**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 the 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, so in order to avoid potential biases brought by the drastically different sample sizes, we decided to randomly sample 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, or influenced by the same culture (e.g., Confucianism; Islam) and 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, because most of the time, countries close to each other tend to have similar cultural background, political systems, and economy conditions. However, there are a lot more countries/regions in PISA 2012 than in Saucier et al. (2015), and we felt that geographical proximity alone was not enough for obtaining the most representatively distinctive groups for the current study involving academic achievement and personality traits. The dissimilarity between North and South Korea, although the former didn’t participate in PISA 2012, had led us to doubt if proximity necessarily indicates resemblance. 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. Not surprisingly, most countries fell in the same group with nearby countries with similar cultural background, economy conditions, and politics. We’ve come up with two new groups missing from Saucier et al. (2015). One is a group made up of the Middle East countries, and the other includes only the Nordic countries. During the grouping process, when there was a big difference in culture, politics, or economy conditions among countries in the same region, we went with the non-geographical factors. For example, according to The United Nations Statistics Division (UNSD), Estonia, Latvia, and Lithuania are all Northern European countries, but considering from both the political and economic perspectives, we found it more appropriate to put them in different groups from the Nordic countries, UK, and Ireland, all of which are located in Northern Europe (UNSD, 2014). Therefore, these three former USSR countries ended up in the same group with other former communist countries, such as Russia, Poland, and Czech. Kazakhstan, a former USSR country and the only country in Central Asia participating in PSIA 2012, was also assigned to the Former Communist Countries group. Ireland and UK were put in the Western Europe category, whose members these two countries bear a resemblance to. In this group, one can find some of the well-developed Western European countries, such as France, Germany, and Austria. Nordic countries, due to their geographic proximity, and unique Nordic capitalism, formed a group themselves.

*Measures*

Since we were interested in the relationships of conscientiousness with truancy and math performance, we included in the analyses 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 the control variables.

The Math Achievement test was divided into 21 booklets (PISA 2012 codebook, 2012), 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 score of all items on the booklet that a participant was assigned to as her overall math performance. Responses were graded as 0 (“no credit”), 1 (“partial credit”), or 2 (“full credit”). If a participant failed to give an answer, the response will be coded as 0, instead of missing, so all missing values in the data were due to the item not being administered to a participant, except when someone missed all items, which only occurred when the participant was “not reached”, as indicated in the data file. The percentage of unreachable participants in the complete data is 0.6%.

The Truancy scale consisted of 3 items asking how often students were (1) late for school, (2) skipping whole school day, and (3) skipping classes within school day. The participants indicated the frequencies for each of the three behaviors on a scale from 1-4, where 1 stands for none, and 4 refers to 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 represent participants’ overall levels of truancy, and in the complete data, the missing rate of this variable is 1%.

The measurement of conscientiousness in PISA 2012 is called Perseverance, which contains 5 items, tapping into three 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. Unlike the common Likert-type scale, where the larger the number, the higher the degree of agreement, on the Perseverance 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 a larger number corresponds to a higher degree of agreement.

Response rates in the complete data for the 5 items were between 64.1% to 64.5%, so the missing rates were around 35%. Respondents’ overall perseverance level was also given by PISA 2012, computed based on the 5 items using an IRT approach. The response rate of this variable in the complete data is 64.5%. The overall perseverance score was coded as missing when responses to at least 4 of the 5 items were missing. We excluded such cases before the random sampling. In the remaining dataset containing 313,172 participants, the missing rates for each of the 5 items dropped from more than 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 Perseverance. This gave us a sample of 33,500 participants in total. There were no missing values for any of the 5 perseverance items, and 4% of all cases have missing values for 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 sample for the analyses. The samples were identical for analyses of ME and the prediction of math and truancy by perseverance. Across the 9 cultural groups, the percentage of female ranged from 48.2% to 53.9%.

*Analyses*

To obtain a reasonable measurement model for the Perseverance scale to assess ME, we started with obtaining the factorial structure of it. The exploratory factor analysis indicated a two-factor solution, with the first two factor accounting for 73% of the variance. The first factor was loaded onto by only the positive items (i.e., “Exceed expectation”; “Continue to perfection”; “Remain interest”), with large positive loadings, and thus was named “Positive”, while the second one was called “Negative”, because only the negative items loaded positively 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. Therefore, we calculated the Cronbach’s alpha for each of the 9 groups using the items loadings onto the two factors separately. Most of the Positive and Negative scales showed acceptable reliability (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 measurement in Mplus 7.4 (Muthén & Muthén, 1998-2010), using multigroup structural equation modeling (SEM). According to the results of the factor analyses, we chose a correlated factor model for the multigroup SEM, whose diagram can be found in Figure 1. In the initial multigroup analysis, all parameters (i.e., loadings, and intercepts) were constrained equal across groups, as if the measurement is completely equivalent to all groups. However, a full invariance model is too stringent to achieve, and unnecessary, too. A partial invariant model with at least one metrically invariant item, on the other hand, is more likely to be found, and is believed to be sufficient for cross-group analyses, such as comparisons of groups means (Steenkamp & Baumgartner, 1998; Byrne, Shavelson, & Muthén, 1989). Therefore, based on the results given by Mplus for the constrained model, we freed some of the parameters in certain groups that had large modification indices (MIs), in order to obtain a partial invariant measurement model. Next, mean values of the two factors in the constrained and the partial invariant models for each of the groups were compared.

