The effect of not using plausible values when they should be: An illustration using TIMSS 2007 grade 8 mathematics data

--- DRAFT June 1, 2010 ---

Ralph Carstens, IEA Data Processing and Research Center, ralph.carstens@iea-dpc.de
Dirk Hastedt, IEA Data Processing and Research Center, dirk.hastedt@iea-dpc.de

Abstract

Analyses using large-scale survey assessment databases such as those collected by TIMSS, PIRLS, ICCS, or PISA should use the data and the included plausible values (PV) as provided by the publishing organizations. Further, they should adhere to appropriate and intended computational procedures in the accompanying documentation to compute unbiased population estimates of student achievement that account for the administration of rotated instruments in which each students responds to a subset of items and not the entire item pool. Working with plausible values may seem overwhelming or cumbersome though and some researchers may be tempted to or actually do use analytical shortcuts or inappropriate scoring methods that lead to biased results and/or results that identify statistical significance in the comparisons of group-level statistics or regression analysis where the appropriate analytical procedure, given a survey's design, would not have.

Building on the convincing theoretical argument in favor of plausible values (PV) by von Davier, Gonzalez and Mislevy (2009), this paper seeks to illustrate the practical relevance of using the plausible values on the database file as intended in the context of the real-world, grade 8 mathematics data collected by TIMSS 2007. The paper addresses the adverse effects of incorrectly using the provided plausible values, more specifically using averages of plausible values or just one out of the five provided values, in the context of country-level and group-level comparisons. In addition, we use common IRT scoring methods and compare estimates based on these to the score estimates included in the TIMSS 2007 international reports (Mullis, Martin, & Foy, 2008).

The results show that inappropriate use of the plausible values or alternative scoring methods can entail a substantial departure from the estimates obtained from correct and intended analysis and consequently inferences to the studied populations. The findings are used to convince researchers to use the plausible values correctly and as designed, that there is no need to rely on computational shortcuts and that user friendly software to achieve this is readily available.

Keywords: *Plausible values, analytical shortcuts, imputation variance, scoring methods, incorrect use*

Introduction

In response to the increased demand of content coverage without increasing the per-student testing time, complex rotated test forms and the plausible value methodology were introduced in large-scale surveys, starting with NAEP in the U.S. (see Mislevy, Johnson & Muraki, 1992), to arrive at better population and group-level estimates of achievement. Each randomly drawn value is considered a representative value from the distribution of possible (plausible) scores for all respondents who generate a specific pattern of item responses (see Mislevy, Beaton, Kaplan & Sheehan, 1992). Before this, percent correct or, at best, maximum likelihood or EAP score estimators have been used in international assessments.

Recent research by von Davier, Gonzalez and Mislevy (2009) has again clearly illustrated the theoretical foundations, scientific utility, and justification of plausible values. Using simulated data generated from known distribution parameters and student "true" scores for two groups, the authors compared population and group-level estimates based on two commonly used score estimators, *Expected a posteriori estimation* (EAP) and *Warms' weighted maximum likelihood estimation* (WML), with those based on five plausible values. For the latter and a variant of the EAP method, the scaling combined students' cognitive item responses with information about their background, which reflects the "condition" they are in (in this case: SES), to measure and incorporate measurement error. This process is known as "conditioning". While the plausible values method added a random component to each individual estimate, the authors concluded that, from the point of view of groups, "they add exactly the right amount of variability to make the distribution of the plausible values in the group match the distribution of the true values in the group" (page 35). In a series of comparisons and considering varying test lengths, only the correct application (labeled as "PV-R" in von Davier et al., 2009) of the plausible value methodology "recovered" the generating parameters with a high degree of precision and for each group separately. In contrast, the two studied point estimate methods (EAP and WML) demonstrated unwelcome effects that adversely affected their utility for accurate overall and group-level statistics. In summary, the aggregation of WML score estimates overestimated the distribution's variance and the aggregation of EAP estimates underestimated these (see also concurring findings in Mislevy, Beaton, Kaplan & Sheehan, 1992; Wu, 2005). Similarly, aggregations of incorrectly averaged plausible values (labeled as "PV-W" in von Davier et al., 2009) underestimated variances. In another simulation (Monseur & Adams, 2009), only the PVs with conditioning were found to recover generating within and between school variance components, intra-class correlations, and latent correlations accurately.

