

Teaching character

Harrison, Tom; Burn, Emily; Moller, Francisco

DOI:

[10.1080/00131911.2018.1538937](https://doi.org/10.1080/00131911.2018.1538937)

License:

Other (please specify with Rights Statement)

Document Version

Peer reviewed version

Citation for published version (Harvard):

Harrison, T, Burn, E & Moller, F 2018, 'Teaching character: Cultivating virtue perception and virtue reasoning through the curriculum', *Educational Review*. <https://doi.org/10.1080/00131911.2018.1538937>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

This is an Accepted Manuscript of an article published by Taylor & Francis in Education Review on 9/11/2018, available online: <http://www.tandfonline.com/doi/full/10.1080/00131911.2018.1538937>

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Teaching Character; Cultivating Virtue Perception and Virtue Reasoning Through the Curriculum

Given the increased interest in character education across the world, educators are seeking guidance for their practice. This paper seeks to add to the evidence base about if and how character can be taught by discussing the results of a pilot evaluation (n 527) of a curriculum intervention designed to enhance two components of character: virtue perception and virtue reasoning. After the intervention the participants had improved virtue perception and reasoning scores compared to the control group. Further, female students scored higher in the pilot than their male counterparts. The results demonstrate how educating these components of character might be possible, providing evidence for a debate that goes back to Aristotle and today is keenly debated in academic, policy and practice circles. The findings are significant as they provide evidence as to how character might be taught through and within the programmes of study of existing curriculum subjects.

Keywords; Character; Virtue; Reasoning; Curriculum

Character Education Policy and Practice

The roots of character education arguably go back to the Greek philosophical tradition and are most notable in the writings of Plato and Aristotle. Recently, character education has witnessed a revival and intentional efforts to cultivate character qualities and traits in children and young people have become increasingly prominent in policy discussions in Britain and across the globe (Kristjánsson, 2015). Policy makers in the UK, from all political parties, have given major presentations on character education and most recently the former Secretary of State for Education, the Rt. Hon Nicky Morgan MP, has authored a book on character education entitled *Taught not Caught* (2017). The renewed or perhaps more explicit focus on character is seemingly broad enough to encompass theorists of different political persuasions (Ecclestone 2012) and ‘can no longer simply be written off as a project of the conservative right or the religious lobby, and has attracted supporters from across the political spectrum’ (Suissa, 2015 : 106/7). Kristjánsson (2013: 278) has sought to debunk the myth that character education is ‘conservative’ and ‘supportive of the *status quo*’. He cites Martha Nussbaum (1990) to argue that Aristotelian virtue-and-well-being theory, if transposed to the modern world, would have radically reformatory and progressive implications. Perhaps the most pressing point here is that some form of emphasis of character education, from Aristotle to the social reformer Robert Owen to more recent focus, has been present in discussions of varied provenance about the purpose of education.

Character education has also attracted academic, as well as political, attention in recent years. As character education has been generally conceived as a multi-faceted concept that embodies a range of explicit and implicit activities, previous studies have warranted contributions from across a range of disciplines, including: philosophy (see for example Carr, 2008); psychology (see for example Peterson and Seligman, 2004); and education (see for example Author et al, 2016). This work demonstrates the breadth of character education and the range of lenses through which it can be viewed. Despite this political and academic interest, a common point of departure is about the actual practice of character education, leading onto two important questions: *can* character be taught, and if so *how* should it be taught?

Aristotle was concerned that ‘we are inquiring not in order to know what virtue is, but in order to become good’ (Aristotle, 2009: 24). The implication is that attempts at virtue development would be futile if they did not issue in some explicit changes in conduct, in individuals actually behaving virtuously. A pertinent question then is how the possible gap between knowing the good and doing the good might be bridged and what would deliberate educational efforts that encourage young people to both know and do the right thing consist? This complex question has troubled educationalists, philosophers, psychologists and well as those writing from other disciplinary perspectives for many years. A clearer picture is being built up over time as new evidence emerges; the research reported on in this present paper has the modest ambition of providing a few more pieces to the complex jigsaw. The pieces are evidence about how virtue perception and virtue reasoning, two components of virtue (Jubilee Centre, 2017) might be taught. **Virtue Ethics and Character Education**

In order for character education to be successful, firm philosophical foundations are required from which to develop approaches to educating character and virtue. Questions about what is morality and what it means to be a moral person have persisted throughout history; most notably in the deontological traditions as characterised by the work of Kant and utilitarian traditions as characterised by the work of Bentham and J.S. Mill. There is not space within this paper to unpack the well-established arguments between these traditions in any depth,

and the paper focuses primarily on the most visible recent movement in the character education field; namely, the convincing case being made by moral and educational philosophers (see, for example, Kristjánsson) as to the advantages of virtue ethics as a moral theory on which character education interventions can build. Although Aristotelian conceptions of character education are not without their challenges, it is arguably well suited to provide a theory of change that makes a direct link from the education of virtues to human flourishing, and crucially emphasises the importance of human agency and critical reasoning in appropriate moral thoughts and actions. At its heart, virtue ethics emphasises flourishing as the widely accepted goal of human life and that character education should be a deliberate attempt to inculcate the virtues necessary for, not only individuals, but also for societies to thrive. For the purposes of this paper, character education should therefore be understood as any form of moral education focusing on the development of virtues as stable traits of character with the aim of promoting human flourishing (*Eudaimonia*) and founded on the general theory of virtue ethics (Kristjánsson, 2015). Virtues, such as compassion, honesty and courage, might be considered the building blocks of character as they are concerned with morally praiseworthy conduct (Author et al, 2017). The principles of Aristotelian character education present both an opportunity and a challenge for those concerned with character education. They invite innovation and creativity to develop new approaches as well as a challenge to understand how these new approaches can be applied in current educational systems where priorities often lie elsewhere.

An emerging practice-based model for character education

Educational models, built on sound theory, are important in education. They can also be popular, for example the ‘moral development model’ (Kohlberg, 1981) although heavily critiqued, is still used by some today. Theoretical models help educators understand seemingly complex conceptual frameworks in accessible formats. However, few in the field of character education are without challenge. This is because character is an inherently complex construct. Curzer (2016) argues that moral development should not be viewed as straight forward and moral progress cannot be understood as movement from one stage to the next. For Curzer

People’s moral abilities develop at different rates with respect to different spheres of human life. Each person is not virtuous, continent, incontinent or vicious overall, but rather he or she is virtuous with respect to some spheres of life, continent with respect to other spheres, and incontinent and/or vicious with respect to yet other spheres. (Curzer, 2016: 121).

Curzer reminds us that overly simplistic models are likely to be flawed due to the unquantifiable number of components or what he calls ‘centi-virtues’ that are associated with growing in virtue (Curzer, 2016). To put these ideas into context, we need to be mindful of where the Kohlbergian stage-theory project failed. It was undone partly by its methodology, which equated an individual’s moral maturity with the ability to offer solutions to far-fetched dilemmas, and partly by its ambition of supposing that students would act morally as a simple consequence of knowing how to act morally, or even of just knowing how to articulate convincing moral judgements. In 1980 Augusto Blasi published a much-cited meta-analysis of empirical studies that gauged the relation between moral reasoning and moral behaviour. What Blasi found was that moral judgement plays at best a modest role in motivating moral action (Blasi, 1980). Thus, a ‘moral gap’ had been identified between cognition and action that theorists – including Blasi himself – have been trying to bridge ever since.

