced primary school teachers, is supporting this work. The following research questions were agreed after preliminary analyses and discussions with these teachers.

- RQ1. Does the Knightly Virtues' programme increase children's
- a. Understanding of the meaning of virtue terms?
 - b. Interpretation of virtue concepts?
 - c. Application of virtues terms in historical and modern day contexts?
 - d. Application of virtue terms in own personal contexts?

- RQ2. Do potential findings in RQ1 apply equally
- a. In faith and non-faith schools?
 - b. For boys and girls?
 - c. In Year 5 and Year 6?
 - d. In inner city and rural schools?

Here we consider how to assess RQ1 effectively. Note that we're looking for understanding and use of virtue language. There is no attempt to assess changes in attitudes or behaviour as we felt this would be unrealistic. After much deliberation and piloting, we discounted the use of questionnaires as it is so difficult to assess depth of understanding in this way. Consequently, the assessment was based upon the Key Stage 2 English Reading test and contained stories similar to those found in the Knightly Virtues programme. The style of the assessment was designed to be familiar to the children so that little explanation from teachers would be required. Two very similar assessments were developed so that pre- and post-tests followed the design in Table 1. Each trial school has an equal number of experimental and control classes, usually one of each. Mostly, one of these classes is in Year 6 and the other in Year 5.

Table 1: Knightly Virtues experimental design

Group	Pre-test	Intervention	Post-test
Experimental 1	Test A	Knightly virtues	Test B
Experimental 2	Test B	Knightly virtues	Test A
Control 1	Test A	Normal teaching	Test B
Control 2	Test B	Normal teaching	Test A

The consultation group also were integral to the development of a mark scheme that was appropriate for assessing research question one. As it was important to separate children's reading skills from their understanding and application of virtue concepts, reading comprehension is also being assessed. Reading comprehension should be scarcely affected by the Knightly Virtues programme as it takes more than a few lessons to improve in this area; therefore it should act as a control variable to see if the experimental and control groups have similar levels of ability, motivation etc.

Currently the consultation group is piloting the mark scheme; then they will mark a sample to estimate inter-rater reliability, before marking all the scripts.

My Character: The practical aspects of running an RCT in schools

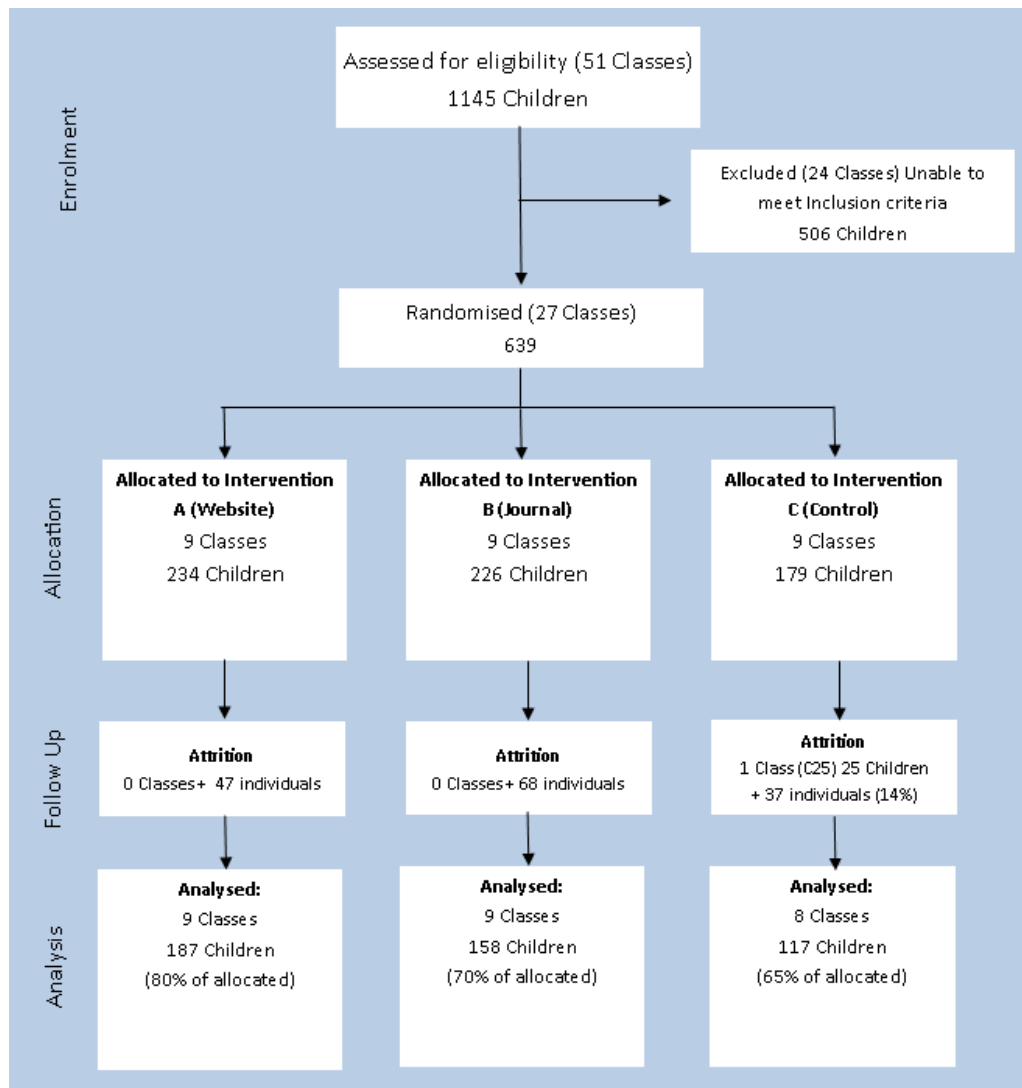
In this section we draw on our experience of running a “Pilot and Feasibility Cluster Randomised Trial into the Effectiveness of a Website or Journaling to Enhance Future-mindedness”. The My Character project consisted of two educational interventions, a website and hard copy journal, which were designed to develop the future-mindedness of 11-14 year olds. The interventions were developed in collaboration with young people and educational professionals from across Britain. The journal and the website contain activities and stimulus materials that encourage young people to explore virtues associated with future mindedness, such as delaying gratification and being thrifty. After development, over 1000 young people in six schools were involved in piloting the materials over the course of a year. Here, we explore the lessons learnt from running a Randomised Control Trial as part of this pilot.

