



THE
JUBILEE CENTRE
FOR CHARACTER & VIRTUES

UNIVERSITY OF
BIRMINGHAM

Applying the Virtues: similarities and differences in responses to the Adolescent Intermediate Concept Measure within and across cultures

Stephen J. Thoma and David Walker

This is an unpublished conference paper for the 4th Annual Jubilee Centre for Character and Virtues conference at Oriel College, Oxford University, Thursday 7th – Saturday 9th January 2016.

These papers are works in progress and should not be cited without author's prior permission.



Stephen Thoma
David Walker

Assessing adolescent's application of the virtues across multiple cultural contexts: An empirical summary of studies using the Adolescent Intermediate Concepts Measure.

Introduction

The purpose of this paper is to present a cross-cultural description of responses to the Adolescent Intermediate Concepts Measure (hereafter the AD-ICM). Since the introduction of the Ad-ICM, the measure has been used multiple times in the US, in The Republic of Macedonia, Taiwan, and the UK. Focusing on these data, this paper addresses what has been learned from the measure about adolescent applications of the virtues across cultures and settings. Specifically, we attend to what is common and what is unique to each setting. The paper describes and interprets these findings and concludes with a discussion about the development of character over the adolescent years and sketches out some of the educational implications of the work.

Perspective(s) or theoretical framework

Intermediate concept measures in the moral domain were first suggested as a response to critics of Kohlbergian-based professional ethics programs. To these critics, Kohlbergian-based measures were unable to adequately assess ethics training due to the highly abstract properties of the stages they measure. In this view, moral stages/schemas¹ are necessary but not sufficient explanations of moral reasoning in context. By contrast, ICMs represent the primary topics within professional ethics education programs and have significant face validity in the professions. For example intermediate concepts in the helping professions include topics such as informed consent, privacy, professional autonomy, among others. These concepts are nested within the profession and are claimed to be understood by the individual based on his/her moral judgment processes interacting with contextual factors including training, experience, precedent, and cultural definitions. At the level of professional ethics education, there is general consensus on what constitutes adequate and inadequate responses to the different intermediate level concepts. Specifically, although it has been found that professionals find it very difficult to arrive at a single best response or definition, there is surprisingly good agreement on the set of acceptable and unacceptable applications of the concept (e.g., Bebeau & Thoma, 1999). The fact that there is some consensus among experts about what is appropriate and inappropriate choices/justifications form the basis of ICM measurement systems. In each ICM measure the primary interest is the match between the participant's responses and the expert's interpretation of the intermediate concepts. The inclusion of different dilemmas in a single measure allow for assessment in a variety of contexts where different ICs are emphasized.

The more recent Adolescent ICM, discussed in this paper, adds to the support for the ICM approach by demonstrating that ICMs need not be tied to a professional context. Prior to the Adolescent ICM there were concerns that ICMs had limited utility as a generalized approach to assess moral reasoning because the measurements focused on a narrow set of concepts fully nested within a specific professional context (e.g., Walker, 2002). To transition from specific

¹ For example, a bedrock moral schema of 'self-interest' refers to moral judgments that are made largely on the basis of self-interest. A 'maintaining-norms' or 'conventional' schema involves a tendency to concord with norms surrounding a person even when these norms are morally suspect. A person with a 'post-conventional' moral schema can make their own considered moral judgements.

professional ICMs to a generalized adolescent ICM required a definition of intermediate concepts within a general adolescent population. Following the tradition of using the curricular focus of educational programs to help identify the most central intermediate concepts (e.g., Bebeau & Thoma, 1999), Thoma Derryberry and Crowson (2013) reviewed current character education programs to develop the core concepts of the measure. They found that across programs, the vast majority focus on the virtues. For instance Likona (1991) notes two core concepts that should reflect character: respect and responsibility. He further suggests that character incorporates concepts of honesty, fairness, tolerance, prudence, self-discipline, and courage. Comparable lists have been incorporated in other character education programs (e.g., Arthur, 2008). Preliminary work in the professions and work in adolescent populations support the utility of ICs as forming the basis for an assessment of discipline/context specific moral reasoning (e.g., Thoma, Derryberry & Crowson, 2013).