An alternative approach to character education model, that might be deemed consistent with Curzer(2016), is seeking to define the components or constructs of virtue growth. These are likely to be many and complicated, but some broad pillars that these features can be grouped around might be sought. The attraction of such an approach is that it does not set out to fully map a developmental path for moral growth, but instead details some important places to visit on the way. For example, the popular ‘Four Component Model’ addresses areas that predicate moral behaviour and provides a conceptualisation of successful moral functioning and the capacities it requires (Rest *et al.*, 1999). At the heart of the model are four inter-related abilities that it is believed are required to be a moral individual – these are moral sensitivity, moral judgment, moral motivation, and moral character.

Drawing inspiration from this work, academics and practitioners at the Jubilee Centre for Character and Virtues have been developing a similar model and one that draws heavily on a virtue ethical understandings of character. Attempts at building a model is founded on a belief that if character education is to flourish, it must have some guiding, structural constructs and concepts that give coherence to the overall curriculum that a student experiences (Narvaez and Lapsley, 2008). A key challenge is to build a model that is accessible to teachers, but not so simplistic it undermines the considerable and widely acknowledged complexities of character. The features of the emerging model are clustered around seven inter-related components: virtue perception, virtue knowledge and understanding, virtue emotion, virtue identity, virtue motivation, virtue reasoning, and virtue practice (Jubilee Centre, 2017: 8). Presently, although there is a great deal of research going on, the seven components that make up these pillars as well as the relationship (or the bridge) between these pillars is under-explored. The aspiration of the current work is to add evidence as to how two of the components, virtue perception and virtue reasoning, might be taught. It is not possible, within the limits of the present research, to show how these components might contribute to the broader aim of bridging the infamous Blasi gap.

Conceptions of *virtue* reasoning can be linked to the construct of *moral* reasoning as understood by neo-Kohlbergians (see Rest *et al.*, 1999). Like moral reasoning, virtue reasoning builds on the idea of rational judgments but places the emphasis firmly on character virtues and in particular the broader concept of *phronesis*; the meta-virtue that is at the heart of Aristotelian-inspired virtue ethical philosophy. Virtue reasoning is concerned with the individual considering the interaction between virtues in any given moral dilemma, as much as the end goal of defining an appropriate course of action.

The problem with many previous attempts to enhance virtue reasoning is that they present students with clear-cut dilemmas, pre-identified by the teacher and handed to them on a plate. That is not how dilemmas present themselves in real life. The initial challenge facing the student – even one which has a decent command of virtue literacy – is to identify the situations to which virtue reasoning needs to be applied. We could call this virtue perception, defined as ‘[n]oticing situations involving or standing in need of the virtues’ (Jubilee Centre, 2017: 8). Perceiving virtue in a situation requires the ability to focus attention on ethical aspects of particular situations. Virtue perception is closely linked to moral sensitivity (Rest *et al.* 1999) in having the ability to recognise the morally relevant and problematic features of a situation. For example, a young person who is unsure if downloading a film from the internet is illegal (Curzer, 2014). Virtue reasoning can only take place once the situation has been recognised to have a moral dimension to it. Reasoning involves making a decision based on virtue and being able to justify the decision through defending a particular judgment and

course of action. It must be reflective to allow for the empowerment of the ethical self through autonomous decision making. The ability to educate children and young people to improve their recognition of morally salient situations and critically engage in virtue reasoning that was the focus of the pilot evaluation described presently.

It is hypothesised that those who are able to perceive a situation that stands in need of the virtues coupled with the *ability* to facilitate virtue reasoning are more likely to act virtuously. That is an *empirical* claim which needs to be tested in practise. However, the empirical claim is based on presumed *conceptual* links between understanding, perception, reasoning and motivation, laid out in an Aristotelian theory of virtue.

Intervention to cultivate Virtue Perception and Reasoning

A naturalist position on virtue ethics allows for the theory to be informed and shaped by empirical research (Kristjánsson, 2015). It was with this in mind that a new intervention, entitled *Making Wiser Choices Online*, was developed and piloted. The intervention built on similar studies that demonstrate the possibilities of teaching character through and within curriculum subjects (Author et al 2016). In this instance the intervention was incorporated into the Computer Science programme of study (England), although it was also deemed to be relevant to the PSHE and Citizenship Education curriculums. The intervention, consisting of a taught course structured across four one hour computer science lessons, required students to be both self-reflective about their own Internet use and its impact on others, as well as to imagine new ways of using the Internet.

At the heart of the *Making Wiser Choices Online* programme was twelve moral dilemmas, developed in conversation with 11-14 year olds, that young people might face in their daily lives; these related to concerns such as cyber-bullying, plagiarism, piracy and extremism amongst others. Exposure to dilemmas, through the taught programme, was seen to be a way in which students could be supported to at first notice and then progress towards resolving conflicting moral demands and developing critical reflection. Repeated exposure to dilemmas might be seen as a form of advanced habituation where students are gradually brought to more critical discernment through the practise of *cyber-phronesis* with the guidance of an outside instructor. The advantage of this approach, and its focus on critical reflection, is that moral character education need not be indoctrinating as it is about ‘helping students grasp what is ethically important in situations and to act for the right reasons, such that they become more autonomous and reflective’ (Jubilee Centre, 2017: 2). Examples of two of the dilemmas in the programme are provided below:

Jack is late with his homework. He has been asked to write a short essay about the virtue of courage that is to be entered into a national competition. He does some research online and finds a really good short essay. Although he knows he can't copy it all he knows he will never be able to write anything better. His homework is due in the next morning and it is already late at night. What should Jack do?

Anna is a very loyal to her best friend Rachel. However, Anna realises that Rachel has been sending bullying messages over the Internet to another girl at school. Anna does not really like the other girl, but she also thinks the messages that Rachel has been sending are not kind. One day at school Anna's teacher asks her if she knows who has been sending the nasty messages. What should Anna do?

Alongside other aims, it was hoped that after experiencing the *Making Wiser Choices Online* programme, the pupils would be better equipped to notice when they are in an online situation that demands a virtuous response (virtue perception), Having achieved this perception they would then be able to determine a course of action that took into account virtues such as compassion, courage and honesty (virtue reasoning). Consequently, of the two concepts this intervention sought to enhance, virtue perception was seen to be prior to virtue reasoning. This meant that the evaluation of the intervention was interested in exploring the extent to which pupils developed these two concepts both independently and within a combined measure.

Methods

The four-lesson intervention, entitled *Making Wiser Choices Online*, was subjected to a pilot evaluation. A pilot is appropriate for the evaluation given; i) well known concerns about conducting experiments in education (Thomas, 2016); ii) challenges associated with ‘measuring’ character (Author et al, 2016); and, iii) the relative immaturity of the intervention. Brown-Urban *et. al.* (2014) describe four evolutionally stages of intervention evaluation. The *Making Wiser Choices Online* programme was deemed to be at the ‘development stage’ as the programme is still undergoing changes or revisions. The authors recommend a ‘change’ evaluation which:

‘generally examines a program’s association with change in outcomes for participants in a limited and specific context (the focus is not yet necessarily on generalizability to other contexts, settings, etc.). Evaluations in this phase are generally correlational studies that use either matched or unmatched pre- and post-tests. This phase also generally includes greater focus on verifying the reliability and validity of measures’ (Brown-Urban *et al*, 2014 : 132).