While working with plausible values may seem overwhelming or cumbersome at first, comprehensive guidance, literature, and software that allows for design and model consistent analysis are publicly available, readily accessible, and typically free of charge (in the context of TIMSS 2007, see Foy & Olson, 2009 and IEA, 2009/2009a). However, experiences from analysis workshops or the review of research papers submitted for publication show that still a good number of researchers analyze the data

incorrectly and try to resort to shortcuts like averaging plausible values or using just one of them to simplify the calculation of means and variances or other analyses. Some others may still not see the advantage or psychometric validity of plausible values and prefer to work with other Item Response Theory (IRT) or non-Bayesian score estimators resulting from re-scaling of the original response data.

Considering the above, this paper seeks to investigate the impact of not using the provided PVs data from a practical point of view when working with real-world TIMSS 2007, which employed a complex matrix-sampling booklet design, IRT scaling, and plausible value methodology to obtain proficiency scores for all students. A comprehensive account of the theoretical framework, methodology, and the computational procedures for TIMSS 2007 can be found in Olsen, Martin and Mullis, 2008.

The paper addressed the following research questions:

(1) What is the effect on estimated means, standard deviations and standard errors when, instead of using the internationally provided plausible values correctly and as designed, i) just one of them is used, ii) the 5 international plausible values are averaged up, or iii) other IRT estimates are used?

(2) Are the methods changing any inferences, i.e. international comparisons of population means between countries and, within countries, between girls and boys and between subgroups defined by the number of books at home? If changes are observed, are these consistently directed?

(3) What is the effect of scale length on these changes?

The fundamental difference between our study and the previous work (von Davier et al., 2009; Wu, 2005) is the absence of “true” student scores in the TIMSS data in contrast to the helpful presence of these in simulations. The social and behavioral sciences involve abstract (latent) qualities and measures of these. TIMSS is no exception and information about a quality such as mathematics achievement can only be obtained through the interaction of students with assessment items associated with the latent quality. The observed manifest responses are then used to estimate a measure of the quality. Consequently, the TIMSS IDB includes “estimates” of plausible student achievement rather than “true” values, which are unknowable. The TIMSS data is authentic data with all the noise and peculiarities that occur in real testing situations and not simulated data that perfectly fits a theoretical model. Nonetheless, the previous research on the behavior and characteristics of plausible values led us to consider them as the closest possible approximation of a student’s true ability given the study’s constraints, yet acknowledging that they are still estimates just as those produced by any other method, i.e. including uncertainty and error. The plausible values, as included in the TIMSS IDB, are hence used as the “reference” method and all other methods presented in this paper are gauged against it.

Further, plausible values were designed as intermediate estimates and not individual-level test scores. Our main interest is hence in the utility of the various methods for population and group-level estimates, the main policy-relevant interest in TIMSS and other large-scale studies that seek to report on the

school system as a whole, rather than on individuals.

Methodology

The analysis in this paper is based entirely on publicly available data and materials released by the IEA, more specifically the TIMSS 2007 international database (IEA, 2009), which is accompanied by a user guide (Foy & Olson, 2009). For the illustrative purpose in this paper, we have limited the analysis to the background and mathematics achievement data collected from students at the 8th grade in all 49 countries that were i) full participants in TIMSS 2007 at that grade level and ii) reported, i.e. excluding any benchmarking participants and Mongolia. Individual data files from these populations were merged using the IEA IDB Analyzer (IEA, 2009a) and raw data files as well as analysis databases were derived from this merged database with the help of SPSS.

The estimators studied in this paper are presented in Table 1. They comprise three different ways of using plausible values (one correct: PV-R, two incorrect: PV-1, PV-W) and two alternative methods (EAP and WML).

[Take in Table 1 about here]

The acronym "PV-R" is used for the correct ("right") application of the five plausible values for the overall mathematics achievement from the TIMSS IDB (variables BSMMAT01 to BSMMAT05), originally computed using the ETS MGROUP software, the predecessor of the current DESI software (ETS, 2010). For the two incorrect applications of plausible values, "PV-1" is used for the first (and only the first) PV, "PV-W" for the simple average of the aforementioned five PVs. Von Davier et al. (2009) studied another method, the Expected a posteriori estimation (with group membership), with acronym "EAP-MG". These four methods are *Bayesian* estimators and take the student's group membership, defined by the combination of a student's responses to the background questionnaire, into account. Estimates are "conditioned" on this prior information about a student's group and integrated with the observed item responses to form a posterior distribution from which estimates are derived. As a consequence, all four methods are capable of estimating a score for a student that did not respond to the assessment item at all, resulting in 0 % missing data, a very desirable aspect for the later analysis of data. It must also be noted that only the PV-R method takes the measurement uncertainty of a single (point) value into account by selecting five PVs. This distinct feature will prove to be important later on when we discuss the aggregated group-level results that include an imputation variance component for PV-R. All three ways of using the PVs employ random sampling from the posterior distribution and produce smooth distributions, a helpful effect during the estimation of, for examples, percentiles. In contrast, the EAP-MG method takes the mean of the posterior distribution as the estimate instead of a random draw.