Methods, techniques, or modes of inquiry

This paper focuses on the Adolescent ICM (hereafter the AD-ICM). The measure presents adolescents with real-life situations that have been designed to highlight the cognitive aspects of virtues. For example, in the honesty story a student is in a panic because she has not been able to study for an exam to the degree she would like. Another male friend has already completed the exam in an earlier session and offers to help identify the material covered on the exam. The protagonist in the story - also a friend - observes the exchange and is shocked because all students who took the test in the earlier session have pledged not to say anything about the test and its content. The protagonist is shocked and wonders what to do. After reading this and the other stories, the respondent must select the best action choice for the protagonist and then the best justification for an action. Additionally, participants are asked to identify the worst action choices and justifications. High scores on the measure represent a response pattern in which the good and bad choices and justifications match expert's interpretation of the items as acceptable, unacceptable or neutral. Low scores by contrast, are the result of ranking items as good that the experts judge to be unacceptable as well as the reverse. Current empirical evidence suggests that the Ad-ICM is a reliable measure that produces trends consistent with theoretical expectations for an ICM measure (Thoma, Derryberry & Crowson, 2013).

Data sources, evidence, objects, or materials

The data sets used for this proposal were all part of larger assessments completed in the past few years. From the US, there are three samples: a multi-age adolescent sample (n= 125) roughly equal by gender; a high achieving private high school sample also balanced by gender (n=49), and a late adolescent sample (n=173) and a college sample (n=70). From the United Kingdom there are two samples: the first is a sample of 14&15 year olds (n=4053) from 39 geographically spread schools and the second is a slightly smaller group aged 10 to 18 (n=3223) from a smaller number of schools. From Taiwan is an early adolescent sample 13-15 years old (n = 1430, male 49%). From the Balkans is a Macedonian sample N=68 (early adolescent < 15 years old) and a middle adolescent group n=86.

Measures. Two measures of intermediate concepts are summarized here: the original Adolescent ICM (Thoma, Derryberry & Crowson, 2013) and a short version developed for the UK and Taiwan studies. The basic measurement approach used by both versions included rating and ranking tasks focused on both action choices (e.g., what should the protagonist do/not do?) and

justifications (what is an appropriate /inappropriate justification for action?). The AD-ICM in its traditional form includes seven stories and takes about 45 minutes to complete. The short version includes three stories derived from the original seven. The stories are the same in all respects with two exceptions: the courage story used in the UK and Taiwan samples uses an alternative setting while keeping the basic issue constant. Second, some wording changes were made to reflect culturally specific terms and informal language. The scoring of participant's choices is accomplished by attending to the degree of fit to a key developed by experts who evaluated each story and items and judged each item as acceptable, unacceptable and neutral. Specifically, items were rated as acceptable if the experts agreed that the ideas expressed by the item could be justified. Similarly, unacceptable items represented ideas that the experts agreed were not justified. Experts were defined as graduate students in Adolescent Psychology with training in moral psychology (see Thoma, Derryberry & Crowson, 2013 for a rationale for this choice of experts). Following these scoring methods, high scores represents a ranking pattern that reflects the expert key both in terms of acceptable and unacceptable items. In addition to an overall score, this scoring process also yields scores for "bad choices" (the degree to which the participant identifies action choices and justifications the experts see as unacceptable) and "good choices" (the degree to which the participant identifies action choices and justifications the experts also see as acceptable). Scores are represented as percentages so for example a "good" score of .80 indicates that 80% of the choices match the expert derived key.

Results and/or substantiated conclusions or warrants for arguments/point of view

Age/educational trends: Taken together, older adolescents achieve higher scores than their younger peers (Figure 1), although there is a contradictory trend in the UK data mostly for the dilemma emphasizing honesty (Arthur, Kristjánsson, Walker, Sanderse & Jones (2015). These findings are generally consistent with the view that the ability to apply the virtues within context is developmental. Within these general trends, there is some variation in average scores across groups. It is unclear whether these mean differences reflect real performance variations or issues related to the relevance of the stories in non-US contexts and/or weaknesses in the translation process. However, it is interesting to note that the Taiwan sample is associated with the highest scores for the assessed age ranges. This group is plausibly the most affected by story relevance and translation issues and thus it is unlikely that measurement issues are major contributors to these group findings. Clearly, US populations do not have an advantage on the measure despite having been the context emphasized in the development and validation process.

Gender Differences. Gender differences favoring women are large and pervasive across contexts and age groups (see Figure 2). Consistent with other measures of moral and character measures (e.g., Thoma, 2006; Wang, et al., 2015), females appear to be better able to respond in a manner consistent with expert opinion. This finding extends to the subscales as well. Females were consistently better at identifying best and worst choices and justifications.

