Adolescent Moral Judgement: A Study of United Kingdom Secondary School Pupils

David Ian Walker, Jubilee Centre for Character and Virtues, University of Birmingham

(corresponding author: d.i.walker@bham.ac.uk, address as below)

Stephen J Thoma, University of Alabama/ Jubilee Centre for Character and Virtues

Chantel Jones, University of Oxford

Kristján Kristjánsson, Jubilee Centre for Character and Virtues, University of Birmingham

Work carried out at: Jubilee Centre for Character and Virtues, University of Birmingham, Muirhead Tower, Birmingham, UK

Abstract: Despite a recent world-wide upsurge of academic interest in moral and character education, little is known about pupils' character development in schools, especially in the UK context. The authors used a version of the Intermediate Concept Measure for Adolescents, involving dilemmas, to assess an important component of character – moral judgement – among 4,053 pupils aged 14–15. Data were generated in 31 UK schools of varying types between February 2013 and June 2014. Results showed that compared to USA samples, the pupils' scores were, on average, low, suggestive of tendencies towards 'self-interest', 'not getting involved' and 'conformity/loyalty to friends'. Judgements varied by subscales assessing 'action' and 'justification' choices; pupils more successfully identified good actions than good justifications, but generally struggled more to successfully identify poor actions and poor justifications. Highest scores were for a dilemma emphasising 'self-discipline' and lowest for 'honesty', with 'courage' in-between. Overall average results were significantly and positively associated with being female, having (and practising) a religion, and doing specific extra-curricular activities. Differences in schools were also noted, although the kinds of schools (e.g., public/private; religious/secular) were unrelated to student scores.

Keywords: adolescence; moral; ethical; character; UK schools

Introduction

The idea that children ought to be equipped by their educational experiences to lead flourishing lives has been gathering momentum among academics and educationalists across the world (Walker, Roberts, & Kristjánsson, 2015). While this is also the aspiration of countless teachers who intuitively want to develop the 'whole child' (Arthur, Kristjánsson, Walker, Sanderse, & Jones, 2015; Sanderse, Walker, & Jones, 2015), there remains in many countries a pre-occupation with academic attainment, where teachers are overly concerned with technical tasks and 'pre-determined output' in the shape of exam scores (cf. Biesta & Miedema, 2002; Exley & Ball, 2014). When character educationalists urge schools to help pupils pass exams *and* become good people, they are calling for a balance between cultivating pupils morally and guiding them towards educational attainment (Berkowitz, 2012a). A growing number of research studies have highlighted a positive relationship between various strengths of character and higher academic attainment (Benninga, Berkowirtz, Kuehn, & Smith, 2003; Caprara, Barbaranelli, Pastorelli, Bandura, & Zimbardo, 2000; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Park & Peterson, 2006; Snyder, Vuchinich, Acock, Washburn, & Flay, 2012; Weber & Ruch, 2012), but all too often this relationship is discussed in mere instrumentalist terms, especially regarding amoral strengths such as 'grit' and 'resilience' (Penn Resiliency Project, 2015; Seligman, 2011; Tough, 2013). These sorts of performance virtues are frequently commended in the name of academic attainment, obscuring the view that a balanced character may be worthy of development for its own intrinsic moral worth.

Conceptions of character from a virtue-ethics framework are based on the principle that an objective notion of human flourishing is possible and that its attainment depends on the possession of distinctively human *virtues*: moral, civic and intellectual as well as performative. 'Virtues' are here understood as settled states of character, concerned with praiseworthy conduct in significant and distinguishable spheres of human life (Kristjánsson, 2015). Each virtue typically comprises a unique set of emotion, reason, attention and conduct. These learned qualities, constitutive of flourishing, are considered by many recent educationalists to be the ultimate ends of the education system (White,

2011; Walker et al., 2015), and the cultivation of this dispositional conception of character – by a range of direct and indirect means – is typically seen as the task of character education. However, the multifaceted expressions of character in perception, feeling, thought and action, appropriate to specific situations and the person involved (e.g. finding the 'golden mean'), present immense assessment and 'measurement' difficulties, not yet overcome (cf. Kristjánsson, 2015, Ch 3.).

In our 'age of measurement' (Biesta, 2010), and for character education to gain traction, there is a clear need to understand and assess the current situation in schools pertaining to the development of character and virtue among young people. In addition to 'measurement' challenges, there is often insufficient clarity about what is actually meant by 'character' and its development in educational policy and practice, as well as among academics (Berkowitz, 2012b). Central to the promotion of character is the view that with its development the individual will make better and more informed moral judgements, and so the current study emphasises this aspect of character among adolescents, which is also an area of recent methodological advancement that is capitalised upon in this study.

Adolescence is a time when issues of character become more central to the person and a time when a young person's focus will ideally shift from a preoccupation with the self to a realisation that there is a wider, morally salient community around them (Colby et al., 1987; D. T. Narvaez & Bock, 2002; Rest, Narvaez, Bebeau, & Thoma, 1999). If this shift in emphasis occurs in individuals, then their moral judgments will advance beyond a consideration of themselves and those in their everyday interactions. This desired swing away from the self and the self's interests permits increased influence from the adolescents' social groups such as friends, peers at school, sports teams, churches, teachers and family members, etc.

Attempts to cultivate character in schools often involve a number of different methods such as discrete lessons or paying attention to and improving indirect processes of socialisation through aspects of school culture, including the way teachers interact with each other and with children. Much has been written about the cultivation of character in schools (Arthur, 2010; Berkowitz & Bier, 2006; Durlak et al., 2011; Lickona, 1992; Seider, 2012; Walker et al., 2015), and from this research

we know that it is not enough to have well written plans and an ordered curriculum, nor is it sufficient to teach character in specific lessons alone. Character education needs to take place within supportive, encouraging and informative relationships that reach all features of school life (Seider, 2012). Attempts to educate for character in schools are best informed by a comprehensive assessment of pupils in terms of the cultivation of their characters, but this is beyond existing methodological expertise. The present study aims to contribute an important component towards understanding character among UK adolescent pupils. It uses a recent point of convergence between neo-Kohlbergian and neo-Aristotelian traditions to assess moral judgement and the application of virtue among adolescents. This approach - explained fully below - is used to address specific research questions that are asked of data generated from a large research project in the UK. Set in schools, the study was designed to assess characteristic ways that adolescents in the UK reason about a number of moral dilemmas that are nested within the adolescent experience and which address virtues of honesty, courage and self-discipline.

