**Method**

*Samples*

In the current study, we used the latest wave of the PISA study, PISA 2012. PISA 2012 contained responses from 485, 490 15-year olds students from 68 countries and regions all over the world. Among all participants, 50.5% were female.

In PISA 2012, there is a wide variety of sample sizes, ranging from 293 for Liechtenstein to 33806 for Mexico, and in order to avoid the imbalanced influence of different sample sizes on the results, we decided to sample randomly 500 people from each country/region for the analyses. Therefore, Liechtenstein, with a total sample size smaller than 500, was excluded from the analyses.

*Grouping*

As discussed in the Introduction, we did not conduct the analyses using individual countries, which we believed would probably overstate the sometimes trivial differences between countries, especially when those countries are geographically proximal, influenced by the same culture (e.g., Confucianism; Islam), and/or ideology (e.g., communism). Therefore, we categorized the 67 countries/regions into 9 major cultural groups.

The categorization of the groups majorly derived from Saucier et al. (2015), where the authors grouped 35 countries into 9 groups based on locations. Their grouping seems reasonable also from the cultural, political, or economic perspective, and we believe that it’s because most of the time, countries close to each other tend to have similar cultural background, politics, and economy conditions.

However, PISA 2012 had 32 more countries/regions than in Saucier et al. (2015), and by looking at theses individual countries and regions, we felt that geographical proximity alone was not enough for obtaining the most representatively distinctive groups for the current study involving education conditions and personality traits. A good example of this would be North and South Korea, although the former didn’t participate in PISA 2012. The two countries are next to each other, but have drastically dissimilar conditions in almost every ways. The Korean Peninsula is for sure not the only part in the world where such differences exist. As a result, we decided to take into consideration the influence of culture, politics, and economy conditions, above and beyond region. Exactly which country or region belongs to which group can be found in Table 1, according to which, not surprisingly, most countries fell in the same group with nearby countries, with similar cultural backgrounds, economy conditions, and politics. We came up with two new groups missing from Saucier et al. (2015), one is a group containing Middle East countries, and the other including only the Nordic countries. Also, when there was a big difference in culture, politics, or economy conditions among countries in the same region, we went with non-geographic factors. For example, Estonia, Latvia, and Lithuania are all Norther European countries, but considering from both the political and economic perspectives, we figured it more appropriate to put them in different groups from the five Nordic countries, UK, and Ireland, all of which are also located in Northern Europe (**UNSTATS)**. Therefore, these three former USSR countries ended up in the group containing other former communist countries, such as Russia, Poland, and Czech. Kazakhstan, as a former USSR country and the only Central Asian country in PSIA 2012, was also assigned to this group. Ireland and UK were put under the “Western Europe” category, where one can also find other well-developed Western European countries, including France, Germany, and Netherlands. Nordic countries, due to their geographic proximity, and unique Nordic capitalism, became an individual group themselves.

*Measures*

Since we were interested in the relationships between conscientiousness, truancy, and math performance, we include in the analysis the 5 items from the Perseverance scale, the 3 items on the Truancy scale, as well as the Math Achievement test. We also chose SES (measured by the item asking about the highest education level of parents) and Gender as control variables.

The Math Achievement test was divided into 21 booklets (PISA 2012 codebook), and each participant was randomly assigned to one of them. That is to say, each participant responded to a set of the the total items. We treated all booklets equally in the current study, as PISA 2012 intended, and used the mean of all items on the booklet that a participant was assigned to as her overall math performance. Students’ responses were graded as 0 (“no credit”), 1 (“partial credit”), and 2 (“full credit”). If a participant failed to give an answer, the response will be coded as 0, instead of missing, which means all missing values in the data were due to either the item not being administered to a participant or being dropped for being inaccurate in translation. If a participant missed all items (possible when the participant was “not reached”), her overall math achievement couldn’t be computed, and thus would be recorded as missing. The percentage of this kind of missing values in the complete dataset is 0.6%.

The Truancy scale consisted of 3 items asking about how often students were (1) late for school, (2) skipping whole school day, and (3) skipping classes within school day. The participants were asked to indicate the frequencies for each of the three behaviors on a scale from 1-4, where 1 stands for none, 2 stands for one or two times, 3 stands for three to four times, and 4 means five or more times. The response rates for the three truancy items are all around 1.3%. The mean of the three items were computed to stand for participants’ overall truancy, and in the complete data, the missing rate for this variable is 1%.