Identifying acceptable and unacceptable choices and justifications. Across contexts and samples participants appear to have an easier time identifying acceptable choices and justifications (see Figure 3). Good choices and justifications appear easier to identify than similar decisions about bad choices and justifications. This difference may diminish with age and educations, but it is evident in each sample. Interestingly, the trend toward smaller difference by age/educational levels was particularly evident in the US and Macedonian samples-both samples that used the full seven story version of the AD-ICM.

The disparity between identifying acceptable and unacceptable choices also raises the question of why is it that the inappropriate responses appear more difficult for participants? Focusing on the possibility of an experiential explanation for these findings we wonder whether there is less of a focus on the “bad” when children and youth come to understand the implications of moral concepts. It appears that we cannot assume that a focus on the good or positive applications of a concept automatically translates into the ability to fully understand its negative aspects. That is, perhaps we are instructed on what one ought to do and what are appropriate choices and must then infer what is inappropriate and poorly conceived. If so, moral education may benefit by expanding its focus on inappropriate choices in order to provide more guidance in developing an experiential base to support more optimal moral decisions. Consistent with this view is a moral educational focus that brackets the concept by highlighting both negative and positive manifestations of the concepts. Whether the overall finding of “good” ”bad” difference reflects instruction or a normative pattern of concept acquisition, it is a now a well-established finding worthy of further study.

Differences by Story and Context. Focusing on the three stories in common across studies, statistically significant differences were noted between the dilemmas (Self-discipline > Courage> Honesty) (see Figure 4). However, there were context by story differences. Particularly noteworthy was the relatively consistent performance across the three stories found in the Taiwan sample. By contrast, the other samples indicate variation across stories primarily due to the low performance on the honesty story. Interestingly, this story asks participants to prioritize friendship over honesty. Participants in the Western samples –particularly boys were unwilling to set aside friendship considerations (see Figure 5A). Although also sensitive to friendship considerations, girls (Figure 5B) did not show the same decline in performance.

Conclusions

Results from studies using the AD-ICM suggest adolescents in the four contexts are remarkably similar in their application of the virtues to specific contexts. This similarity is particularly interesting in that the stories and items were developed within a US adolescent population. It may be that the role of context is less central when the intermediate concepts are based around the virtues because these concepts are recognizable within a variety of Eastern and Western ethical traditions (Shanahan & Wang, 2003) and relate to situations likely to be experienced by all youth. The observed similarities notwithstanding, it is important to note the difference between Taiwan and the other groups on the honesty story where a tension is developed between friendship and honesty. As mentioned above, participants must weigh a pledge to a teacher not to divulge an exam’s content against reporting a friend and the damage this action might do to their friendship and the friend’s academic standing.

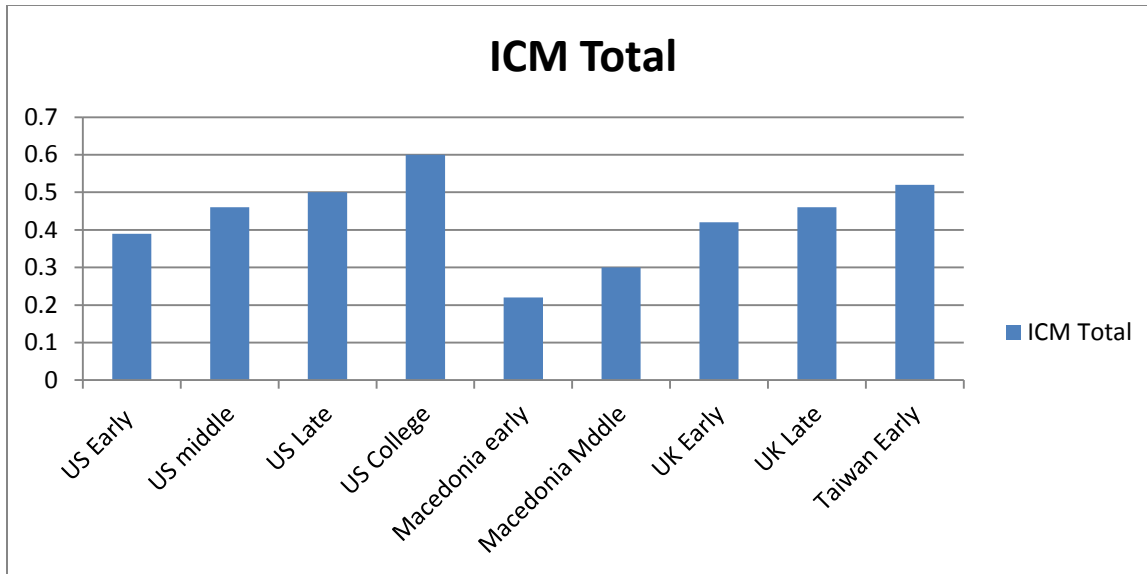
One wonders whether the focus on friendships is a reflection of the relative priority placed on the collective and roles within the group (e.g., the teacher) rather than individual relationships. In addition, two generalized findings need further attention. First, why is it that females are more advanced than their male peers? Early in the construction of the measure it was thought that the female advantage was due to the possibility of having a stronger voice in the test development process. The consistent pattern across contexts suggests instead that the effect may be a more interesting gender difference in performance. Also of interest is the difference in acceptable and unacceptable responses. These differences have implications for moral/character education by suggesting that knowing good choices does not translate into knowing what is inappropriate. That is, educational interventions ought to emphasize both positive and negative manifestations