Relating to UK pupils in Year 10, the following specific research questions were investigated: How do pupils understand and apply virtues in specific contexts, especially in relation to the virtues of honesty, courage and self-discipline? What are their strengths and weaknesses with respect to these? How do adolescent reasoning processes differ between their *action* choices and their *reasons* for acting? How do adolescent reasoning processes differ between the recognition of *best* and *worst* choices? How do results vary in relation to types of school attended? And, finally, how do results vary in relation to important demographic categories?

Initial studies from the US and elsewhere that used the methodology used in this study (Thoma, Derryberry & Crowson, 2013; Fritzhand, 2013) found a consistent pattern of findings indicating that: a) girls achieve higher scores than boys; b) determining the appropriate (*good*) choices and justifications is associated with higher scores than identifying poor choices and justifications; and c) students who "act out" at school achieve lower scores than their peers (who do not). Additional evidence showed that scores are higher for action choices than the related justification scores (Thoma,

Derryberry, & Crowson, 2013). These previous studies, however, are confined to smaller, more specific samples (mostly in a US context), whereas the current study is designed to extend this work using a comprehensive sample in different cultural and policy contexts. More generally, we also know that better responses to moral dilemma tests have been associated with being female (Malti & Buchmann, 2010; Nunner-Winkler, 2007; Pan and Sparks, 2012; Sparks, 2014; Thoma, 1986; Van der Graaf et al., 2014; White, 1999) and it is suspected that having a religion might also translate into an advantage on such tests (Arthur, 2010; Pike, 2010, 2011). The participation in various kinds of extra-curricular activities such as charity work, music, and drama has also been associated with improved moral functioning (Arthur, 2010; Hill, Russell, & Brewis, 2009; Lies, Bock, & Brandenbeger 2012; Adderley, Kennedy, & Berz, 2003; Campbell, Connell, & Beegle, 2007; Carr, 2008; Bouchard, 2002), and there is a widely held belief that sport builds character (Doty, 2006; Shields & Bredemeier, 2008).

In what follows we describe the study design and procedures, before presenting the results firstly in terms of individual responses to the moral dilemmas, and secondly in terms of patterns of response to the dilemmas by school. Overall, the pupils' scores were quite low and these results are discussed at individual and school levels leading to detailed interpretation and discussion. The final section makes recommendations for future character education initiatives as well as suggestions for future research.

Methods

Participants

The participants in the ~~main~~ study were 4,053 Year 10 (England, Wales and one school in Northern Ireland) and S3 (Scotland) pupils, aged 14 and 15. Participants were from 31 secondary schools across the UK (these pupils also completed a self-report survey, which is not discussed in this paper).

Purposive sampling was used across the four nations, in that different types of school were deliberately recruited. These included: state (comprehensive), grammar and independent (covering selective and non-selective); different types of faith-based schools; single-sex and co-educational; rural and city locations; those in affluent and deprived areas, and so on. Researchers were partially

successful in accessing schools for the sample that appeared to have no overt statement or commitment to character education alongside those that did. Schools in London and Northern England were especially difficult to recruit, however, for unknown reasons. It was important to the study design that all pupils in year 10/S3 were surveyed in each school to minimise selection bias, resulting in only a handful of missing students of this age in each school if they were absent when the survey was administered or did not consent to participate. This particular year group was chosen because older pupils, preparing for GCSE examinations, would have more (academic) demands on their time and thus be harder to access, whilst younger pupils would have spent less time within the school setting. Moreover, it was reasonable to expect that most of the young people in that age bracket would be able to cope with reading and responding to the moral dilemmas without support. The participants had a mean age of 14.3 (sd=.53). Demographic differences are described in Table 1 below. An anonymised list of all schools is at: http://www.jubileecentre.ac.uk/userfiles/jubileecentre/pdf/Anonymised_Participating_Schools.pdf

(web link to be added – see file a).

Table 1 - here

Procedures

Ethical approval was granted by the University's Research Ethics Committee, and pupils and their parents gave informed consent to take part in the study. Electronic – or very occasionally paper – surveys comprising Ad-ICM (UK) moral dilemmas were completed between February 2013 and June 2014. There were no differences between electronic and paper versions of the survey in terms of their administration; the difference only existed due to lack of computer facilities in some secondary schools. Furthermore, both versions were very similar in appearance (the online version was simply transformed into a print-friendly PDF file). Audio-delivered surveys were also used with pupils who, for various reasons, preferred this to reading text alone. All participants, grouped by classes or whole year groups, were supervised by researchers and teachers and completed the survey under exam-like

conditions (i.e. quiet, with spaces between pupils where possible). Participants were able to ask questions/seek clarification throughout the process, and they were free to withdraw from the study at any point during their completion of the survey.

Measure

Adolescent Intermediate Concept Measure (UK). This is a reduced version of the full Intermediate Concept Measure for adolescents (Thoma et al., 2013). The full version of the measure includes seven moral dilemmas, each emphasising a particular virtue concept, such as honesty, whereas the UK version included only three. Participants read each dilemma and rated approximately ten action choices *and* ten reasons/justifications on a scale from 1 ('I strongly believe that this is a GOOD choice/reason') to 5 ('I strongly believe that this is a BAD choice/reason'). Following the rating task, participants selected and ranked (first, second and third) best and worst (and second worst) options for actions *and* reasons. A shortened measure was used mostly because the length of time required to complete all seven dilemmas exceeded the time limits imposed so as not to cause excessive disruption to school routines. In addition to these practical concerns, the selection of dilemmas was guided by the research team's judgment about the relevance of dilemma content for a UK audience, their apparent cultural universality and for other research purposes not covered in this paper. The dilemmas chosen for inclusion were then modified for the intended audience and age-group. Specifically, American terms were replaced with British ones and one dilemma, emphasising *courage*, was changed to a story about a female protagonist competing as a gymnast, in order to avoid possible confusion because the dilemma it replaced was about a fictitious situation: a play. This story had also been fully developed by the original expert panel. The structure of the original Ad-ICM measure was retained. Following these changes the proposed dilemmas were pilot tested with a representative group of UK adolescents and judged to be both relevant and compelling. The resulting set of three dilemmas captured the virtues of *honesty* (should one report a cheating incident involving peers in a school context?), *courage* (should one stand-up to a gymnastics coach to uphold personal values?) and *self-discipline* (whether to attend a final year trip or to prepare for a maths exam). The full Ad-ICM

can be viewed at http://www.ethicaldevelopment.ua.edu/adolescent-icm, and the shortened UK version at http://www.jubileecentre.ac.uk/userfiles/jubileecentre/pdf/Online_Survey.pdf. (Web links to be added – see file b &c).