The EAP-MG method is not studied further in this paper for three key reasons: 1) it is unduly demanding to compute for the mere purpose of illustration[1], 2) any applied researcher who is able to set up and operate the conditioning in, for example, the DESI software (ETS, 2010) will most likely resort to plausible values anyway given their added value, and 3) the average of the five PVs (PV-W), which are drawn from a posterior distribution with a known mean μ, are a suitable proxy for the mean of the posterior distribution itself. So whenever we discuss the characteristics of the PV-W method in what follows, we implicitly also get an arguably uncertain yet close-enough idea about the characteristics of the EAP-MG method.

Three additional estimators were computed for the intended comparisons using the Parscale 4.1 software (Muraki & Bock, 2003; see also Du Toit, 2003): "EAP", "WML" and "MLE". For this, item information and calibrated parameters were extracted from the released IDB documentation, transformed to the format expected by Parscale, and used as anchor values (i.e., the calibration was skipped). The three-parameter logistic (3PL) model was used for multiple-choice items, the two-parameter (2PL) model for dichotomously scored constructed-response items, and the generalized partial credit model for polytomous constructed-response items (see Chapter 11 in Olsen, Martin, & Mullis, 2009; for a general description of models, see Van der Linden & Hambleton, 1997). Consistent with the international scaling, we treated omitted and not-reached responses as incorrect. Scores were estimated and extracted from Parscale output and transformed to the metric of the IDB's plausible values by using the average of the reported transformation constants across the five values (see Exhibit 11.15 in Olson, Martin, & Mullis, 2008).

None of these three methods consider group membership or the imputation variance and all produced point estimates (i.e., a single value per student). The EAP method is Bayesian and therefore produced the smallest amount of missing data (0.2 %) for the 469 students in the IDB that did not complete the cognitive portion of the test. While the WML estimator should theoretically be able to produce an estimate for all students that worked on at least a portion of the cognitive test (and does precisely that in simulations), it exhibited a substantial amount of missing data (5.9 %). It turned out that authentic student response vectors in TIMSS 2007 can cause the Newton-Raphson algorithm used by the Parscale software to drift away from convergence rather than towards it in presence of the guessing parameter in the 3PL. When the algorithm detected that it will not be able to get a stable estimate of the student score, it stops and assigns a missing value of 999. Nonetheless, some researchers may unquestioningly consider 5% of missing data to be "ignorable". In that light, we decided to retain the estimator in the following comparisons despite the possibility of biased estimates in case that the above problematic response vectors are not observed by chance alone. Since the MLE method is known to be biased (Warm, 1989), not commonly used in professional settings, and incapable of handling so-called zero (all incorrect) or perfect (all correct) response vectors by design, it produced a recognizably high

[1] Also due to time constraints.

amount of missing data (13.3 %) due to zero/perfect response vectors and the said convergence problems. The method was excluded from the comparison on all of these accounts.

Means, standard deviations, and their standard errors were computed for all 49 countries and for all score methods independently using the SPSS macros (JB_PV and JB_GEN) included with the IEA IDB Analyzer (IEA, 2009a) and accounted for the sample design by using the jackknifing (JK2) algorithms and final student weights (TOTWGT). Deciles were computed for each score method separately and the sum of squared differences across the nine deciles was computed per country. For the multiple comparisons of country means, no adjustments (e.g., Bonferroni) were made to account for the concurrent comparisons. This is consistent with the international reporting strategy (see Exhibit 1.2 in Mullis, Martin, & Foy, 2008).

For the regression analysis of mathematics achievement on gender, the variable ITSEX in the IDB was recoded to a 0/1 dummy variable (0 = Girls). For the regression of math achievement on books at home, the IDB variable BS4GBOOK was used to derive four predictor variables using a coding scheme known as “backward difference coding” (for the values, see Table 2; for the rationale, see UCLA, n.d.). In this coding system, the mean of the dependent variable for one level of the categorical variable is compared to the mean of the dependent variable for the prior adjacent level. In our analysis, this translates to the change in math achievement between level 1 and 2, 2 and 3, 3 and 4, and 4 and 5.