of the concepts. We also wonder whether the finding of a shrinking gap between acceptable and unacceptable choices with age/education implies that maturity can be assessed by the ability to understand and apply both the good and bad aspects of a concept. Perhaps this integration of both understandings is a hallmark of development during the adolescent years. Finally, these findings further support the AD-ICM as a cognitive measure of intermediate concepts defined by the virtues.

References:

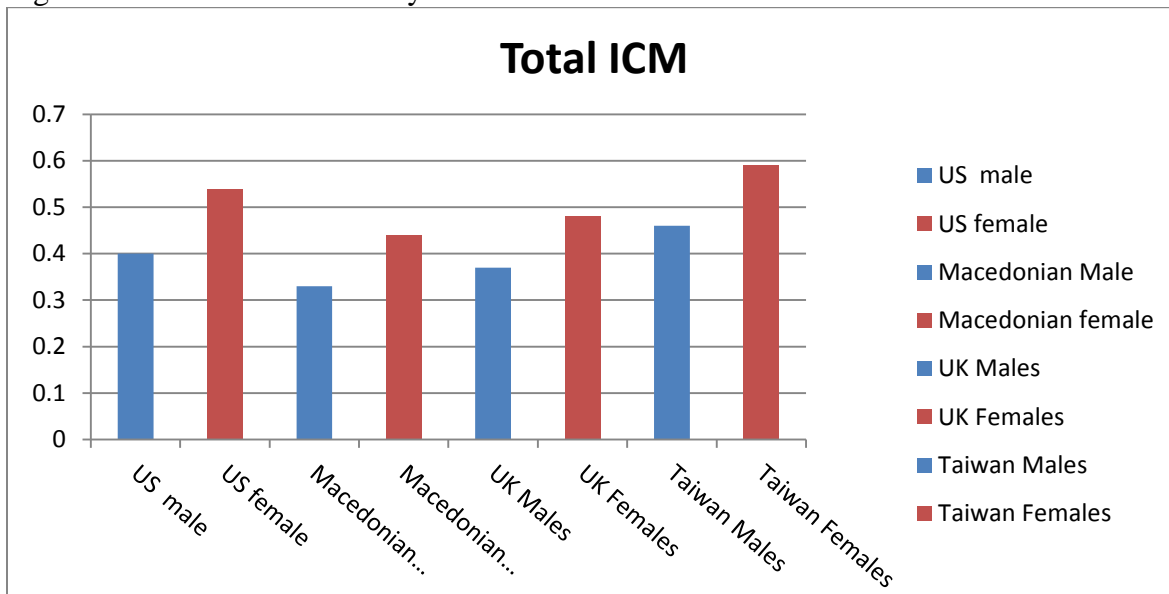
- Arthur, J. (2008). Traditional approaches to character education in Britain and America. In L.P. Nucci & D. Narvaez (Eds.). *Handbook of Moral and Character Education*. Routledge, New York.
- Arthur, J., Kristjánsson, K., Walker, D., Sanderse, W., & Jones, C. (2015). *Character Education in UK Schools*. Birmingham: Jubilee Centre for Character and Virtues, University of Birmingham.
- Bebeau, M., & Thoma, S.J.(1999). “Intermediate” concepts and the connection to moral education. *Educational Psychology Review*, 11, 343-360.
- Lickona, T. (1991). *Educating for Character*. New York :Bantam.
- Shanahan, T. and Wang, R. (2003, second edition). *Reason and Insight: Western and Eastern Perspectives on the Pursuit of Moral Wisdom*. Thomson Wadsworth.
- Thoma, S.J. (2006). Research using the Defining Issues Test. In Killen and Smetana (Eds.), *Handbook of Moral Psychology*. L. Earlbaum: Mahwah, NJ.
- Thoma, S.J., Derryberry, P., & Crowson, H. M. (2013). Describing and testing an intermediate concept measure of adolescent moral thinking. *European Journal of Developmental Psychology (special issue)*. 10, 239-252.
- Walker, L. (2002). The model and the measure: An appraisal of the Minnesota approach to moral development. *Journal of Moral Education*, 31, 353-367.
- Wang, J., Hilliard, L.J., Hershberg, R.M., Bowers, E.P., Chase, P.A., Champine, R.B., Buckingham, M.H., Braun, D.A., Gelgoot, E.S., & Lerner, R.M. (2015). Character in Childhood and early adolescence: models and measurement. *Journal of Moral Education*, 44. 165-197.