The Intermediate-Concept approach was first developed by neo-Kohlbergians Rest, Narvaez, Bebeau and Thoma (1999). The neo-Kohlbergian tradition – foundational to this approach and often seen as gap-bridging between Kohlbergian rationalism and a character-based take on moral development – considers moral functioning to involve four component processes operating together and in interaction: ethical judgement, reasoning, sensitivity and action (Narvaez & Rest, 1995). Moral judgement – the process by which one identifies the morally ideal choice – is further segmented by three levels. At the most general and abstract level are 'bedrock' schemas' (referring to moral schemas that serve as default interpretive systems which are engaged when more content specific strategies fail to yield an appropriate judgment (c.f., Rest et al.. 1999)), in contrast to a third level of 'highly contextual norms' (e.g. professional codes) which prescribe action in specific circumstances. 'Intermediate concepts', as assessed by Ad-ICM, sit between these two levels as specific to daily life and related to virtue-based concepts (Thoma, Derryberry, & Crowson, 2013). Although in terms of bedrock moral schemas, three levels of moral judgement (personal interests, maintaining norms and post-conventional) may be equated to developmental stages, the cross sectional nature of this study precludes a developmental interpretation. Nevertheless, the three (bedrock) schemas are relevant because they have been found to influence responses to the measure in theoretically consistent ways.

The premise of the ICM measure is that patterns of ratings and rankings in response to the dilemmas reveal information about participants' ability to interpret and apply virtue concepts. Ad-ICM is not designed to assess 'moral schemas' *directly*, but respondents need a system-wide perspective to achieve higher scores and, thus, a mere personal-interest schemas will be a liability. Assumptions about a link between measures of moral schemas and intermediate-concept measures have received empirical support. For instance, Thoma and colleagues (2013) found that a reliance on a personal-interest schema was associated with low scores on the Ad-ICM measure, supporting the limitations of

a personal-interest moral conception on the application of virtue concepts within context. Additionally, Thoma and colleagues (2013) found that Ad-ICM scores related to behavioural outcome measures in a manner consistent with more traditional moral-schema measures. Taken together, these findings indicate that the Ad-ICM could perhaps be seen as a measure of moral functioning in general (rather than just moral reasoning), albeit with an emphasis on the cognitive aspects of virtuous character, and as such be relevant for evaluations of a virtue ethical conception of character.

The Ad-ICM measure has been through an extensive development and testing process, fully described, including psychometric properties, in Thoma et al. (2013). In their paper, initial data are presented supporting the Ad-ICM as a viable measure of moral reasoning in adolescent populations. Validity is claimed for this purpose based on the following: the measure distinguished between age educational groups as well as between individuals who 'act out' or behave poorly in school compared to other students. Scores achieved by Ad-ICM related to the Defining Issues Test (DIT) which is a well-established test of moral judgement, using dilemmas to target moral schemas at a predominantly non-verbal and intuitive level (Rest et al., 1999). This provided preliminary support for the claim that both the Ad-ICM and DIT assess 'the moral domain'. Thoma et al. concluded that an interpretation of the measure is that it is sensitive to the transition from personal interest to conventional thinking, based on the statistically significant correlations between Ad-ICM and DIT results. Furthermore, the findings in the study offered support for the possibility of intermediate concepts being identified in generic and non-professional settings.

Psychometric properties of the Ad-ICM measure indicate acceptable internal consistency reliability (alpha= .85) and there is evidence for construct validity including age/educational trends, relationships with established measures within the moral domain and to moral behaviour (Thoma et al., 2013).

Analysis

Data generated from the surveys were processed to allocate individuals scores on different dimensions of the measure. Such scores are in relation to pre-existing expert panel judgements (see Thoma et al., 2013). Following methods suggested in the development of intermediate-concepts measurements designed for the professions (e.g., Bebeau & Thoma, 1999), experts for the original Ad-ICM were defined as individuals who had an understanding of the intended population and context, coupled with a familiarity with ethical theory and practice. Specifically, for the Ad-ICM, experts were defined as professionals in the field of adolescent development with training in moral psychology. The UK research team acted as a second group of experts who were familiar with the UK context and evaluated each of the original item statements only for language use and acceptability, relying on the original US expert panel for other aspects. As already mentioned, this process resulted in some wording changes but the set of items were judged to be appropriate and were retained.

Participants' choices and justification rankings are identified as being 'acceptable', 'neutral' or 'unacceptable' as defined by the experts (original US experts and confirmed by the UK research team). An overall high score on the measure is achieved when a participant selects acceptable items in ranking good choices and justifications as well as unacceptable items when ranking bad choices and justifications. Additionally, scale scores are provided that represent the component parts of the assessment. These include independent scores capturing a participant's ability to identify acceptable and unacceptable items for actions and justifications, as well as scores on each dilemma. The summary and scale scores have a theoretical range of -100 (no match) to +100 (all matches), which can be presented as a percentage. These are reported in the results section in numerical form such that an ICM score of 51% shows this level of agreement with the expert panel and is reported in numerical form (.51). Typically, participants have a majority of choices in the appropriate direction so a few misidentifications can be absorbed and the summary score remains positive.

Results

Individual-level analysis

Mean percentages for the primary ICM indices are presented in Table 2. These findings suggest that, on average, adolescents scored less than fifty percent (M=.43) and results were evenly distributed across percentiles (25^{th}= .25; 50^{th}= .46; 75^{th}= .65). Inspection of the means and associated standard errors indicate that adolescents found it easier to select best actions (M=.51) and justifications (M=.41) than worst actions (M=.35) and justifications (M=.36). In other words, participants found it harder to identify poor options both in terms of what the protagonist should do and justifications for doing so. Adolescents were also better at picking best choices than at selecting best justifications (M=.51 versus M=.41), suggesting that they could identify more easily what should be done rather than explaining why. These within-subject differences are indicated by a significant main effect using the Greenhouse-Geisser correction for absence of sphericity (F (2.69, 8983.74) =236.84; p<.001; η_p2=.066). All subsequent repeated measure ANOVAs will be subject to the same procedures to test and correct for the absence of sphericity. Inspection of the individual contrast between means confirms that the best choices were better than the worst choices and best choices had higher means than justifications. No statistically significant differences were observed between the worst choices and the worst justifications.