[Take in Table 2 about here]

All regression parameters were calculated independently for all score methods with the IEA IDB Analyzer (macros JB_REGP and JB_REG) to correctly estimate the standard error, again using jackknifing as well as, final student design and replicate weights. Significance was determined based on an absolute value of the t-test statistic being greater than 1.96, so using a normal distribution to identify the critical value for a 95 percent confidence level, also commonly referred to as the ".05 level of (statistical) significance."

Findings and Discussion

To address RQ 1, Table 3 presents the relative size of means, standard errors, and standard errors in comparison to the PV-R reference method.

[Take in Table 3 about here]

While there is no change for the PV-W method by design and only random variation in the means estimated using PV-1 (reflected in the average of percent values), country means for EAP and WML change by as much as 14.8 percent for Qatar with an average absolute change of 2.2% for EAP and 2.8% for WML. The most noteworthy changes occur for the standard error which can “shrink” by up to

50% (Qatar/PV-W: -50.8%; Saudi Arabia/EAP: -53.4%), reflecting the missing imputation variance component when analyzing a point estimate rather than a set of five values. On average, the largest reduction can be observed for the standard errors estimated by the EAP and WML methods (about 17% each) in contrast to the smaller yet consistent changes observed for the PV-W and PV-1 methods (about 7% each). Standard deviations, again mostly consistent with the stated theory and simulations, are underestimated to some degree by all methods but most pronounced by the EAP method, least by PV-1 since a single plausible value still represents the same spread of values as the 5 plausible values do.

Analogous to Table 18 in von Davier et al. (2009), Table 4 below presents the sum of squared differences over the nine distribution deciles estimated for each country and method separately. This gives us an idea of how precise different estimates are over the distribution. As can be seen, the PV-1 method only shows trifling deviations from the deciles estimated under PV-R due to the random component (including the extreme 1^{st} and 9^{th} deciles), whereas the PV-W estimates are, on average, 5 times larger due to the premature averaging. Both EAP and WML show a substantially higher distortion of the distribution (highest for Qatar/EAP and Chinese Taipei/WML), which is also largest for the extreme percentiles (not shown in Table 4).

[Take in Table 4 about here]

One of the questions that we had (RQ 2) was if any country comparisons would change if an inappropriate or alternative method is used. For this purpose, country means and standard errors of the means were calculated as described above. The following two tables 5a and 5b show the internationally used method that includes all five plausible values and contrasts them to the comparisons based on the average of the five plausible values and the first plausible value only. Since a multiple comparisons table would be symmetric (see Exhibit 1.2 in Mullis et al., 2008), we used the area above the diagonal in Table 5a to contrast the comparisons based on the first plausible value only and the correct method, the area below the diagonal to contract the comparisons based on the average of the plausible values with the correct method. Plus "+" signs indicate comparison for which the difference between the row and column country is significant (higher or lower) for both methods. Cells marked as "o" indicate a nonsignificant difference, again for both methods, whereas those marked with "o+" indicate that an nonsignificant difference became statistically significant under the contrast method.

[Take in Table 5a about here]

As can be seen, most of the comparisons are not affected as expected. However, some countries comparisons are affected. When, for example, comparing Ukraine and Romania based on the first plausible values only, Romania would score higher than Ukraine – but the comparison remains nonsignificant. Also, when using the first plausible value only, the comparisons between England and Czech Republic, Syria and Algeria, and Algeria and Colombia became statistically significant whereas originally they were not. When using the average of the plausible values as the basis for country

comparisons, of course no changes in the order of the countries can be observed because the mean is computed equivalent to the PV-R method. However, four comparisons became statistically significant (Bahrain vs. Georgia, Egypt vs. Bahrain, Algeria vs. Syria, and Oman vs. Morocco) although they were not under the correct method.

The next table 5b shows the comparisons of the EAP (above the diagonal) and WML scores (below the diagonal) with the correct comparisons in the same way as above.

[Take in Table 5b about here]

Here, the number of comparisons that changed is larger. When using EAP scores instead of the five plausible values as the basis for comparisons, five countries change order, but none of these comparisons becomes statistically significant. The biggest differences between the EAP scores and the plausible values can be seen for top and especially low performing countries. Chinese Taipei "looses" about 10 score points on the EAP scores. The 10 lowest performing countries gain more than 10 score points on the EAP scores with the lowest performing country, Qatar, gaining as much as 45 score points on the EAP scale. 11 comparisons become statistically significant and 10 comparisons become nonsignificant (indicated by "o+").