To predict math achievement and truancy respectively with the perseverance factors, we added a path model to the partial invariant model, while controlling for SES and gender. With the partial invariant measurement model included, we were able to predict the response variables with a relatively equivalent measurement, which is a prerequisite for meaningful comparisons across groups.

With the means of the latent factors, regression solution, and the estimated correlations between the latent factors and the response variables, we were able to compare self-reported conscientiousness, and its ability to predict educational outcomes across different cultures, while gender and SES were controlled for.

**Results**

Descriptive statistics of math, truancy, and perseverance can be found in Table 2, along with the zero-order correlations among these three variables within each of the 9 cultural groups. 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 relating negatively. This pattern was consistent what we found within each the 67 countries/regions, where higher perseverance scores were associated with better math performance, and lower truancy rates. However, such trends were reversed both across countries and across groups, where math and perseverance were found to correlate negatively (across groups: *r* = -.76; across countries: *r* = -.48), whereas positive correlations were found between truancy and perseverance (across groups: *r* = .67; across countries: *r* = .36).

The constrained measurement model, where all parameters were kept equal across group, yielded acceptable but not great model fit (*RMSEA* = 0.068, *χ2 (84)* = 1483.93, *SRMR* = 0.043). After some of the parameters with large MI were allowed to be freely estimated for some of the groups, the model fit improved (RMSEA = 0.048, *χ2 (70)* = 664.09, *SRMR* = 0.032), indicating the existence of measurement nonequivalence in the Perseverance scale.

For each group, the mean values for the latent factors can be found in Table 3, along with the changes of the latent means and the rankings for the means before and after some of the constraints were removed. The first group, Southern Europe, was set automatically as the reference group by Mplus. As shown in Table 3, compared with the constrained model where full measurement invariance was assumed, the partial invariant model led to different latent means. Southeast Asia had the greatest change in the means on the Negative factor. The mean dropped from 0.076 to -0.629, and the ranking became the ninth (i.e., the lowest), whereas in the constrained model, it was the second highest. Means of the Negative factor also increased by more than 100% for Latin America and Western Europe, after a partial invariant model was applied, although the magnitude of the change was rather small. The mean values of the Negative factor in Latin America and Southern Europe did not differ significantly. Compared with the Negative factor, smaller changes were observed on the Positive factor. Only the Northern America/Oceania group’s mean dropped by 80%. No great changes were found on the rankings.

Participants from the Middle East, scoring one of the lowest on math, and the highest on truancy (see Table 2), had the highest self-ratings on both perseverance factors. Participants from Nordic countries, who scored the third highest on math, and the fourth lowest on truancy, had the second highest estimated means for Negative, and the lowest for Positive, and thus in general rated themselves the most negatively on conscientiousness. The East Asia group had a similar pattern as the Nordic Country group, for the group generally rated themselves pretty negatively on conscientiousness (i.e., the third highest on Negative, and the third lowest on Positive), whereas scoring the highest among the 9 groups on both math and attendance. Overall, the Southeast Asia group seemed to have the most positive view of themselves on conscientiousness, as under the partial invariant model, the group was the second lowest on the Negative factor, and the second highest on the Positive one. In the North America/Oceania group, respondents rated themselves the second lowest on Negative, but in the middle (the fifth highest) on Positive, which was similar to what was observed in the Former Communist Countries group. The group of Western European countries with UK and Ireland, however, showed an opposite pattern. Participants in this group rated themselves as the second lowest on Positive, yet in the middle (the fourth highest) on Negative.

Standardized regression coefficients of the two factors for the prediction of overall math performance and overall truancy can be found in Tables 4-5. In most groups, the Negative factor, with greater magnitude of the standardized coefficients, turned out to be a much better predictor of math performance than the Positive factor. The East Asia group was the only exception, where the coefficients of the two factors were close in size (i.e., -.144 for Negative, and .112 for Positive). Moreover, the Negative factor was a significant predictor of math performance in all groups, while the Positive factor predicted math performance significantly only in three groups: Northern America/Oceania, Nordic Countries, and East Asia.

When the response variable was truancy, neither predictor was obviously better than the other. Although in 7 out of the 9 cultural groups, the magnitude of the standardized coefficients was greater for the Positive factor, the differences were rather small in most groups. In the Former Communist Countries group and the East Asia group, Negative was an insignificant predictor, while in Southeast Asia, the Positive factor failed to predict truancy. In all the other conditions, both factors significantly predicted the response variable.

Correlations between the latent factors and the two response variables (Table 6) were all in the same direction as the sample correlations (see Table 2): The Negative factor was negatively associated with math performance, and positively related to truancy, whereas the correlations between the Positive factor and the response variables exhibited a reversed pattern. The two perseverance factors were negatively correlated. Correlations were all significant except for those between the Positive factor and math performance in the Middle East group and the Southeast Asia group.