Table 2 here

Gender differences. Table 2 also provides information on the ICM findings by gender. As expected from previous findings (e.g., Thoma, 2013), girls (M=.47) significantly outperformed boys (M=.37) (F(1,3319)=103.10, p<.001, d=.36, η_p^2 = .030). Extending the analysis to the four ICM subscales we again found a between-subject main effect (F(2.69,8930.26)=235.59, p<.001, η_p^2 =.066). The subscale findings highlight the magnitude of the difference between boys and girls; for instance, in detecting worst choices females are close to the overall average (M=.41) whereas boys are much below that (M=.28). The main effect was conditioned by a gender by subscale interaction effect (F(2.69,8930.26)=3.13, p<.05, η_p^2 =.001) which though small did suggest that the magnitude of the gender differences were not uniform across subscales. Specifically, the differences between the subscales were smaller for girls than for boys.

Performance by ICM dilemma. Each dilemma is designed to emphasise a specific intermediate concept. Scores, shown in Table 3, were highest for self-discipline (M=.63), dropping lower for courage (M=.44), but results for the dilemma emphasising honesty were very low indeed (M=.20) as indicated by a significant repeated measures ANOVA with dilemma as the within-subjects factor (F(1.879,6280.109) =1293.51; p<.001, η_p^2 =.28). This main effect was conditioned by a gender by story interaction (F(1.879,6237.92)=3.88; p<.05, η_p^2 =.001). Inspection of the story means suggests, much like the subscale results, that the interaction effect is due to smaller differences between stories for the girls.

Table 3 here

Demographic categories and ICM performance. Overall scores, shown in Table 4 below, were also related to a number of factors asked of the adolescents completing the survey. Given the presence of gender effects on the ICM measure, gender was included in all cases in order to assess whether these effects were moderated by this demographic category. Inspection of the table indicates that participants stating that they were religious scored higher (M=.44 versus M=.41) than those who selected either 'atheist', 'preferred not to say' or did not provide a religion (F (1,3254) =15.031; p<.001, η_p^2 =.005). The difference increased between pupils who said that they *practised* their religion (M=.48 versus M=.42) and those who did not (F (2,3277) =11.47; p<.001, ηp^2 =.007). These main effects were not conditioned by gender.

Parental education was found to relate to ICM scores (F(4,3291)=9.86 p<.001, η_p^2 =.012). Post hoc comparisons indicate that adolescents who said that either both of their parents or their father had attended university scored higher (Ms=.47) than those who did not know (M=.39) or said neither of their parents went to university (M=.42). These main effects were not conditioned by gender.

Adolescents who were asked how their school grades/results compared to their class mates, differed in their ICM scores (F(3,3275) =37.45; p<.001, η_p^2 =.033). Post hoc comparisons indicate that highest ICM scores were associated with those claiming to have 'mostly better' grades (M=.49), followed by

those selecting 'about the same' (M=.42), 'better' (M=.38), with 'somewhat lower' (M=.28) being associated with the lowest scores. These main effects were not conditioned by gender.

Each extra-curricular activity was assessed in isolation (from other activities). Fourteen percent of participants claimed to be involved in charity work, and these pupils scored higher on the measure (M=.50 versus M=.41) than those that did not participate in charity work (F (1,3317) =25.28; $p<.001$, η_p^2 =.008). Similarly, pupils involved in 'music/choir' outside of required school lessons had higher results (M=.48 versus M=.41) than those who were not involved (F (1,3317) =24.08; $p<.001$, η_p^2 =.007). This was also the case for those doing 'drama' outside of lessons (M=.48 versus M=.42) compared to those who were not (F (1,3317) =5.90; $p<.05$, η_p^2 =.002). Adolescents who claimed to participate in 'sports', 'debating' 'art/photography' or 'other' did not significantly differ from their non-participating peers on total ICM scores. However, there was a gender interaction effect for sport (F(1,3317)=8.70 $p<.003$, η_p^2 =.003) such that boys claiming to do it scored less well (M=.37) than boys who said that they did not (M=.40), whereas sport was associated with better scores for girls who participate than those who did not (M=.48 versus M=.46).

Table 4 here

School level analysis

To further explore relationships between different contexts and performance on the ICM, attention was given to variation associated with experiences within particular schools in the main sample. A mean ICM score was calculated for each school. This separated schools considerably (M=.29 to M=.58), with an overall school average of M=.43. When UK schools were ordered hierarchically by their mean scores, various types of school remained in both top and bottom quartiles, including independent and state schools; faith and non-denominational schools; schools with both grammar and academy statuses; schools from varied regions (including rural and city); as well as schools with varied rates/percentages for Free School Meals (FSM) eligibility, Ofsted (Office for Standards in Education, Children Services and Skills) and attainment histories. One school was excluded from both the original top and bottom quartiles because sample sizes were too small for analysis (n = <80). By chance, both of the removed schools had independent status.

Table 5 here

Table 6 here

Pinpointing why pupils at particular schools performed as they did in moral dilemma tests was inevitably difficult. A number of factors were assessed for their effect on individual results to assess whether a school's status on a variable of interest (e.g. faith school or not) related to how individual pupils within that school performed. Among factors with no (statistical) relationship are: the size of school; if the school has a city or rural location; if the school has independent (fee paying) or a non-fee paying status; and the percentage of pupils achieving five General Certificate of Secondary Education (GCSE) at grades A* to C in England or Level 4 in Scotland (r=.032. n=3100,p=.089). Ofsted ratings for English schools, shown in Table 6, did show statistically significant differences for ICM scores at the student level but the number of schools in each category was highly skewed making the results inconclusive. Some difference was also found for pupils going to a school that is classified as a faith school. These pupils achieved slightly, but statistically significantly, better total ICM scores (M=.46, n=628, sd=.29) than those going to non-faith schools (M=.42, n=2693, sd=.28), perhaps not surprising given that faith was associated with higher scores at an individual level ($F(1,3317)=10.887$ $p<.001$, η_p^2 =.003). Gender did not moderate this result. Very small, but significant, negative correlations were also found between pupils' total average ICM scores and (a) their school's percentage of FSM eligible pupils, (r=-.095, n=3343, p=>001), and (b) their school's local authority's unemployment rate (r=-.104, n=3037, p=>001).