The Conscientiousness scale in PISA 2012 is called Perseverance, which contains 5 items, tapping into different facets of conscientiousness (MacCann et al., 2009), including industriousness (“Exceed expectation”), perfectionism (“Continue to perfection”), procrastination refrainment (“Put off difficult problems”), and perseverance (“Give up easily”; “Remain interested”). Participants reported how well these statements described them on a scale from 1-5. This scale is the reversal of the common Likert-type scale, where the larger number indicate the higher the degree of agreement. On this scale, 1 stands for the highest degree of agreement (“Very much like me”), while 5 refers to the lowest (“Not at all like me”). Therefore, before running the analyses, we reverse coded the responses to all 5 items, so that the larger the number, the higher the degree of agreement with the statement.

Response rates in the complete data for the 5 items were between 64.1% to 64.5%, with the missing rates around 35%. A variable called “Perseverance” was also given by PISA 2012, which was the overall level of perseverance computed based on the 5 items using an IRT approach. The response rate of this variable in the complete data is 64.5%. The score for Perseverance was coded as missing only when responses to at least 4 items failed to be recorded, and therefore, we excluded such cases before random sampling. In the remaining dataset containing 313,172 participants, the missing rates for each of the 5 items dropped from about 35% to less than 1% (0.2%, 0.4%, 0.6%, 0.3%, and 0.3% for the five items, respectively).

Next, we randomly sampled 500 people from each of the 67 countries, using the dataset mentioned above, where there was no missing value for overall perseverance. This gave us a sample of 33,500 participants. There were no missing values for any of the 5 perseverance items, and 4% of all cases have missing values for either overall math performance, overall truancy, or both. Given the low missing rate, we dropped these 4%, and ended up with 32,794 people in the final random sample for the analyses. Therefore, the samples were identical for studying ME and prediction of math and truancy by perseverance. Across the 9 cultural groups, the percentage of female ranged from 48.2% to 53.9%.

*Analyses*

Using the random samples, we started with obtaining the factorial structure of the Perseverance scale. The exploratory factor analysis indicated a two-factor solution, with the first two factor accounting for 73% of all variance. The first factor was loaded onto by only the positive items (i.e., “Exceed expectation”; “Continue to perfection”; “Remain interest”), and thus was named “Positive”, while the second one we decided to call it “Negative”, because only the negative items loaded onto it (i.e., “Put off difficult problems”; “Give up easily”). Following the EFA, a 2-factor CFA with direct oblimin rotation showed that the two factors correlated at -.154.

The existence of a clear 2-factor structure indicated that the scale is not unidimensional, and that there were two sub-facets, positive and negative. Therefore, we calculated the Cronbach’s alpha across the 9 groups for the two facets separately, and most of them showed acceptable reliability across the 9 groups (Negative: α ranged from 0.63 to 0.79, with 2 of them below but close to 0.7, i.e., 0.63, and 0.68; Positive: α ranged from 0.73 to 0.79).

We assessed the measurement equivalence of the Perseverance scale in Mplus 7.4 (Muthén & Muthén, 1998-2010), using multigroup structural equation modeling (SEM). According to the results of EFA and CFA, we chose a correlated factor model, the diagram of which can be found in Figure 1. In this initial multigroup analysis, all parameters (i.e., loadings, and intercepts) were constrained across groups, as if the measurement is completely equivalent across the 9 groups. Then based on the modification indices (MI) computed for these parameters, we freed some of the parameters in certain groups that had large MI, in order to obtain a partial invariant measurement model. Mean values of the two factors in the constrained and the partial invariant models for each of the groups were compared.

Next, we added a path model to the partial invariant model, so that the two factors of perseverance were used to predict the overall math performance, and overall truancy, one at a time, while controlling for SES and gender. With the partial invariant model included, we were able to predict the two outcomes separately with a relatively equivalent measurement, which is a prerequisite for meaningful comparisons across groups.

With the regression solution, estimated correlations between latent factors and the outcomes, and the change in latent means, we were able to compare self-reported conscientiousness and its effects on math and truancy across 9 different major cultures, when effects of gender and SES were controlled for.