The contrasting of the WML scores shows even more difference to the PV-R method. Five countries change the ranks and two countries (Bulgaria and Lithuania) even "loose" two rank positions. But again, all of these changes occur within the statistically expected error. Again, the most severe changes of country means occur for the highest and lowest performing countries. The five top performing countries each loose more than 20 score points on the WML scale with Chinese Taipei losing more than 35 score points. The 15 lowest performing countries on the other hand gain more than 10 score points with Qatar gaining more than 38 score points. 18 originally statistically nonsignificant comparisons become significant and 16 comparisons changed to statistically nonsignificant ones.

In summary we can say that in terms of the country averages and consequently the ranking, the average of the five plausible values gives of course the same result as the correct method, simply because this is how the method is defined (see equation 5.17 in Little & Rubin, 2002). The results based on the first plausible value only comes close but the other two methods (EAP and WML) show severe differences, although looking at statistically significant comparisons only, no rank order changes occur. When looking at the differences of comparisons becoming statistically significant or losing the statistical significance, the method that makes use of the first plausible value only comes closest to the international results, the average comes next and the other two methods show severe differences.

Pursuant to RQ 2, we looked at the gender comparisons when using the different methods. The next Table 6 shows the difference between average scores of boys and average score of girls based on the correct methods that makes use of all five plausible values, based on the first plausible value only, based

on the average of the five plausible values and based on the EAP and WML scores. A negative number indicates that on average the boys score higher in that country, a positive number that the girls score higher. All statistically significant differences are highlighted.

[Take in Table 6 about here]

As can be seen, most comparisons do not change when using a method other than PV-R. Interestingly, the numbers – especially for the EAP and WML scores – are much closer to the differences observed with the PV-R method than the differences in country means. As expected, none of the comparisons change their direction, which would mean that on one scale the boys outperform the girls and on the other it is the opposite. The biggest differences of the boy-girl differences can be observed for Oman and Qatar for the EAP and WML scores. Most differences become smaller when using the EAP or the WML scores. But also the standard errors of the differences become smaller in nearly all cases when not using the correct method. Consequently, four of the boy-girl comparisons become statistically significant when using the first plausible value only. Five of the comparisons become statistically significant for the average of the five plausible values. The comparisons of the EAP scores result in the same statistically significant comparisons than the correct method. Three comparisons become statistically significant when comparing the WML scores.

Using “books at home” as a predicting variable, we also analyzed smaller subsets of students for the mathematics scale as presented in Table 7 but did not find the differences to being different in number or magnitude from the ones presented for the grouping by gender above.

[Take in Table 7 about here]

Finally we studied one of the subscales of the TIMSS mathematics test (RQ 3). Since also von Davier et al. (2009) found that the fewer the number of items in a scale, the bigger the influence of a different scoring and analysis method, we chose to analyze one of the shortest subscales in TIMSS. The geometry subscale included only 47 items out of the 214 mathematics items in TIMSS 2007. Table 8 displays the comparison of the international results for the Geometry subscale and the results when averaging the five plausible values or taking only the first plausible value into account.

[Take in Table 8 about here]

In line with previous research, we found more deviations for the geometry scale than for the mathematics scale. We did not only find statistically nonsignificant differences that became statistically significant but also comparisons that changed in the opposite direction, i.e. from being significant to becoming nonsignificant. By using only the first plausible value, eight comparisons became statistically significant and nine changed to non-significant. For the average of the five plausible values nine of the comparisons became statistically significant although they were not.

In terms of the ranking, for the averaged plausible values no changes occur since this is not possible by definition. For the first plausible value, three pairs of countries change their ranks (Norway and Cyprus, Morocco and Indonesia and Oman and Kuwait) and the group of Tunisia, Israel and Jordan changed ranks. None of these changes were statistically significant.

For the geometry scale, we focused on the two incorrect way of using PVs only and did not rescale the data with the other methods described. Our assumption (not proved) is that a the number of changed comparison increases in a similar way for the EAP and WML methods.