A Hierarchical Linear Modelling approach (unconditional means model) was used to more systematically estimate the amount of variance in ICM scores that exists between schools. This approach accounts for the nested data structure where individuals (level 1, pupils) are nested within schools (level 2, schools), and both are assumed to potentially affect the outcome variable (ICM score). Using this approach, a statistically significant (p=>.001) variability was found between-schools as well as within schools respectively ($\tau 00 = 0.003921$ and $\sigma 2 = 0.076451$). The intra-school correlation coefficient was computed as 0.003921/[0.003921+ 0.076451]) = 0.48, which suggests that 5% of total moral score variability occurred between schools in the sample.

Summary of results

In summary, overall ICM results show that participants scored less than 50 percent. Scores were higher for dilemmas emphasising self-discipline and courage than for the one about honesty. Important differences were noticeable for the selection of 'best' and 'worst' choices as well as between 'action' and 'justification' selections. Girls outperformed boys in all ways. When ICM results were used to separate schools, many school factors such as Ofsted results, size of the school, or their percentage of GCSE A* to C results etc. did not seem to affect results, but encouragingly, a variety of kinds of school were present in the top performing schools and a number of factors were associated with higher ICM scores, such as extra-curricular activities and religion.

Discussion

The growing interest in character and its development calls out for better understanding of the strengths and limitations of adolescent applications of virtue concepts. The current study attempts to fill this gap by sampling a broad cross-section of 14-15 year olds on a measure assessing the application of virtues within realistic situations. The picture of young adolescents in this study presents adolescents as having difficulty applying the virtues in ways that represent an informed view of the virtue concepts. Within these overall trends, adolescents were found to have a relatively easier time identifying appropriate choices than justification, and identifying poor choices was more difficult than positive ones. Across all indices girls outperform boys.

Although these findings have implications for practitioners interested in developing character in this population, the question of why these scores fall in the middle to low range is more speculative. It is interesting to note that inappropriate items tend to represent a narrower framing of the situation that tends to highlight the interest of the protagonist rather than a focus on a virtuous action. In support of this contention we would note that in validating the Ad-ICM Thoma et al., (2013) found low Ad-ICM scores associated with personal interest moral judgments as measured by the DIT. Only when adolescents developed a system-wide perspective that explicitly acknowledged the moral basis of laws, norms and practices did Ad-ICM scores increase and better represent the informed view of the items as represented by the expert panel.

Low ICM scores occur when the individual tends to miss-specify items by placing bad items as best and good items as worst. Typically bad items tend to prioritize the self's interest and narrow conceptions of the situations. The implication of this is that many adolescents in British schools could be motivated by an understanding of cooperation that highlights self-interest, and this may be a useful starting point for informing character-education efforts.

Variations by dilemma were also found. Highest scores were achieved when self-discipline or courage was at stake compared to situations involving honesty. One explanation for such low scores for the honesty dilemma is that this creates for pupils a (hypothetical) pull between honesty on the one hand and loyalty to peers on the other. The desire to be well thought of by peers is probably a particular wrench for pupils aged 14 and 15 and this could be reflected in lower scores (cf. Steinberg & Monahan, 2007). However, suggestions of a more general decline in moral judgement in this age group may not be surprising to many parents or teachers given that these young people are beginning to attend to an adult identity, including the re-evaluation of traditional values and ideologies. (See also Nucci and Turiel (2009) for suggestions of a U-shaped pattern in moral development among children).

Across all dilemmas, two clear findings, replicated in other Ad-ICM research (USA), stand out, namely that pupils were better able to say *what* the protagonist should do than *why*, and that they could select 'best' options more easily than 'worst' ones. To our knowledge, this finding is the first of its kind in the UK context, especially involving so many schools and pupils. Knowing what to do more than being able to say why seems a likely reality among many young people who are conceivably habituated to some extent in the ways of good character, but who have not yet grown this into a reflective pattern for themselves. Social routine, modelling and habit may well help pupils determine the right thing to do, but the development of an experientially learned capacity to make good moral judgements supported by sound reasoning (or what virtue ethicists call 'phronesis') might be lagging behind. An inability to recognise poor choices or poor justifications in the face of difficult situations is an unfortunate deficit in life because individuals under pressure can act out of character or make poor choices with potentially negative repercussions. Although good moral judgment improves through life experience, both of these weaknesses - identified by the ICM survey - ought to be addressed directly in schools. Sensitising children to poor choices could be beneficial and would

represent an important amendment to many recent approaches to character education, especially those inspired by positive psychology (Peterson & Seligman, 2004) that tend to emphasise the positive over the negative. Prior sensitisation to a range of poor hypothetical choices, reasons and their consequences might in fact offer young people a degree of protection or buffering from making poor choices, especially at times of heightened pressure or stress.

Variances in ICM scores based on various demographic categories were also found, although these were generally small in magnitude. For example, positive correlations with having a religion (and more so for practising religion) may indicate the influence of the explicit teaching of virtue in many religions (Arthur, 2010; Pike, 2010, 2011). Religion could also be operating for some pupils in much the same way as a 'maintaining norms' moral schema might. In other words, a religious effect among pupils of this age could occur due to a 'system-wide' religious perspective (religious norms), rather than so much an individually reasoned moral choice. Nonetheless, Pan and Sparks (2012) urge caution about religion-moral-development links, and refer to other research showing no such relationship or to studies where the relationship is less significant than claimed. Participation in some extra-curricular activities was also associated with higher ICM scores. This concurs with other empirical and philosophical literature for: charity work (Arthur, 2010; Hill et al., 2009; Lies et al. 2012); music (Adderley et al., 2003; Campbell et al., 2007; Carr, 2008) and drama (Bouchard, 2002). Self-declared participation in sport was not, however, matched by higher ICM results, although a gender interaction effect shows that girls claiming to do sport achieved better scores than girls who did not. This main result is unsurprising because qualities developed in sport do not necessarily transfer to other domains and negative behaviour can also occur (Omar-Fauzee et al., 2012). More common is the view that sport *can* build character (Doty, 2006; Shields & Bredemeier, 2008) and has even more potential to help adolescents build qualities of good character than generally realised (Arnold, 1994, 1999). The present study suggests that we need to learn more about precisely how this might work, especially in terms of developing moral judgement or virtue.