The pilot evaluation sought evidence that helped to answer the following research questions: i) to what extent does the programme enable 11-14 year olds to perceive issues relating to moral virtues on the Internet; and, ii) to what extent does the programme enhance the virtue reasoning of 11-14 year olds? If the pilot produces statistically significant findings, and the measures were deemed to be robust and valid, then it is anticipated the programme would be put forward for a larger experimental trial including randomisation.

Pilot Evaluation

Three schools in England delivered the *Making Wiser Choices Online* intervention between February and April 2017 and in total 527 pupils completed both pre- and post-tests. In each school, year groups were divided between classes who received the lessons as part of the *Making Wise Choices Online* programme, the ‘intervention’ group (n=262) and those who did not participate in the programme and were in the ‘control’ group (n=265).

The intervention and control groups were from the same school for two reasons: i) within-school matching increases the likelihood of the control group having a similar profile to that of the intervention group and so minimises ‘imbalance across treatment groups’ (Campbell *et al.*, 2004: 705); and, ii) it is difficult to recruit schools to provide purely control groups, as there is no immediate benefit for them. The measurement instrument sought to test the ability of pupils to perceive the moral salience of situations and assess pupils’ ability to undertake virtue reasoning. The instrument consisted of a written dilemma related to social media use and asked pupils to respond to three questions. The first asked students to explain the dilemma and sought to assess pupils’ perception of the moral situation and how virtue

knowledge and understanding could be applied within the dilemma. The second asked pupils to explain what the protagonist should do in response to the scenario. This question was not assessed but was incorporated as an initial step in order to encourage pupils to consider why the recommended course of action should be taken. Pupils were asked to provide three reasons for why the protagonist should act in a particular way and to identify which reason was the most appropriate and this question was intended to assess pupils' virtue reasoning. This focus on attempting to assess virtue reasoning addresses concerns from some quarters regarding the believed behaviouristic nature of character education, in which it is suggested that character education seeks to promote a limited view of what it is to be a moral person. Virtue reasoning is dependent on critical reflection and the development of tools for the individual to ascertain the most appropriate course of action in a given situation. Consequently, while there is a general assumption at work on what it is to be a moral person, the way in which this is demonstrated will be dependent on the specific context in which the individual finds him or herself.

Two versions of the instrument were created so that they could be administered at time 1, before the start of the intervention and at time 2, after pupils had participated in the *Making Wiser Choices Online* lessons. The two instruments (referred to as test A and test B) were designed to be of equal style, length and intended difficulty, this was assessed through considering the Flesch Reading Ease Score, which at 78.8 for Test A and 75.7 for Test B, was identified to have a Flesch-Kincaid grade level of 5.2 and 5.7 respectively and so was seen to be appropriately targeted to the expected reading level of a child aged between 10 and 11 years old. The suitability of the test were further established through undertaking a pilot with pupils aged between 12 and 13 years in two schools. A sample of pupils' responses within 41 Test A and 40 Test B papers were considered by the researchers and was used to refine the mark scheme and also determine prototype answers. In addition, feedback from teaching staff on the suitability of the intervention materials was also received and amendments were made based on these comments.

The allocation of Tests A or B to either time 1 or 2 was randomised on a school basis, according to whether a class was in the control or intervention group. Pupils completed the instruments independently and in silence. In order to control for pupils' literacy levels, teaching staff were asked to indicate whether their pupils' literacy level were considered in the top, middle or lower third of the class. While this request would necessitate the class being split into thirds, analysis indicated that teachers assessed pupils individually as to whether they perceived them to have higher, middle or lower levels of literacy which led to an uneven distribution of literacy levels. However, arguably, this would reflect a more accurate assessment of literacy levels rather than imposing a grouping on a class of students. Pupils were also asked to identify their gender to allow for the effects of this variable to be considered as part of the analysis (table 1 provides a breakdown of pupils by school, age and gender).

Table 1: Pupils by school, age, and gender

Gender		School Year		
Female	Male	Year 7	Year 8	Year 9
Count	Count	Count	Count	Count

School 1	69	69	138	0	0
2	46	65	111	0	0
3	141	137	0	149	129
Total	256	271	249	149	129

All teachers were provided with a pack of materials that included instructions regarding the delivery of the instrument. Contact was also maintained with the teachers charged with implementing the intervention in order to provide clear instructions and clarification as required. Nevertheless, despite these safeguards, it was recognised that there may be variation in the delivery of the intervention.

To assess pupils' responses to the instrument, a scoring rubric was devised and revised after piloting. Although the rubric relied on some interpretation by the markers, prototype answers drawn from the pilot were included to improve the reliability of the measure. The scale used within the marking of each of the two questions within the tests was based on a limited number of categories (from a score of zero if the pupil did not attempt a response to the question to a score of 4 if the pupil demonstrated an answer that was compatible with the top score within the rubric) in order to increase the reliability of the measure. The attribution of scores was dependent on how the pupils expressed the identification of virtues, along with their reasoning as to why the protagonist should act in a particular way. To further support the reliability of the measure, a conservative approach was taken with regards to the marking of the test papers which meant that pupils had to express the virtue dimension of the dilemma, rather than focussing their responses on the consequences of the protagonist taking (or not taking) a particular course of action.

The pre- and post-tests were marked by the research team, with a quarter (n=263) of the 1054 tests second marked. Two-way mixed average measures with absolute agreement were used to assess inter-rater reliability (IRR) by test A and test B across the two questions and the overall total score (table 2). The intra-class correlation varied across both tests for Q1 and Q2b and this could perhaps be accounted for by the need for further interpretation of how a student has expressed the reasoning behind their stated actions within Q2b. It was felt that the measures of reliability in the scoring were robust enough to proceed with marking the tests according to the rubric.

Table 2: Intraclass correlation Coefficients (ICC)

	N	Q1	Q2b	Total Score
Test A	88	0.86	0.72	0.79
Test B	175	0.96	0.87	0.92

Analysis

As has already been noted, the rubric was structured to employ a limited scale as this would allow for greater consistency in the marking of the test papers, thus increasing the reliability of the measure. The requirement for a reliable measure meant that the scale that was used to

mark both questions 1 (identifying virtue perception) and 2b (identifying virtue reasoning) was restricted to an ordinal variable (with scores ranging from 0-4 for both questions). As has already been indicated, the scores for the two questions were then combined in order to create a rounded measure of virtue perception and virtue reasoning and this created a continuous scale (with scores ranging from 0-8). The justification for combining the scores was that in order to be able to give virtuous reasons for undertaking a particular course of action, a person must first perceive the situation to stand in the need of a virtuous response.

The scores were combined in order to respond to theory as discussed above that virtue perception is a necessary pre-condition of virtue reasoning. Consequently, it was felt to be an interesting area of analysis to consider the combined scores. Those pupils who did not attempt either question received an overall score of zero, these data were removed from analysis as, due to the tests being administered by the schools, there was no way of knowing whether the pupils did not understand what was asked of them, or whether they decided not to engage with the test. In addition, pupils who only completed a pre or post-test, or for whom the research team did not have data on their gender or teacher-assessed reading and writing level, were also removed from the dataset. All statistical analysis was completed using SPSS 22.