**Results**

**Descriptive statistics for overall math, overall truancy, perseverance~~, and the 5 C items~~ for each of the 9 groups, with sample sizes;**

**Correlation matrices: correlations between math, truancy, and perseverance within each of the 9 groups;**

**Correlations among the 3 variables across the 9 groups, across countries – different direction!**

Descriptive statistics of the overall math performance, overall truancy, and perseverance, and zero-order correlations among the three variables within each of the 9 cultural groups can be found in Table 2. All correlations were statistically significant at the 0.05 level. The directions of the correlations were pretty intuitive, with math and perseverance positively related, while truancy and perseverance negatively related. This pattern was consistent what we found within each the 67 countries/states/provinces, where the correlations were all positive between math and perseverance, ranging from xx to xx, while negative between truancy and perseverance, ranging from xx to xx). However, such trends were reversed both across countries and groups, where math and perseverance were found to correlate negatively (across groups: *r*s between xx and xx; across countries: *r*s between xx and xx), whereas positive correlations were found between truancy and perseverance (across groups: *r*s ranged from xx to xx; across countries: *r*s ranged from xx to xx).

**ME: all constrained – model fit; some freed using MI > 4 rule – model fit; in order to show that nonequivalence exists, but not severe; put latent means before and after some of the parameters were freed in the same table, along with the change.**

Multigroup SEM for the measurement model of perseverance, where all parameters were kept equal across group, yielded acceptable model fit (*RMSEA* = 0.068, *χ2 (84)* = 1483.93, *SRMR* = 0.043). When some of the parameters with extremely large MI were allowed to be freely estimated for some of the groups, the model fit became more satisfactory (RMSEA = 0.048, *χ2 (70)* = 664.09, *SRMR* = 0.032), indicating that there was measurement nonequivalence in the perseverance measurement, but was not severe.

For each group, Mplus also estimated the standardized means for the two latent factors which can be found in Table 3, along with the changes in the means before and after some of the constraints were removed. Table 3 also contains a complete list of model fit indices provided by Mplus, including the ones reported above. Here, the Southern Europe group was used as the reference group, so the latent means of the other groups were relative to the reference group. As shown in Table 3, compared with when all paramters were constrained, i.e., when the conscientiousness measurement was assumed equivalent across cultures, when we freed some of the parameters that were obviously nonequivalent, the latent means didn’t change by much, neither did the ranks, except for Southeast Asia on the Negative facet. With the constrained measurement model, the Southeast Asia group was the fourth highest for the Negative facet, whereas in the partial invariant model, the group was estimated to be the eighth highest (i.e., the second lowest) for the same facet.

Middle East participants got the highest self-ratings on both the negative and positive factors. Participants from Nordic countries had one of the highest means for the negative factor, but the lowest for the positive facet, and thus in general rated themselves the most negative on conscientiousness. The East Asia group had a similar pattern as the Nordic Country group, but not as extreme, as the group ended up the third highest for the negative facets, and the third lowest on the positive facet. Participants from Southeast had the most positive view of themselves on conscientiousness in general, as they were the second lowest on the negative facet, and also the second highest on the positive facet. North America/Oceania also had an interesting pattern, in that the students in this cultural group rated themselves the lowest on negative facets, but in the middle (the fifth) on the positive facet, as if they do not think they are unconscientious, but don’t think that they are the most conscientious, either. The group of Western European countries, however, showed an opposite pattern from the North America/Oceania group. Participants from this culture group rated themselves as one of the lowest (the second lowest) on the positive facet, yet somewhat in the middle (the fourth) on the negative facet, as if they didn’t see themselves as conscientious at all, yet not very unconscientious, either. The Latin America group had very similar latent means for both factor to the Southern Europe group.

**SEM: partial invariance model + path model – standardized coefficients in a table; figure out which of the 5 items is the best predictor for math and truancy, respectively (use standardized coefficients); report the correlations between latent variables and the outcome variables in a table**

In terms of prediction of the overall math performance and truancy, respectively, standardized regression coefficients can be found in Tables 4-5. For math, the magnitude of the standardized coefficients of the negative factor were much larger than that of the positive factor, indicating that the negative factor was a much stronger predictor for math than the positive factor. For truancy, we couldn’t find a clear pattern. Although in 7 out of the 9 cultural groups, the positive factor was a stronger predictor than the negative factor for the prediction of truancy, the difference in magnitude was very small.

Correlations between latent factors and the outcome variables were all in the same direction as the sample correlations, meaning that the negative factor was negatively associated with math, while positively associated with truancy, while the positive factor exhibited a reversed pattern.