Better responses to moral dilemmas have long been associated with females, who generally outperform males in tests of moral development (Malti & Buchmann, 2010; Nunner-Winkler, 2007;

Pan and Sparks, 2012; Sparks, 2014; Thoma, 1986; Van der Graaf et al., 2014; White, 1999). Explanations for these differences include role-based ones, socialisation theories (Pan & Sparks, 2012) and identity identification explanations (Nunner-Winkler, 2007). Why females achieved better ICM results in this particular study is somewhat unclear, although it does seem as if there is a female advantage in this domain. A degree of social-desirability bias should not be ruled out (if perhaps girls were more concerned than boys to identify choices in tune with what adults consider acceptable), but this can only occur when a participant is also able to apply appropriate moral judgement. Taken together with the finding that scores across the three dilemmas were marginally closer together (more similar) for girls than boys is suggestive that gender differences are worthy of further study.

Small negative correlations between ICM scores and certain socio-economic factors were also found. Although pupils were not asked directly whether they received FSM or about their parents' occupational statuses, reference to proxies (such as having parents who did not go to university, going to a school with a higher percentage of FSM eligibility, or if the school was in an area with a high unemployment rate) showed a significant but small effect on ICM scores. Many of the teachers at the schools also stressed a negative impact of socio-economic factors on character in many different ways, while also underscoring a rather obvious point: that families really matter for the development of character (cf. Sanderse et al., 2015). Context variation, as represented by different school ICM scores, was noticeable too, including the finding that different kinds of schools were among top performing schools. This suggests, rather encouragingly, that character - or in this case its moral judgement component - may be thriving in a variety of school types and in different ways. Given this variety it is likely that the underlying mechanisms that promote growth in ICM scores will be contextual rather than by kind. To explore this possibility we would encourage research designed to assess whether aspects of the school culture moderates the link between character educational activities and moral/character outcomes. Studies specifically targeting school climate and how this might be associated with moral outcomes seem especially warranted by our findings of meaningful differences among schools. The finding that high scores on the measure are not associated with a particular type of school, or indeed low scores, suggests that if larger national contexts influence

school performance in the moral domain, the effect will be equally subtle. Our sampling was designed to be sensitive to broad-based national and regional differences by ensuring a varied sample of participating schools but we recognize that these regions also differ on socio-economic, political and curriculum dimensions. For example, in Scotland the 'Curriculum for Excellence' creates a specific policy environment not replicated for other nations. Although using our methods we were likely to identify major distinctions between regional groupings, our design is not sensitive to the specific structural differences underlying the four nations, for example. Further studies attending specifically to differences between schools within different nations are also encouraged by our results.

A further methodological point deserves mention. As described earlier the current study is the first to use a shortened version of the Ad-ICM. One potential limitation in using a reduced set of dilemmas and associated virtue concepts is that meaningful information may have been lost. As a partial response to this concern we note that despite using fewer situations and target virtues the current findings are quite consistent with previous studies on all of the major outcomes including gender differences and the relative performance on the subscales (Thoma, Derryberry, & Crowson, 2013; Fritzhand, 2013). More practically, this observation suggests that future users of the Ad-ICM now have a justification to explore and use short forms of the measure, having established short forms will increase the utility of the measure for users interested in a broader range of applications and settings. More generally, our findings suggest that assessments like the Ad-ICM are more sensitive to variation in how the virtue is being applied rather than to different understandings of specific virtue concepts. Thus whether one concept is included or excluded appears less important to a measure than in assessing the characteristic way the adolescent applies the concept within a real-life situation. Our findings suggest that research designed to clarify the central features of contextually based measures of moral thinking seem especially warranted.

Conclusion

In terms of adolescent application of the virtues in realistic contexts, this study - using moral dilemmas - shows that many of the adolescents tended in their moral judgement towards an over-emphasis of 'self-interest', 'not getting involved' and 'conformity/loyalty to friends'. They tended to

know ‘what to do’ more than ‘why’ and often struggled to identify poor actions and poor justifications. Important age and gender variations were found as were associations with specific extra-curricular activities, but not with sport. Although pinpointing precisely how these results vary by individual schools has been difficult, it is nevertheless clear that a number of different kinds of schools were associated with high scores, suggestive that higher levels of moral judgement among adolescents can be associated with a variety of school, socio-economic and geographical circumstances, without wanting to underestimate the salience of those factors. Future efforts to improve moral judgement among adolescents in schools, as part of character-education programmes, will do well to address the key weaknesses suggested by the ICM measure. This will involve a certain amount of bucking the trend for character / moral education, in terms of challenging its tendency to emphasise positives to the detriment of exploring in more detail poor choices and reasons and their consequences. Early indications suggest that these main findings are replicable in multiple cultural contexts (Thoma & Walker, 2016). Relevant differences and similarities to samples in the USA, Macedonia and Taiwan are being explored further.

Tables 1 to 6

Table 1: Demographic Information for Pupils completing Ad ICM (UK)

Category	Options	(%)
Gender	Male	50.7
	Female	49.3
Ethnicity	White British/Irish, or other White	85.4
	British Asian (Indian/Pakistani/ Bangladeshi)	5.1
	Chinese, or other Asian	1.9
	‘Mixed’	2
	Black Caribbean,	1.1

	African, or other Black	
	Other/don’t know/ rather not say	3.4
Religion	Atheists/non-believers	34
	Christian	34.8
	Muslim	3.1
	Hindu	1.9
	Jewish	1.8
	Sikh	1.1
	Buddhist	0.5
	Other religion	3.3

Table 2: Group level ICM scores for main sample by Gender

Variable	Categories	Sample size	Total ICM	Subscales			
				Action choices		Justification choices	
				Best	Worst	Best	Worst
		3343	.43(.28)	.51(.32)	.34(.43)	.41(.35)	.36(.41)
Gender	Male	1596	.37(.28)	.46 (.33)	.28(.44)	.37(.35)	.31(.42)
	Female	1725	.47(.27)	.56 (.30)	.41(.42)	.45(.34)	.41(.39)

Note: Standard deviations are included in parentheses.

Table 3: Group level ICM scores for main sample by dilemma

Variable	Categories	Total ICM
By Dilemma	Honesty	.20(.45)
	Courage	.44(.35)
	Self-discipline	.63(.39)

Note: Sample size total is 3343. Standard deviations are included in parentheses.