There is a hierarchical structure of Pupils nested within Classes which are nested in Year groups and within Schools. To take part in the trial, there must be a multiple of 3 classes within a single school and year group that could be randomly allocated to the website, journal and control groups i.e. it is a stratified design. This constraint was imposed “to minimise imbalance across treatment groups” (Campbell et al., 2004, p705). Also, no school would want to participate as a control.

To assess the impact of the programme, the pre- and post-tests were identical, consisting of a) ‘I Believe’ Questionnaire which asks children to rate their agreement to 40 statements related to the 8 virtues addressed during the project; and, b) a monetary choice questionnaire to assess children’s propensity to delay gratification (Kirby et al., 1999). Figure 1 gives recruitment and retention data for one of these (I Believe questionnaire); it shows that 27 classes in the 6 schools were allocated to the RCT; 24 other classes were not in multiples of 3 and so were not enrolled in the trial. Attrition was caused by one class not completing the post-test questionnaire and by children being absent or otherwise not completing either the pre- or post-questionnaire. The primary aim was to evaluate the feasibility of conducting such a trial. Clearly one class out of 9 dropping out of the Control

group adds substantially to the attrition. Also, more children did not complete the post-trial (119, 19% excluding the missing group) than the pre-trial (63, 10%) questionnaire. This may be because of more absences at the end of the school year, more young people chose not to participate or some other reason.

Figure 1: Flow diagram of classes and children in the My Character trial (I Believe data)



A secondary aim is to explore possible differences between the website, journal and control groups: these analyses have yet to be undertaken, but we intend to use random effects, multilevel (hierarchical) analyses in MLwiN (Rasbash et al., 2009).

The 6 schools that volunteered to take part in the trial, wanted to be involved in what they saw as a high quality, well developed programme with sound educational aims. Some of the schools also saw the benefits of the trials, but for most this was a secondary motivation. Ensuring the schools both understood and were able to meet the challenges of administering the trials was a priority for the research team. Even so, many of the excellent lead teachers, who were responsible for running the research in their school, took on this commitment blindly as they were unfamiliar with running trials and associated protocols. The teachers were expected to gain approval from the students and their parents/guardians, explain how to administer the trial to the other teachers involved, and to ensure the trial protocol was observed which sometimes meant making changes to timetables and rooms. All these factors created work for the teachers, in addition to their normal busy workloads. In several schools, the research team had difficulty encouraging the lead teacher to fulfil their role, either because they had under-estimated the expectations or, in the case of two schools, there was a replacement lead who was less enthusiastic. This led to some complications, most noticeably late or non-return of questionnaires. Unsurprisingly, the R.C.T. is a relatively low priority for schools, so was in danger of being side-lined for example by an Ofsted inspection.

To avoid potential contamination, each teacher could only teach website, journal or control classes. Teachers in the different groups were asked not to discuss or share the intervention with each other. This made curriculum planning meetings particularly difficult and goes against the collaborative culture found in schools. Furthermore, whoever was selected to teach the website group needed regular access to a computer suite. Therefore, involvement in the trial required re-organisation of staffing and rooms on the timetable in some schools.

Finally, ensuring that the pre and post questionnaires were completed as determined by the protocol was also an issue. The size and geographical spread of the trial meant it was impossible for the researcher to be present whilst the questionnaires were being completed; therefore we had limited control over when and how they were completed.