Conclusions and Implications

We found that the findings reported by von Davier et al. (2009) have a sizable, practical importance for the analysis of TIMSS data. Although the TIMSS data, due to the length of the scales, seems to be quite robust towards using an inappropriate analysis method, in areas where results are relatively close to each other, the usage of an inappropriate analysis method resulted in differences to be considered as statistically significant although they shouldn't be. Most changes were in the direction that a statistically not significant difference became significant, usually because the "wrong" methods presented in this paper underestimate variances and standard errors. While the differences observed for the PV-W and PV-1 methods mostly result from the absence of the imputation variance component, the higher number of deviations for the EAP and WML methods result from the both the missing imputation variance as well as the absence of the "conditioning" on students' group memberships, which leads to a substantial "distortion" of the score distribution in comparison to PV-R. Further, shorter scales – as for example Geometry in TIMSS – seem to be more affected by using an inappropriate method of analysis. We also analyzed smaller subsets of students for the mathematics scale (using "books at home" as a predicting variable), but did not find these differences to be different in number of magnitude from the ones presented for the grouping by gender. It would be of interest if the group size creates significantly more inconsistencies when they become much smaller and whether plausible values are still appropriate estimators for very small sub-groups.

In summary, analysts should not rely on computational shortcuts (PV-1 or PV-W) or alternative methods (such as EAP or WML), which require complete item data for each student and are therefore inappropriate under matrix-sampling designs as that used in TIMSS. Besides the resmapling varience estimation methods, the sophistication of many large-scale databases, including IEA/TIMSS, IEA/PIRLS, US/NAEP, OECD/PISA, and IEA/ICCS requires researchers to apply the PV procedures correctly to estimate the imputation variance (uncertainty) in population and group-level estimates. Ignoring this entails the risk of producing biased estimates, underestimates of standard errors, inferences, or artifacts that are not supported by the data.

Instead of using software that is incapable of doing this correctly (such as SPSS 'out-of-the-box'), the

procedures developed in the relevant literature and user guides should be followed. Even though this requires analysts to use specialized software, there is no alternative. Luckily, several programs are capable of conducting various types of analyses with complex sample and imputed survey data and many of them are easy to use. Examples are the latest versions of HLM, Westat's WesVar, AIR's AM software, StataCorp's Stata, the IEA's free SPSS-based IDB Analyzer (IEA, 2009a), or ACER's free SPSS Add-In (Gebhardt & Daraganov, 2010). They provide analysts with appropriate and accessible procedures for analyzing the information contained in many large-scale survey databases.

References

Du Toit, M. (Ed.)(2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Scientific Software International, Inc., Lincolnwood, IL.

ETS (2010). *DESI. Direct estimation software interactive v3.2.7*. (computer software). ETS Educational Testing Service, Princeton, NJ.

Foy, P. & Olson, J.F. (eds.) (2009). *TIMSS 2007 User Guide for the International Database*, Boston College, Chestnut Hill, MA.

Gebhardt, E., & Daraganov, A. (2010), SPSS Replicates Add-In (version 7.1) for SPSS 18 and older (computer software), Australian Council for Educational Research, Melbourne. Retrieved 1 June 2010 from https://mypisa.acer.edu.au/index.php?option=com_content&task=view&id=87&Itemid=468

IEA (2009). *TIMSS 2009 International Database*. Retrieved April 15, 2010 from http://timss.bc.edu/timss2007/idb_ug.html; alternative download available from IEA's study data repository at http://rms.iea-dpc.org/

IEA (2009a). *IDB Analyzer II* (computer software), IEA Data Processing and Research Center, Hamburg.

Little, R.J.A, & Rubin, D.B. (2002) *Statistical Analysis with Missing Data*. Wiley, New York.

Mislevy, R.J., Beaton, A.E., Kaplan, B., & Sheehan, K.M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29, 133-161.

Mislevy, R.J., Johnson, E.G., & Muraki, E. (1992). Scaling Procedures in NAEP. *Journal of Educational Statistics*, *17*(2), 131-154.

Monseur, C., & Adams, R.J. (2009) Plausible Values: How to Deal with Their Limitations. *Journal of Applied Measurement*, *10*(3), 320-334.

Mullis, I.V.S., Martin, M.O., & Foy, P. (2008). *TIMSS 2007 International Mathematics Report. Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*, Boston College, Chestnut Hill, MA.

Muraki, E., & Bock, R.D. (2003). *PARSCALE 4 for Windows: IRT based test scoring and item analysis*

for graded items and rating scales (computer software). Scientific Software International, Inc., Lincolnwood, IL.

Olson, J.F., Martin, M.O., & Mullis, I.V.S. (Eds.)(2008). *TIMSS 2007 Technical Report*, Boston College, Chestnut Hill, MA.