Results

Virtue Perception, Virtue Reasoning and Combined Scores

Descriptive analysis of the difference between scores within the control and the intervention groups demonstrate that the mean score for the intervention group increased across the three measures (virtue perception, virtue reasoning and overall total score), as demonstrated throughout figures 1, 2 and 3. Across the overall total score for those pupils who received the intervention, there was an increase of 7 percentage points (figure 3), which was comprised of a 5 percentage point increase for measures of virtue perception (figure 1) and an 8 percentage point increase for measures of virtue reasoning (figure 2). At a preliminary level of analysis, this suggests the intervention supported the expression of both virtue perception and reasoning in that, on average, students achieved a higher score after they had received the intervention.

Figure 1: Measure of identification of virtue perception

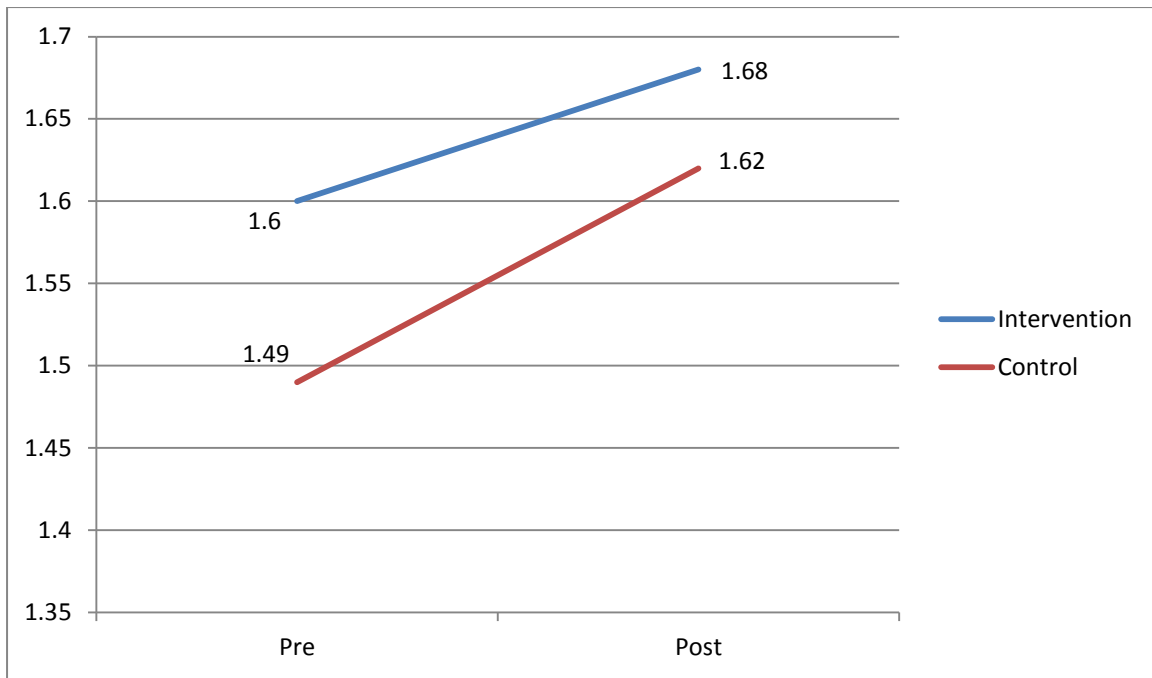


Figure 2: Measure of virtue reasoning

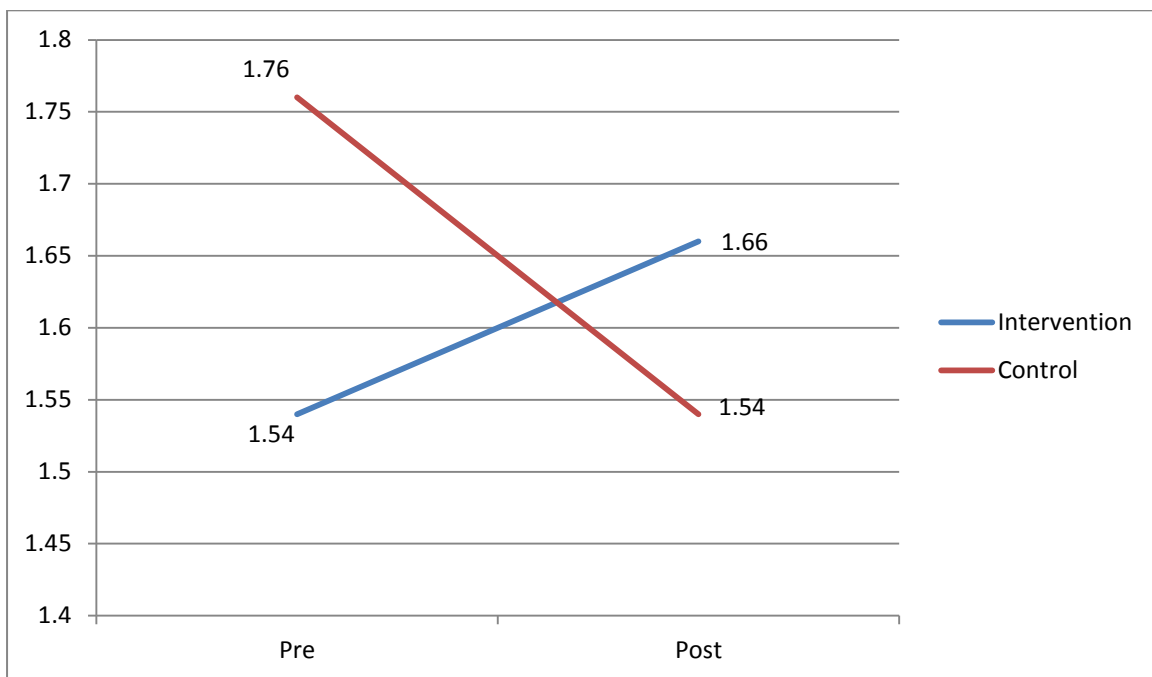
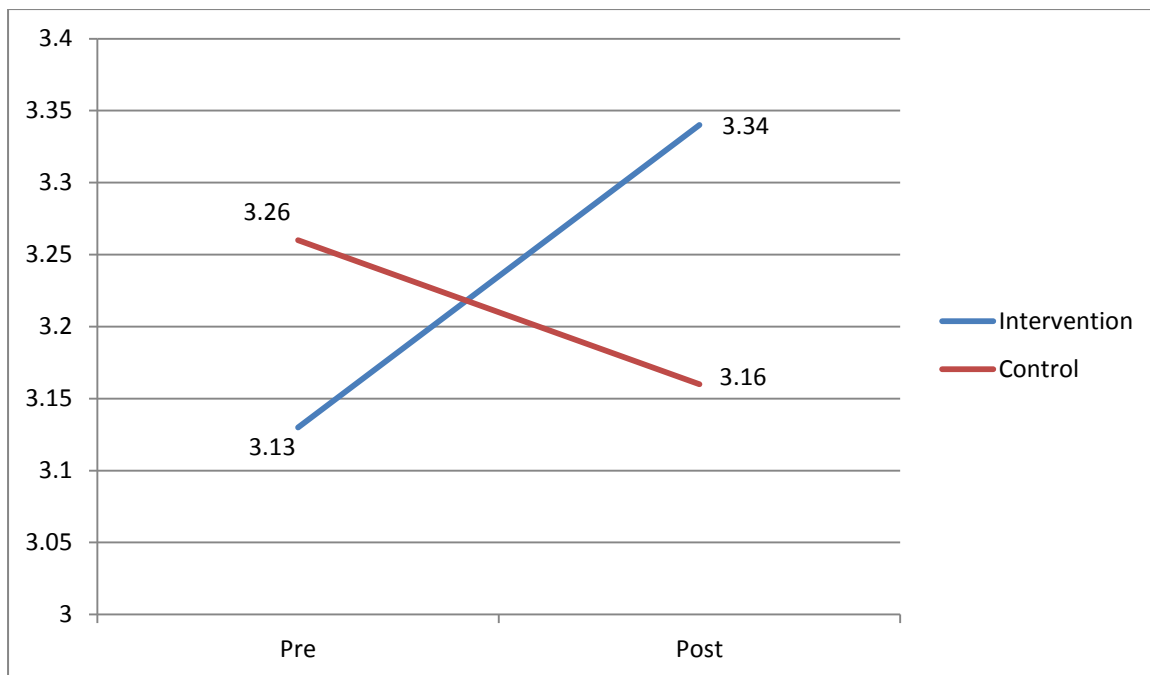