Table 4- Demographic Categories and ICM performance

Variable	Categories	Sample Size	Total ICM
Religion	Selected	1513	.44(.28)
	Not selected	1745	.41(.29)
Religion is practised	Practised	612	.48(.27)
	Not practised	2331	.42(.28)
	Practised - don't know / rather not say	340	.41(.27)
Grades compared to others in class	Better	301	.38(.30)
	Mostly better	963	.49(.27)
	About the same	1816	.42(.28)
	Somewhat Lower	203	.28(.28)
Parents go to university	Both went	716	.47(.28)
	Father only	273	.47(.26)
	Mother only	349	.43(.29)
	Neither	1051	.42(.29)
	Don't know	912	.39(.28)
Extra-Curricular Activity	Charity yes	458	.50(.27)
	Charity no	2863	.41(.28)
	Music/choir yes	665	.48(.28)
	Music/choir no	2656	.41(.28)
	Drama yes	374	.48(.28)
	Drama no	2947	.42(.28)
	Art/photography yes	524	.45(.28)
	Art/photography no	2797	.42(.27)
	Debating yes	101	.44(.29)
	Debating no	3220	.43(.28)
	Sport yes	2121	.42(.28)
	Sport no	1200	.43(.28)

Note: Standard deviations are shown in parenthesis.

Table 5: 'Top 7 schools', by average total ICM score

School Code	Ad-ICM (UK) (%)	Region	Type	Faith	Location	Ofsted (England only (1 = highest))	% Achieving grades A to C or GCSEs/or Scottish Level 4[1]	% Pupils receiving FSM
6	.55	Staffordshire	Academy Converter	Roman Catholic	City	1	81	13.3
8	.52	Yorkshire	Independent	Christian	City	N/A	98	N/A
31	.51	Berkshire	Grammar Academy	No[2]	City	1	95	12.9
13	.50	Sussex	Academy Converter	No	City	2	63	21.6
3	.49	Derbyshire	Academy Converter	Christian	Rural	3	67	14.3
17	.48	Aberdeenshire	Academy Converter	No	Rural	N/A	81	8.3
26	.48	Hampshire	Academy Converter	No	City	2	58	18.3

[1] General Certificate of Secondary Education and Scottish Equivalent.

[2] School 31is a Multicultural school, 7.6% of respondents were white, compared to an average of 80% across the sample.

Table 6: 'Bottom 7 schools', by average total ICM score

School Code	Ad-ICM (UK) (%)	Region	Type	Faith	Location	Ofsted (England only (1 = highest))	% Achieving grades A to C or GCSEs/or Scottish Level 4	% Pupils receiving Free School Meals
22	.36	Ayrshire	State Funded	No	Rural	N/A	77	17
7	.36	Fife	State Funded	No	City	N/A	78	21.1
24	.36	Hertfordshire	Academy Converter	No	City	2	53	15.6
11	.35	Cheshire	Voluntary Aided	Roman Catholic	City	1	85	17
4	.35	Hampshire	Independent	No	Rural	N/A	95	N/A
27	.34	Shropshire	Academy Sponsor Led	No	City	2	61	50.1
32	.29	Durham	Academy Converter	No	City	3	63	26.5

Acknowledgments

This work was made possible through the support of a grant from the John Templeton Foundation.

References

Adderley, C. M., Kennedy, M., & Berz, W. (2003). A home away from home: The world of the high school music classroom. *Journal of Research in Music Education, 51*(3), 190–205.

Arnold, P. (1994). Sport and moral education. *Journal of Moral Education, 23*(1), 75-89.

Arnold, P. (1999). The virtues, moral education and the practice of sport. *Quest, 51*(1), 39–54.

Arthur, J. (2010). *Of Good Character: Exploration of Virtues and Values in 3–25 Year-Olds*. Exeter: Imprint Academic.

Arthur, J., Kristjánsson, K., Walker, D.I., Sanderse, W., & Jones, C. (2015). Character Education in UK Schools. Birmingham: Jubilee Centre for Character and Virtues, University of Birmingham.

Benninga, J. S., Berkowirtz, M. W., Kuehn, P., & Smith, K. (2003). The relationship of character education implementation and academic achievement in elementary schools. *Journal of Research in Character Education 1*(1), 19–32.

Berkowitz. (2011). What works in values education. *International Journal of Educational Research, 50*(3), 153–158.

Berkowitz, & Bier, M. C. (2006). What works in character education: A research-driven guide for educators. Washington: Character Education Partnership.

Berkowitz, M. (2012a). Moral and character education . In K. R. Harris, S. Graham & T. Urdan (Eds.), *APA Educational Psychology Handbook.* APA Handbooks in Psychology

Berkowitz, M. (2012b). Navigating the semantic minefield of promoting moral development Retrieved 20 Sep 2012, from http://amenetwork.org/oped/

Biesta, G. (2010). *Good Education in the Age of measurement: Ethics, politics, democracy*. Boulder, CO: Paradigm Publishers.

Biesta, G., & Miedema, S. (2002). Instruction or pedagogy? The need for a transformative conception of education. *Teaching and Teacher Education, 18*(2), 173–181.

Bouchard, N. (2002). A narrative approach to moral experience using dramatic play and writing. *Journal of Moral Education, 31*(4), 407–422.

Campbell, P. S., Connell, C., & Beegle, A. (2007). Adolescent expressed meanings of music in and out of school. *Journal of Research in Music Education, 55*(3), 220–236.

Caprara, G. V., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P. G. (2000). Prosocial foundations of children's academic achievement. *Psychological Science, 11*.

Carr, D. (2008). Music spirituality and education. *Journal of Aesthetic Education, 42*(1), 16–29.

Colby, A., Kohlberg, L., Speicher, B., Hewer, A., O Candee, Gibbs, J., & Power, C. (1987). *The measurement of moral judgment: Standard issue scoring manual, Vol. II*. New York: Cambridge University Press.

Doty, J. (2006). Sports build character. *Journal of College & Character, 7*(3).

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*(1), 405–432.

Exley, S., & Ball, S. J. (2014). Neo-liberalism and English education. In D. Turner & H. Yolcu (Eds.), *Neo-liberal Educational Reforms: A critical analysis*. London: Routledge.

Fritzhand, Ana (2013). *Moral Thinking and Moral Self-Concept in Adolescence: Results from the Study on Adolescents in the Republic of Macedonia*. OmniScriptum, Saarbrucken, Germany.

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.

Hill, M., Russell, J., & Brewis, G. (2009). Young people, volunteering and youth projects: A rapid review of Recent Literature. London: National Centre for Social Research.