UCLA (n.d.). *Regression with SAS. Chapter 5: Additional coding systems for categorical variables in regression analysis*, UCLA: Academic Technology Services, Statistical Consulting Group. Retrieved May 19, 2010 from http://www.ats.ucla.edu/stat/sas/webbooks/reg/chapter5/sasreg5.htm

Van der Linden, W. J., & Hambleton, R.K. (Eds.)(1997). *Handbook of modern item response theory*. New York/Berlin/Heidelberg: Springer.

Von Davier, M., Gonzalez, E., & Mislevy, R.J. (2009). What are plausible values and why are they useful? *IERI Monograph Series. Issues and Methodologies in Large-Scale Assessments, Vol. 2,* 9-36.

Warm, T.A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*, 427-450.

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation, 31*(2-3), 114-128.

Table 1. Scoring methods

Acronym	Name	Software	Accounts for group membership	Handles zero/perfect vectors	Type of estimate	Bayesian	Accounts for uncertainty	Complete data	Percent of scores not estimated	Studied in von Davier, Gonzalez, Mislevy (2009)	Studied in current project
PV-R	Plausible values - correct ("right")	MGROUP/DESI (TIMSS 2007 IDB)	Yes	Yes	Random draws (5) from posterior distribution	Yes	Yes (imputation variance component)	Yes, including students lacking cognitive item responses	0 %	Yes	Yes
PV-1	First plausible value only	MGROUP/DESI (TIMSS 2007 IDB)	Yes	Yes	Random draw (1) from posterior distribution	Yes	No	Yes, including students lacking cognitive item responses	0 %	No	Yes
PV-W	Average of 5 plausible values ("wrong")	SPSS, based on MGROUP/DESI draws	Yes	Yes	Based on random PV-R draws	Yes	No	Yes, including students lacking cognitive item responses	0 %	Yes	Yes
EAP-MG	Expected a posteriori estimation with group membership	DESI (not computed, PV-W is proxy)	Yes	Yes	Point estimate (mean of posterior distribution)	Yes	No	Yes, including students lacking cognitive item responses	0 %	Yes	No
EAP	Expected a posteriori estimation	Parscale 4.1	No	Yes	Point estimate (mean of posterior distribution)	Yes	No	No. Students lacking item data are not estimated	0.2 % (469 of 220,788)	Yes	Yes
WML	Warms' weighted maximum likelihood estimation	Parscale 4.1	No	Yes, theoretically	Point estimate (maximum of likelihood function)	No	No	No. Students lacking item data and certain response vectors are not estimated	5.9 % (12,991 of 220,788)	Yes	Yes
MLE	Maximum likelihood estimation	Parscale 4.1	No	No	Point estimate (maximum of likelihood function)	No	No	No. Students lacking item data and zero/perfect as well as certain response vectors are not estimated	13.3 % (29,342 of 220,788)	No	No

Table 2. 'Backward difference' regression coding scheme for 4 predictors based on BS4GBOOK (number of books at home, ordinal)

BSG4BOOK	BOOK_BD12	BOOK_BD23	BOOK_BD34	BOOK_BD45
1	-0,8	-0,6	-0,4	-0,2
2	0,2	-0,6	-0,4	-0,2
3	0,2	0,4	-0,4	-0,2
4	0,2	0,4	0,6	-0,2
5	0,2	0,4	0,6	0,8

Table 3. Relative size of means, standard errors and standard deviations in percent of PV-R method (overall mathematics) by country
Sorted by country name