Figure 3: Total measure of virtue perception and virtue reasoning



In contrast, when considering the results for the control group (figures 1, 2 and 3), it can be seen that there was an increase in the scores for the question which identified the perception of virtues (a 9 percentage point increase), whereas there is a decrease within the measures of virtue reasoning (a 13 percentage point decrease). Overall, there is a 3 percentage point decrease in the total overall measure which could be suggested to indicate that increases in pupils' scores are not solely the result of maturation.

On the basis of the positive descriptive analysis results, a more statistically advanced paired-sample t-test was conducted to evaluate the impact of the intervention on pupils' overall scores (measuring both virtue perception and virtue reasoning) on the dilemma test. For those pupils who received the intervention there was a statistically significant increase in the post-test score ($M = 3.34$, $SD = 1.38$) compared to the pre-test score ($M = 3.13$, $SD = 1.27$), $t(261) = -2.25$, $p = .03$ (two-tailed). In contrast, the control group indicated a decrease between the pre ($M = 3.26$, $SD = 1.40$) and post-test scores ($M = 3.16$, $SD = 1.30$) which was not statistically significant, $t(264) = 0.95$, $p = .344$ (two-tailed).

Influence of gender

On the basis of the positive descriptive analysis results, analysis was furthered through undertaking a paired sample t-test and splitting the sample by gender. When the sample was split by gender, across the total measure, it can be seen that within both the pre- and post-scores, females who received the intervention achieved higher scores compared to male pupils who also received the intervention (figures 4, 5 and 6).