Figure 1: Total ICM scores by group and context.



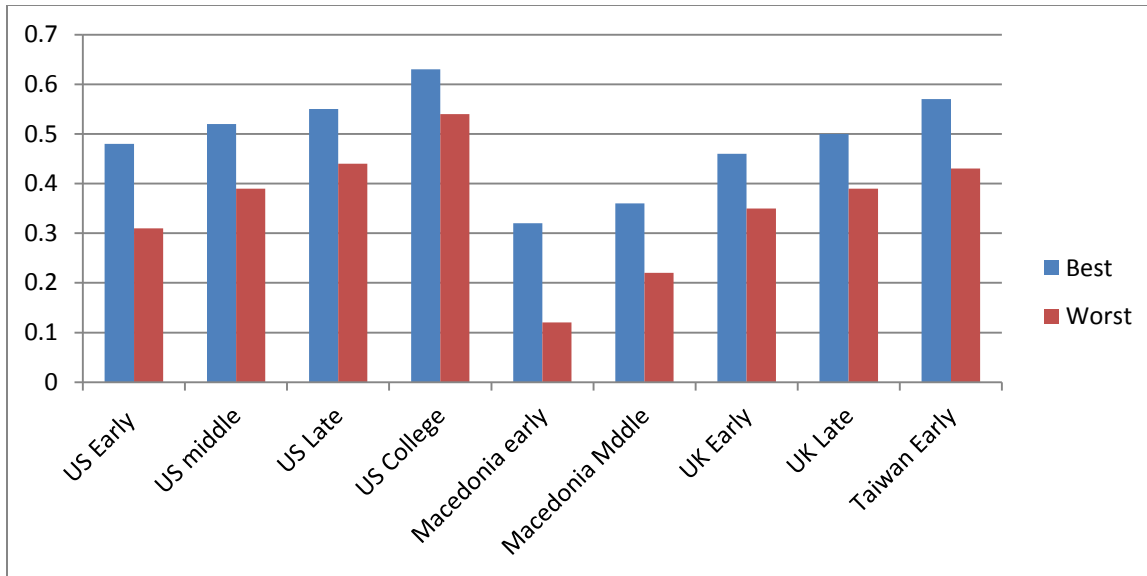
Note: All US groups are statistically different with the exception of the US Middle and US late (a statistical tendency). Both the Macedonian and UK samples are statistically different.

Figure 2. Gender differences by context.