Discussion

We are extremely grateful for the dedication of numerous teachers and the hard work of young people, without which it would be impossible to conduct research of this kind. Fundamental to the success of a trial is a good working relationship with the schools: we have worked closely with the My Character schools for many years, and so were able to persuade these schools to change staffing and rooms. To do this, the researchers had to acknowledge the extra burdens they were imposing and understand the particularities of school cultures and normal working practices. In particular, the research team had to ensure that it was as easy as possible for the lead teacher to administer the trial in their school by preparing all the materials (ethics letters, pre and post questionnaires etc) and communicating clearly how these should be administered.

It was also important for the lead teachers to see their role as 'researcher in situ' and take on the responsibility for the maintenance of the trial protocol in their particular school. Regular face to face meetings between the research team and lead teachers helped the teachers envisage and carry out this largely unfamiliar role. These visits enabled the teachers to both understand the reason for conducting the research and adhering to the protocol. Trust, between the teachers and researchers, was consolidated during these meetings which helped to engender commitment to the success of the trial.

A weakness in the design of the My Character RCT is that the outcome measures were questionnaires which cannot assess complex understanding. In the Knightly Virtues evaluation, we attempted to address this by involving practicing teachers in the development of a more detailed, realistic test. Analysis is yet to begin, but it is likely that further iterations will be required. This consultation group of teachers are highly dedicated and professional; even so, they are finding the change from being a teacher to a researcher challenging. Particularly, they struggle with the idea of assessing work to evaluate the intervention rather than the children.

In the Knightly Virtues trial, we seek to assess changes in understanding of virtue terms. This is a much more modest aim than trying to observe changes in attitudes or behaviour, which would require much larger interventions, probably involving changes to school culture.

Conclusion

Oakley (2006, p64) states that RCTs in education are “ethical and useful”; this can only be true if they work in practice, so this paper has explored some of the practical challenges. If the impact of these challenges can be negated then it is feasible to run trials designed to assess character in schools. We regard appreciation of educational research as a worthy aim, and so are delighted that the lead teachers, many of whom were already advocates of character education, understood the need to test new interventions to show what works and appreciated that this would make a stronger case for other schools to take part. These teachers also saw other benefits from being involved, such as the research being educational for their students, and welcoming the association with the University of Birmingham and Jubilee Centre for Character and Values.

In summary, a high quality teaching programme and excellent relationships with schools are required to recruit and retain schools, teachers and children. The schools need continual encouragement to continue with a trial. Finally, robust, piloted outcome measures are essential for a worthwhile trial. All these elements should be in place before undertaking a randomised controlled trial into character education.

Acknowledgements

We are indebted to countless teachers and young people who have participated in these projects, particularly the teachers in our consultation group. Also, we value the wisdom provided by other members of the project team, especially James Arthur, Jon Davison and David Carr. Finally, we are most grateful to Dan Hayes and Jenny Higgins who have undertaken so many administrative and data handling tasks.

References

- Arthur, J. (2010) **Of Good Character; Exploration of Virtues and Values in 3-35 year olds**. Exeter: Imprint Academic.
- Campbell, M.K., Elbourne, D.R. and Altman, D.G. (2004) CONSORT statement: extension to cluster randomised trials. **British Medical Journal**, 328: 702-708.

Goldacre, B. (2013) "Building Evidence into Education <http://media.education.gov.uk/assets/files/pdf/b/ben%20goldacre%20paper.pdf> [accessed 19/12/13]".

Hutchison, D. and Styles, B. (2010) "A Guide to Running Randomised Controlled Trials for Educational Researchers". Slough, NFER.

Kirby, K.N., Petry, N.M. and Bickel, W.K. (1999) Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. **Journal of Experimental Psychology: General**, 128: 78-87.

Lather, P. (2004) Scientific research in education: A critical perspective. **British Educational Research Journal**, 30: (6): 759-772.

MHRA (2008) **Medicines and Medical Devices Regulation: What you need to know** [online]. <http://www.mhra.gov.uk/home/groups/comms-ic/documents/websiteresources/con2031677.pdf>
The Medicines and Healthcare Products Regulatory Agency [Accessed 08/12/2013]

Moher, D., Hopewell, S., Schulz, K.F., et al. (2010) CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. **British Medical Journal**, 340: (mar23 1): c869-c869.

Oakley, A. (2006) Resistances to 'new' technologies of evaluation: education research in the UK as a case study. **Evidence & Policy**, 2: (1): 63-87.

Oakley, A., Gough, D., Oliver, S., et al. (2005) The politics of evidence and methodology: lessons from the EPPI-Centre. **Evidence & Policy**, 1: (1): 5-32.

Rasbash, J., Steele, F., Browne, W., et al. (2009) **A User's Guide to MLwiN: Version 2.10**. Centre for Multilevel Modelling, University of Bristol.

Torgerson, D. and Torgerson, C. (2001) The Need for Randomised Controlled Trials in Educational Research. **British Journal of Educational Studies**, 49: (3): 316–328.

Torgerson, D. and Torgerson, C. (2008) **Designing randomized trials in health, education and the social sciences: an introduction**. Basingstoke: Palgrave Macmillan.