Figure 4: Virtue perception by gender

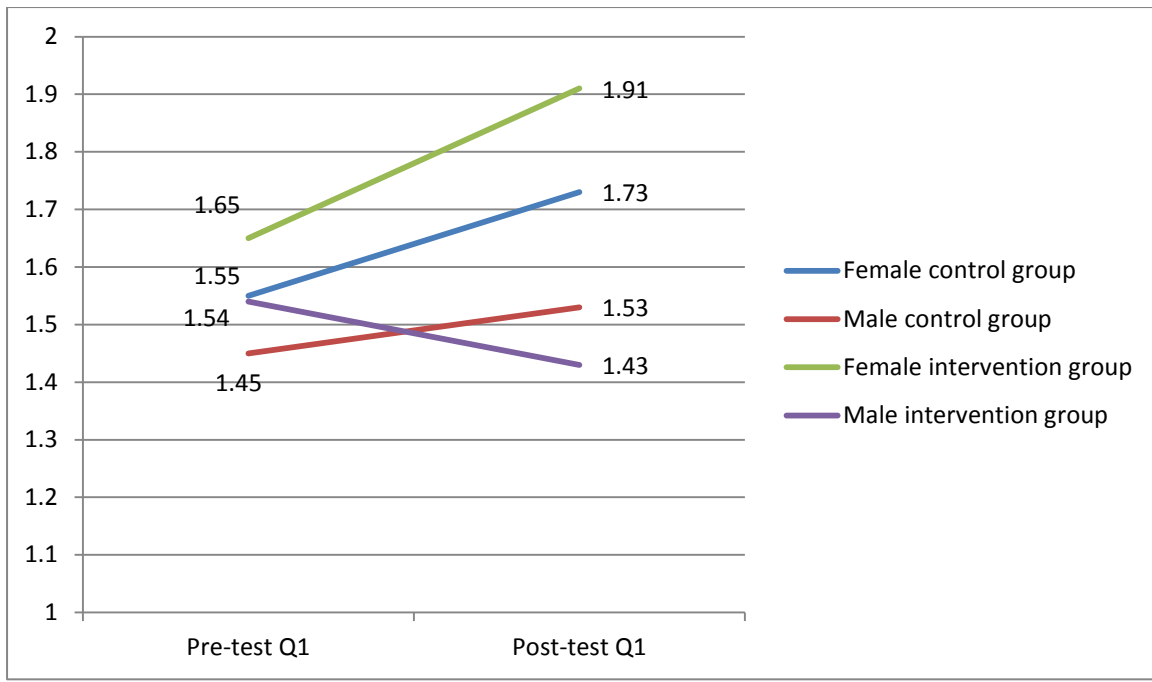


Figure 5: Virtue reasoning by gender

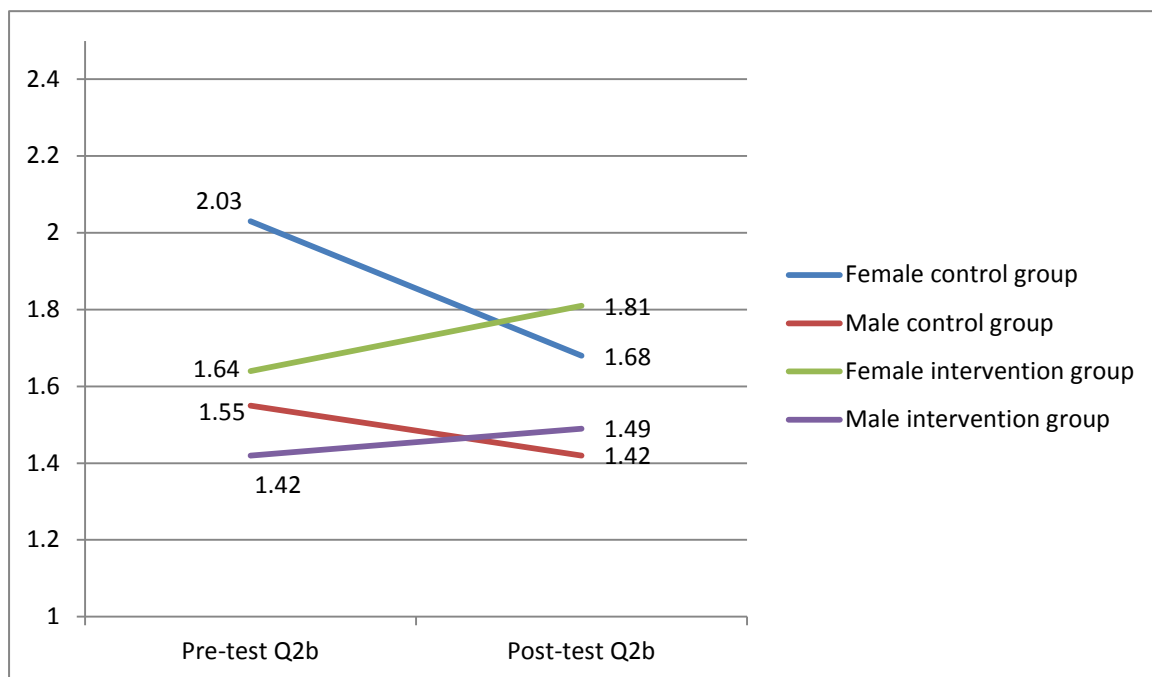
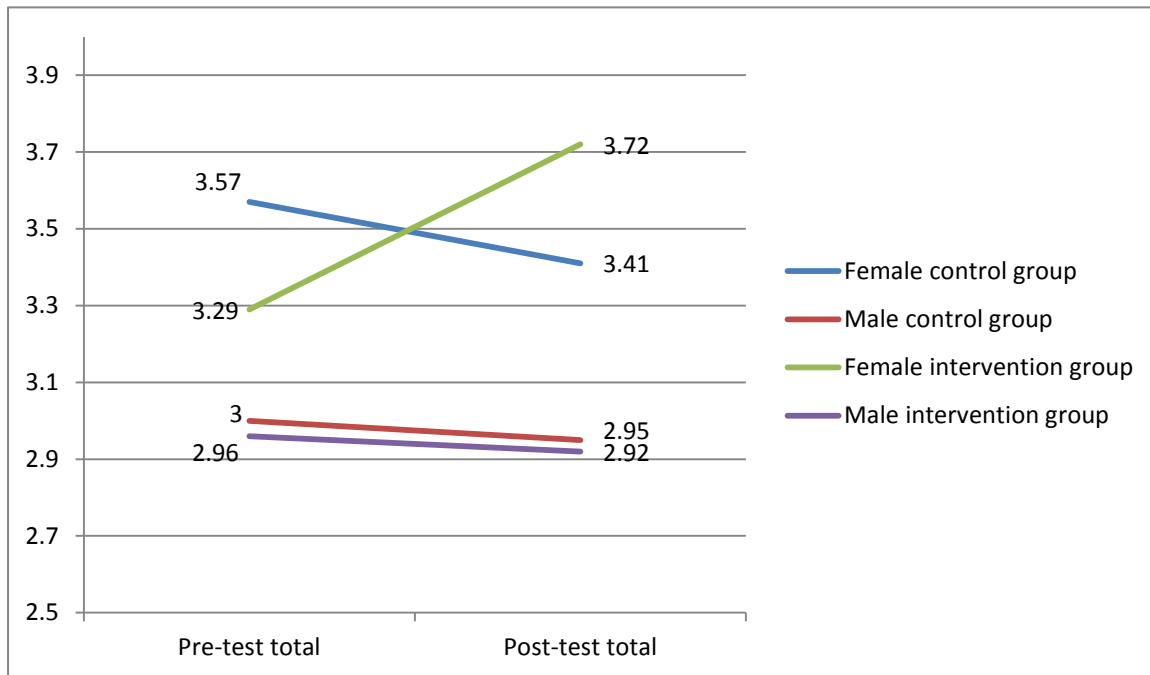


Figure 6: Total score by gender



It is notable when considering the results of female pupils who received the intervention there was a statistically significant increase in the post-test score ($M = 3.72$, $SD = 1.42$) compared to the pre-test score ($M = 3.29$, $SD = 1.26$), $t(136) = -3.14$, $p = .002$ (two-tailed). In contrast, male pupils demonstrated a small reduction when the pre-tests ($M = 2.96$, $SD = 1.26$) are compared to the post-tests ($M = 2.92$, $SD = 1.20$), $t(124) = 0.351$, $p = .726$ (two-tailed), however, this was not statistically significant. This result again be suggestive of the greater engagement of female pupils with the intervention compared to males or, as previously considered, may be indicative of the way in which the evaluation instruments favoured female pupils.

Controlling for literacy levels

The evaluation of the *Making Wiser Choices Online* intervention was grounded in the ability of individual pupils to perceive how the presented dilemma required the expression of virtues, along with providing justification as to why the protagonist should act in the way they have suggested. As this was assessed through considering pupils' written responses to the posed questions, it was felt that there could be the opportunity for pupils' responses to be affected by their aptitude for reading and writing in that their response is very much dependent on the ability to express their thoughts through writing. As identified, it was felt necessary to collect data on pupils' literacy levels in order to control for any potential effects of this on the post-intervention scores. The most appropriate parametric test to undertake this analysis was the ANCOVA statistical test, however, on examination, it was identified that the dependent variable (overall post-test score) data violated an assumption of the ANCOVA test as it had a skewed distribution. Despite this concern, it was decided to follow previous literature (Olejnik and Algina, 1984) and to run the ANCOVA as a secondary test, following the use of paired sample t-tests. In order to control for pupils' literacy levels, this variable was included as a covariate as well as the pre-test scores in the ANCOVA. In doing so, we found that there was a significant effect of the intervention on the post score obtained by the

pupils who were part of the intervention after controlling for literacy level and pre-test scores, $F(1, 523) = 3.68, p=0.05$. This followed the finding of Davison *et al.* (2016) which explored the potential conflation between virtue literacy and general reading skills and identified that improvement in measures of pupils' application of virtue concepts in personal, social and cultural contexts was most likely to be independent of any improvement in reading level. Despite the statistically significant result and the indication that pupils' literacy levels have not been conflated with measures of virtue literacy, there is a need to be cautious when analysing this finding due to the violation of the assumptions of ANCOVA. However, such results should be read as providing further evidence for the effects of the intervention and should allay concerns regarding the effects of pupils' teacher-assessed literacy levels on the post-test scores, and the subsequent analysis of the effectiveness of the *Making Wiser Choices Online* intervention.

Limitations

There are several principle limitations to the pilot evaluation; primarily methodological and practical in nature. Firstly, the trial involved 527 pupils, while this is a relatively large sample, these were purposely sampled from three schools already interested in character education. It is not possible to generalise the results, as the participants cannot be deemed to be representative of the wider population. As the intervention and control groups were not randomly assigned there is a possibility of systematic bias through schools potentially selecting 'better' classes for the intervention group. Also, for practical reasons, both control and intervention groups were in the same schools at the same time meaning contamination of the control group was possible. Furthermore, the dilemmas in the pre- and post-test were very similar so it is possible that completing the pre-test informed pupils' answers within the post-test. Responding to the dilemmas was very much dependent on the way in which the pupil expressed themselves in writing. It may be that future interventions that consider the way in which pupils access the Internet could also consider how other methods which collect a broader range of data could be used so as to allow those pupils who may struggle to express themselves in writing to convey their thoughts. This is a further area for future research to consider.

The schools were in control of many important features of the pilot, especially the length of sessions devoted to the programme, the timings, duration, setting, and conduct of the tests. It is recognised that, while efforts were made to ensure that there was consistency to the delivery of the intervention, there may be some variation as to how each intervention was delivered.

Finally, the issues of measuring character (Kristjánsson, 2015) and the impact of character education interventions (Author, et al., 2016) are well known. Most notably, these are due to the complexity of the construct of character and associated virtues and that measuring how an intervention might develop single or multiple constructs presents considerable methodological challenges (Duckworth and Yeager, 2015).