Kristjánsson, K. (2015). *Aristotelian character education*. London: Routledge.

Lickona, T. (1992). *Educating for character: How our schools can teach respect and responsibility*. New York City: Bantam.

Lies, J. M., Bock, T., & Brandenbeger, J. (2012). The effects of off-campus service learning on the moral reasoning of college students. *Journal of Moral Education, 41*(2), 189–199.

Malti, T., & Buchmann, M. (2010). Socialization and individual antecedents of adolescents and young adults' moral motivation. *Journal of Youth and Adolescence, 39*(2), 138–149.

Narvaez, D., & Rest, J. (1995). The four components of acting morally. In W. Kurtines & J. Gewirtz (Eds.), *Moral behavior and moral development: An introduction* New York: McGraw-Hill.

Narvaez, D. T., & Bock, T. (2002). Moral schemas and tacit judgement of how the Defining Issue Test is supported by cognitive science. *Journal of Moral Education, 31*(3), 297–314.

Nucci, L., & Turiel E., (2009). Capturing the complexity of moral development and Education. *International Mind, Brain, and Education Society and Blackwell Publishing, 3(3),151 – 159.*

Nunner-Winkler, G. (2007). Development of moral motivation from childhood to early adulthood. *Journal of Moral Education, 36*(4), 399–414.

Omar-Fauzee, M., Nazarudin, M., Saputra, Y., Sutresna, N., Taweesuk, D., Chansem, W., . . . Geok, S. (2012). The strategies for character building through sports participation. *International Journal of Academic Research in Business and Social Sciences, 2*(3), 48–58.

Pan, Y., & Sparks, J. R. (2012). Predictors, consequence, and measurement of ethical judgments: review and meta-analysis. *Journal of Business Research, 65*(1), 84–91.

Park, N., & Peterson, C. (2006). Moral competence and character strengths among adolescents: The development and validation of the Values in Action Inventory of Strengths for Youth. *Journal of Adolescence, 29*(6), 891–909. doi: 10.1016/j.adolescence.2006.04.011

Pen Resiliency Project. (2015). Resilience research in children Retrieved 29 Jan 2015, 2015, from http://www.ppc.sas.upenn.edu/prpsum.htm (accessed 29 January 2015).

Peterson, C. & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. Oxford: Oxford University Press.

Pike, M. A. (2010). Christianity and character education: Faith in core values? *Journal of Beliefs Values, 31*(3), 311–321.

Pike, M. A. (2011). The value of christian-ethos schooling for secular students. *Journal of Research on Christian Education, 31*(3), 311–321.

Rest, J., Narvaez, D., Bebeau, M., & Thoma, S. (1999). *Postconventional moral thinking: A neo-Kohlbergian approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Sanderse, W., Walker, D. I., & Jones, C. (2015). Developing the whole child in an age of academic measurement: Can this be done according to U.K. teachers? *Teaching and Teacher Education 47*, 195–203.

Seider, S. (2012). *Character compass: How powerful school culture can point students towards success*. Cambridge, MA: Harvard Educational Press.

Seligman, M. (2011). *Flourish – A new understanding of happiness and well-being – and how to achieve them*. London: Nicholas Brealey Publishing.

Shields, D., & Bredemeier, B. L. (2008). Sport and the development of character. In L. Nucci, T. Krettenauer, L. P. Nucci & D. Narvaez (Eds.), *Handbook of moral and character education* (pp. 500–519). Oxford: Routledge.

Snyder, F., Vuchinich, S., Acock, A., J.Washburn, I., & Flay, B. R. (2012). Improving elementary school quality through the use of a social-emotional and character development program: A matched-pair, cluster-randomized,controlled trial in Hawaii. *Journal of School Health, 82*(1).

Sparks, J. R. (2014). A social cognitive explanation of situational and individual effects on moral sensitivity. *Journal of Applied Social Psychology*. doi: online 10.1111/jasp.12274

Steinberg, L., & Monahan, K. C. (2007). Age differences in resistance to peer influence. *Developmental Psychology, 43*(6), 1531–1543. http://doi.org/10.1037/0012-1649.43.6.1531

Thoma, S. J. (1986). Estimating gender differences in the comprehension and preference of moral issues. *Developmental Review*, 6, 165 -180. (Reprinted in Jacklin, C.N. (1992) *The Psychology of Gender*. London: Edward Elger Publishing).

Thoma, S. (2006). Research on Defining Issues Test. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development*. Mahwah, New Jersey: Erlbaum.

Thoma, S. (2014). Adolescent Intermediate Concept Measure data sets. Unpublished report: available from: the Center for the Study of Ethical Development, University of Alabama.

Thoma, S., Derryberry, W. P., & Crowson, H. M. (2013). Describing and testing an intermediate concept measure of adolescent thinking. *European Journal of Developmental Psychology, 10*(2), 239-252.

Thoma, S., & Walker, D. I. (2016, January). *Assessing adolescent's application of the virtues across multiple cultural contexts: An empirical summary of studies using the Adolescent Intermediate Concepts Measure.* Paper presented at the Jubilee Centre Conference 'Cultivating virtues: Interdisciplinary approaches', Oriel College, Oxford, Oxford. Retrieved January 19, 2016, from www.jubileecentre.ac.uk/1643/conferences/cultivating-virtues.

Tough, P. (2013). *How Children Succeed: Grit, Curiosity, and the Hidden Power of Character*. London: Random House Books.

Van der Graaf, Branje, S., M De Wied, Hawk, S., P Van Lier, & Meeus, W. (2014). Perspective taking and empathic concern in adolescence: Gender differences in developmental changes. *Developmental Psychology, 50*(3), 881–888.

Walker, D., Roberts, M., & Kristjánsson, K. (2015). Towards a new era of character education in theory and in practice. *Educational review, 67*(1), 79–96. doi: http://dx.doi.org/10.1080/00131911.2013.827631

Weber, M., & Ruch, W. (2012). The role of a good character in 12-year-old school children: Do character strengths matter in the classroom? *Child Indicators Research, 5*(2), 317–334. doi: 10.1007/s12187-011-9128-0

White, J. (2011). *Exploring well-being in schools: A guide to making children's lives more fulfilling*. London: Routledge.

White, R. D. (1999). Are women more ethical? Recent findings on the effects of gender upon moral development. *Journal of Public Administration Research and Theory, 9*(3), 459–371.