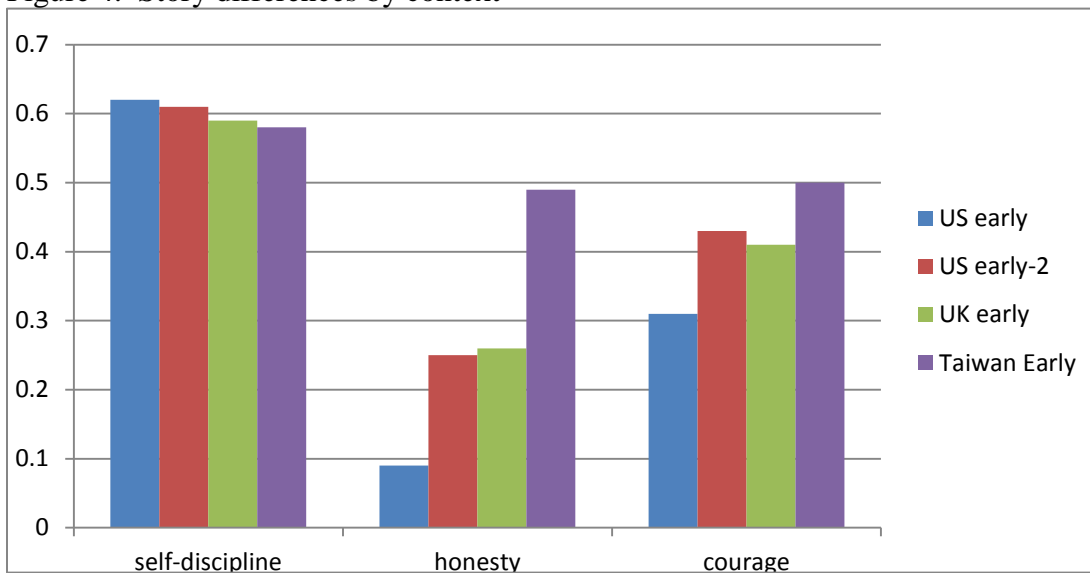
Note: All gender differences are statistically significant and large (effect sizes > .50)

Figure 3. Mean unacceptable and acceptable scores by cultural context.



Note: All mean differences within groups are statistically significant.

Figure 4. Story differences by context



Note: Statistically significant differences by story (Self-discipline > Courage > honesty). The Macedonian sample did not provide story specific data.

Figure 5A. Story difference by context for males.

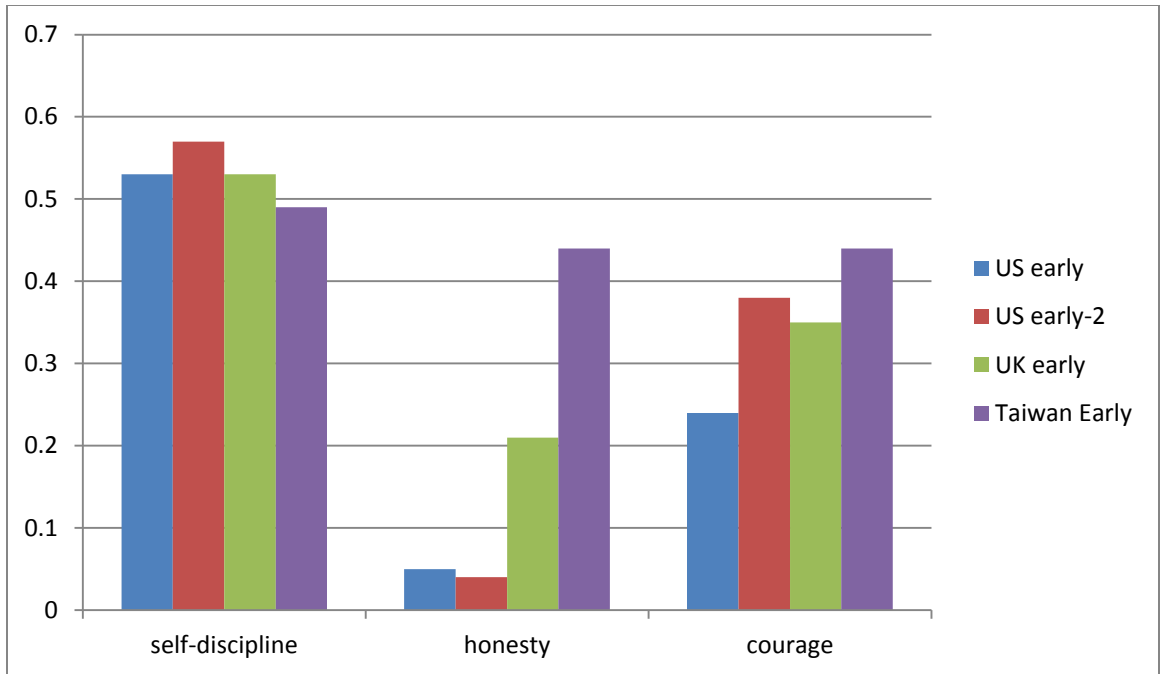


Figure 5B. Story difference by context for females.