Ethical considerations

Ethical approval was granted for the design of the *Making Wiser Choices Online* by the University of Birmingham Ethics Committee and participants were given the opportunity to opt-out of the study and for their pre- and post-tests to not be analysed. Each of the schools

involved in the trial confirmed that the control groups would also receive the intervention after the completion of the post-test; this was to ensure that no student was disadvantaged from their allocation into either the control or intervention groups.

Discussion

The results of the trial are encouraging for those interested in the age-old question can character and, more specifically, components of virtues be taught? Although the results are a long way from providing an answer to the question; they do shed some important light on how two components of virtue (perception and reasoning) might be educated through targeted interventions. Whilst there has been significant, mainly philosophical, developments in understandings about the conceptual nature of the components of character and virtue (see, for example, Rest *et al.*, 1997; Jubilee Centre, 2017), empirical evidence about how these concepts might be cultivated in children and young people through education has been more scarce. The present results, add some new evidence, about how virtue perception and reasoning might be taught to 11-14 year olds in the UK. The picture as to how virtues can be cultivated becomes slightly clearer if the current findings are read in combination with other studies, for example, those showing how virtue knowledge and understanding can be developed (Arthur *et al.*; 2015) and how virtue emotion (Ornaghi, Brockmeier, & Grazzani, 2014) might be taught. If we consider virtues, as Curzer (2016) suggests, to be made up of many parts, it is important to consider these separate components in turn in order to address the much bigger concern about appropriate strategies for the education of virtue.

Given that the study was a evaluation of an intervention in its developmental stage, alternative explanations for the results cannot be ruled out. Firstly, the pre- and post- tests (as well as the programme itself) were likely subject to ‘demand characteristics’ (Orne *et al.*, 2000) in that the participants tried to work out the aim of the study and answer in ways to either support those aims or undermine them. A particular issue with studies on character is a social desirability effect often associated with research methods that rely on self report, where participants strive to give answers that are deemed to be ‘acceptable’. Secondly, given that the intervention was an isolated curriculum intervention it is not possible to determine if the participants carry any enhancement in their virtue perception and reasoning into their wider online lives. To determine this would most likely demand a more substantial longitudinal, study.

Further, there remains the significant problem, much as there has been with previous moral development models (Rest *et al.* 1999), of gaining evidence as to how the components relate to each other as well as, most importantly, how they contribute to the overall goal of virtue practice and action. Narvaez (2002), who was also involved with the development of the four-component model, has previously challenged assumptions that children build moral literacy from reading or hearing moral stories. However, if we understand that character requires autonomous and critical reflection on the part of individuals; and to do this requires the ability to both be able to notice an issue that stands in need of virtues, and to also justify a subsequent course of action that draws on the virtues, then it is possible to see how the enhancement of virtue perception and reasoning might help address the infamous ‘knowing and doing’ gap.

Alongside the results, significance comes from the novel use of the methodology itself. Given that concerns about the measurement of character are well-known (Kristjánsson, 2015), advancements in this area are important. Although trials into character education

interventions are becoming increasingly commonplace (see, for example Education Endowment Foundation, n.d.); challenges to validity are as equally common (see for example, Arthur *et al.*, 2014). In using dilemmas to investigate virtue reasoning, the present research has utilised an established methodology, (Thoma, 2014; Author *et al.*, Morris and Ryan, 2016) which has been prevalent since the work of Kohlberg (1981). Dilemmas have often been used in research that considers how particular populations, for example: teachers (Arthur *et al.*, 2015), children and young people (Walker, Curren and Jones, 2016), and various professions (Bebeau and Thoma, 2013), respond to moral and ethical dilemmas. The present research utilised ethical dilemmas, found online, as the basis of a before and after test to evaluate a character education intervention. The significant results suggest that such an approach is a valid and useful instrument and helps, somewhat, to counter the ‘serious challenges’ (Duckworth and Yeager, 2015: 1) of measuring character. However, one of the main critiques of dilemmas remains, in that they show what a person might theoretically do – but not necessarily how they would respond in a ‘real life’ situation..

It was important within the trial to control for literacy levels (Davison *et al.*, 2016). The research team were keen to address the possible conflation of general and virtue-specific literacy. The ANCOVA results suggest that literacy levels were not a factor in the results, however, given the complexity of running trials of this nature, it cannot be discounted that higher test scores may in fact be an artefact of the way in which the intervention was evaluated, and the emphasis placed on pupils’ ability to express their thoughts in writing. This may also account for the female participants’ higher scores. Research has demonstrated that, at age 11, a higher proportion of male students have reading attainment levels lower than expected compared to female students (Moss and Washbrook, 2016). The extent to which the evaluation, and potentially also the intervention, could have favoured female pupils requires further exploration. As also seen in the *Character Education in UK Schools* study (Arthur *et al.*, 2015) female pupils also scored higher in the Intermediate Concept Measure for Adolescents (Ad-ICM) when compared to male pupils. Potential explanations as to why this may be the case are considered by Walker *et al.* (2017) and it is particularly noted that “A degree of social-desirability bias should not be ruled out” (p. 603) in that it may be the case that female pupils, through the process of assessing their virtue perception and virtue reasoning, may be more likely to identify choices which were coherent to their perceptions of adults’ preferences. Nevertheless, Walker *et al.* (2017) indicate that a female pupil whose response is affected by social desirability bias would still be required to demonstrate ‘appropriate moral judgement’ (p. 603). This is suggestive of the need to explore further the foundations on which the presumed greater expression of social desirability from females within this trial is built. Such an appeal for additional research is further complicated by the fact that the differences in scores between male and female pupils could be potentially attributed to greater engagement with the intervention, or may be a result of the way in which this engagement has been measured through the pre- and post-tests.

Conclusion

This paper has been based on the argument that human qualities and virtues, such as compassion, courage and integrity, are generally admired. It has also been based on the argument that, in some form, education has always sought to cultivate these virtues in children and young people – although not always explicitly. More explicit attempts might be called character education, an approach that has been promoted most recently by individuals and organisations from across the political spectrum. Arguments about what form this character education should take have been presented and the paper has sought to address this issue through an in-depth focus on a single intervention.

Evidence from the pilot evaluation of the *Making Wiser Choices Online* programme with (527) 11-14 year olds is that it might be possible to teach virtue perception and reasoning through deliberate educational efforts. Further, it would appear that males and females may respond differently to either the intervention, or the way in which it is evaluated. This finding appears to mirror prior literature which has considered the differences in male and female pupils responses to moral dilemmas and is an area which is ripe for further study. Most encouraging, despite the limitations of the pilot, was that advancements were seen over the duration of a relatively short programme, targeted at a particular age group. Deliberate efforts that span a child and young person's education will likely have more substantial effects. These findings are reported tentatively given the relative immaturity of the intervention and the limitations associated with evaluating character education interventions which might suggest alternative reasons for the results.

Whilst the paper adds to the evidence in support of claims that two important elements of virtue, perception and reasoning, might be cultivated through deliberate educational efforts; it leaves many questions still to be answered. Most notably is how these aforementioned pillars of virtue can be bridged together and contribute to an individual actually carrying out virtuous actions as a matter of course. The paper concludes with the assertion, that if these gaps are to be ameliorated, then philosophers, psychologists and educationalists with an interest in the area should work together on this challenge.

References

Annisette, L. E. and Lafreniere, K. D. (2017) 'Social media, texting, and personality: A Test of the Shallowing Hypothesis', *Personality and Individual Differences*, Vol. 115, no. 1 pp. 154-157.

Aristotle (2009) *The Nichomachean Ethics*. Translated by R. David, Oxford: Oxford University Press.

Arthur, J. (2003) *Education with Character: The Moral Economy of Schooling*. London: RoutledgeFalmer.

Blasi, A. (1980) 'Bridging Moral Cognition and Moral Action: A Critical Review of the Literature', *Psychological Bulletin*, Vol. 88, no. 1, pp. 1-45.

Brown-Urban, J., Hargraves, M., Trochim, M. (2014) 'Evolutionary Evaluation: Implications for evaluators, researchers, practitioners, funders and the evidence-based program mandate', *Evaluation and Program Planning*, vol. 45, pp. 127-139

Bebeau, M. J. & Thoma, S. J. (2013) 'Moral Motivation in Different Professions', in Heinricks and Oser (Eds.) *The Handbook on Moral Motivation*. Sense Publishers, Rotterdam, The Netherlands.

Campbell, M. K., Elbourne, D. R. and Altman D. G. (2004). 'CONSORT Statement: Extension to Cluster Randomised Trials', *British Medical Journal*, Vol. 328, no. (7), pp. 702-708.

Carr, D. (2008) 'Virtue Ethics and the Influence of Aristotle' in *Contemporary Philosophical and Psychological Perspectives on Moral Development and Education* (ed.). Fasko, D. & Willis, W., Cresskill, NJ: Hampton Press.

Curzer, H. (2016) 'Rules Lurking at the Heart of Aristotle's Virtue Ethics', *Aperion*, Vol. 49, no. 1, pp. 57-92.

Davison, I, Harrison, T, Hayes, D & Higgins, J. (2016) How to assess children's virtue literacy: methodological lessons learnt from the Knightly Virtues programme *Journal of Beliefs and Values*. DOI: [10.1080/13617672.2016.1141526](https://doi.org/10.1080/13617672.2016.1141526)

Department for Education (2016) *DfE Strategy 2015-2020, World-Class Education and Care*, [Online], Available at: www.gov.uk/government/uploads/system/uploads/attachment_data/file/508421/DfE-strategy-narrative.pdf [Accessed: 21 April 2017].

Department for Education (2017) *Developing Character Skills in Schools: Summary Report*, NatCen Social Research [Online], Available at: www.gov.uk/government/uploads/system/uploads/attachment_data/file/634710/Developing_Character_skills-synthesis_report.pdf [Accessed: 07 August 2017].

Duckworth, A. L., and Yeager, D. S..(2015) 'Measurement Matters: Assessing Personal Qualities Other than Cognitive Ability for Educational Purposes' *Educational Researcher*, Vol. 44, no. 4, pp. 237–251.doi:10.3102/0013189x15584327.

Ecclestone, K., (2012) 'From Emotional and Psychological Well-Being to Character Education: Challenging Policy Discourses of Behavioural Science and 'Vulnerability'', *Research Papers in Education*, Vol. 27, no. 4, pp. 463– 480.

Education Endowment Foundation (no date) *School Themes, Character*, [Online], Available at: <https://educationendowmentfoundation.org.uk/school-themes/character/> [Accessed: 15^h September 2017].

Harrison, T., Morris, I, Ryan, J. (2016) *Teaching Character in Primary Schools*. London, Sage.

Harrison, T. (2016) Cultivating Cyber-phronesis. *Journal of Pastoral Care in Education*. <http://dx.doi.org/10.1080/02643944.2016.1202307>

Harrison, T. (2016) Virtuous reality: moral theory and research into cyber-bullying. *Ethics and Information Technology*, 17(4), 275-283

Jubilee Centre for Character and Virtues (2017) *A Framework for Character Education in Schools*. Birmingham: University of Birmingham, Jubilee Centre for Character and Virtues, [Online], Available at: <http://jubileecentre.ac.uk/userfiles/jubileecentre/pdf/character-education/Framework%20for%20Character%20Education.pdf> [Accessed: 21 June 2017].

Kohlberg, L. (1981) *The Philosophy of Moral Development: Moral Stages and the Idea of Justice*. San Francisco, CA: Harper & Row.

Kristjánsson, K. (2013) 'Ten Myths About Character, Virtue and Virtue Education – Plus Three Well-Founded Misgivings', *British Journal of Educational Studies*, Vol 61, no. 3, pp. 269-287.

Kristjánsson, K. (2015) *Aristotelian Character Education*. London: Routledge.

McMahon, J. M. and Cohen, R. (2009) 'Lost in cyberspace: ethical decision making in the online environment', *Ethics and Information Technology*, Vol. 11, no. 1, pp. 1-17.

Morgan, N. (2017) *Taught Not Caught: Educating for 21st Century Character*, Woodbridge: John Catt Educational Ltd.

Moss, G, and Washbrook, L (2016) Understanding the Gender Gap in Literacy and Language Development. Bristol; University of Bristol. Available from <https://www.bristol.ac.uk/media-library/sites/education/documents/bristol-working-papers-in-education/Understanding%20the%20Gender%20Gap%20working%20paper.pdf> [Accessed: 21 June 2017].

Narvaez, D. (2002). Does Reading Moral Stories Build Character? *Educational Psychology Review* 14(2), 155-171.

Narvaez, D. and Lapsley, D. (2008) 'Teaching Moral Character; Two alternatives for Teacher Education', *The Teacher Educator*, Vol. 43, no. 2, pp. 156 – 172.

Olejnik, S. F. and Algina, J. (1984). 'Parametric ANCOVA and the Rank Transform ANCOVA when the Data are Conditionally Non-Normal and Heteroscedastic'. , Vol. 9, no. 2, pp. 129-149.

Ornaghi, V. Brockmeier, J. and Grazzan, I. (2014) 'Enhancing Social Cognition by Training Children in Emotion Understanding: A Primary School Study'. *Journal of Experimental Child Psychology* Mar;119:26-39.

Orne. M.T., Whitehouse, W.G., Kazdin, A.E. (2000) Demand characteristics. *Encyclopaedia of Psychology*. Washington, D.C.: American Psychological Association and Oxford University Press. pp. 469–470.

Peterson, C. and Seligman, M. E. P. (2004) *Character Strengths and Virtues: A Handbook and Classification*. Oxford: Oxford University Press.

Rest, J., Thoma, S., and Edwards, L. (1997) 'Designing and Validating A Measure of Moral Judgment: Stage Preference and Stage Consistency Approaches', *Journal of Educational Psychology*, Vol. 89, no. 1, pp. 5-28.

Rest, J., Narvaez, D., Bebeau, M. and Thoma, S. J. (1999) *Postconventional Moral Thinking: A Neo-Kohlbergian Approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Suissa, J. (2015) 'Character Education and the Disappearance of the Political', *Ethics and Education*, Vol. 10, no. 1, pp. 105-117.

Thomas, G. (2016) After the Gold Rush: Questioning the “Gold Standard” and Reappraising the Status of Experiment and Randomized Controlled Trials in Education. *Harvard Educational Review*: Vol. 86, No. 3, pp. 390-411.

Thoma, S. (2014) ‘Measuring Moral Thinking from a Neo-Kohlbergian Perspective’, *Theory and Research in Education*, vol. 12, no. 3, pp. 347-365.

Walker, D., Thoma, S. J., Jones, C. and Kristjánsson, K. (2017) ‘Adolescent Moral Judgement: A Study of UK Secondary School Pupils’, *British Educational Research Journal*, Vol. 43, no. 3, pp. 588-607.