Row Labels	Mean				Mean SE				SD			
	PV-1	PV-W	EAP	WML	PV-1	PV-W	EAP	WML	PV-1	PV-W	EAP	WML
Algeria	0,2%	0,0%	1,8%	2,8%	-30,3%	-32,9%	-43,6%	-36,6%	-1,4%	-11,0%	-0,1%	17,1%
Armenia	-0,1%	0,0%	-0,9%	-2,0%	-1,9%	-1,7%	-10,2%	-24,7%	-0,8%	-4,4%	2,6%	1,9%
Australia	-0,1%	0,0%	-0,5%	-1,4%	-1,5%	-5,9%	-7,3%	-32,8%	0,1%	-3,4%	5,3%	-1,5%
Bahrain	0,2%	0,0%	2,2%	2,7%	-15,8%	-15,2%	-22,8%	-13,3%	-1,1%	-5,8%	-10,4%	-0,9%
Bosnia and Herzegovina	0,0%	0,0%	-0,1%	0,4%	-0,7%	-2,5%	-8,8%	-8,7%	-0,8%	-4,9%	1,1%	2,9%
Botswana	0,0%	0,0%	4,5%	4,7%	-12,6%	-14,6%	-36,3%	-25,7%	-2,1%	-6,7%	-16,2%	-0,3%
Bulgaria	-0,1%	0,0%	0,4%	-0,8%	-2,2%	-3,7%	-12,7%	-13,3%	-1,1%	-3,7%	-5,7%	-4,2%
Chinese Taipei	-0,2%	0,0%	-1,7%	-6,0%	-4,3%	-3,7%	-11,3%	4,4%	-1,3%	-2,9%	-4,6%	7,0%
Colombia	0,0%	0,0%	3,1%	3,3%	-0,9%	-4,8%	-27,9%	-21,3%	0,2%	-6,0%	-12,1%	-0,4%
Cyprus	-0,1%	0,0%	0,1%	-0,1%	-4,8%	-2,7%	-9,2%	-16,3%	-0,7%	-4,2%	-3,2%	-2,7%
Czech Republic	-0,2%	0,0%	-0,7%	-1,1%	-3,3%	0,7%	0,6%	-12,1%	-2,6%	-3,6%	6,4%	-1,7%
Egypt	0,3%	0,0%	3,9%	3,7%	-4,5%	-4,2%	-31,7%	-23,4%	-1,1%	-5,6%	-17,0%	-8,3%
El Salvador	-0,1%	0,0%	6,8%	6,3%	-3,4%	-4,3%	-36,2%	-23,0%	-1,9%	-8,4%	-20,3%	-0,1%
England	0,0%	0,0%	-0,4%	-1,7%	-0,6%	0,3%	-2,4%	-14,9%	-0,8%	-2,5%	4,4%	0,2%
Georgia	0,1%	0,0%	2,3%	2,1%	-1,0%	-5,0%	-25,2%	-17,7%	-1,3%	-5,3%	-12,8%	-5,6%
Ghana	0,3%	0,0%	13,9%	12,1%	-4,9%	-2,6%	-45,7%	-34,0%	-1,7%	-8,7%	-34,4%	-14,7%
Hong Kong, SAR	-0,3%	0,0%	-1,0%	-5,1%	-2,0%	-0,3%	-5,5%	-0,6%	-1,3%	-3,1%	0,5%	2,2%
Hungary	-0,2%	0,0%	-0,6%	-2,2%	-1,6%	-2,5%	-5,1%	-21,5%	-0,4%	-3,0%	3,6%	-4,4%
Indonesia	-0,1%	0,0%	2,5%	2,8%	-3,1%	-2,7%	-20,8%	-15,0%	0,1%	-5,3%	-11,7%	-3,3%
Iran, Islamic Republic of	0,0%	0,0%	2,0%	2,1%	-3,6%	-4,7%	-18,1%	-19,0%	-0,2%	-5,2%	-8,3%	-2,2%
Israel	-0,1%	0,0%	0,4%	-0,7%	-2,1%	-1,8%	-12,2%	-13,1%	0,0%	-3,8%	-4,4%	-2,5%
Italy	-0,2%	0,0%	-0,4%	-0,2%	-3,7%	-2,9%	-3,3%	-14,5%	-0,4%	-4,2%	3,7%	0,3%
Japan	-0,2%	0,0%	-1,0%	-3,9%	-14,3%	-14,4%	-20,1%	-24,7%	-1,1%	-3,4%	1,2%	-0,9%
Jordan	0,1%	0,0%	1,8%	1,3%	-1,9%	0,1%	-20,9%	-19,5%	-0,9%	-3,9%	-11,4%	-7,5%
Korea, Republic of	-0,3%	0,0%	-1,4%	-5,4%	-8,7%	-12,2%	-19,9%	-3,6%	-0,6%	-3,4%	-2,7%	2,9%
Kuwait	-0,1%	0,0%	5,8%	5,6%	-5,2%	-2,2%	-35,3%	-22,6%	0,8%	-8,4%	-19,1%	-1,0%
Lebanon	0,3%	0,0%	-0,2%	0,5%	-3,9%	-1,9%	-9,0%	-13,9%	-2,3%	-5,0%	1,6%	1,3%
Lithuania	-0,1%	0,0%	-0,6%	-1,7%	-4,6%	0,0%	-0,7%	-17,0%	-1,8%	-3,3%	4,8%	-3,3%
Malaysia	0,0%	0,0%	-0,5%	-0,5%	-0,5%	1,4%	-4,9%	-11,2%	-0,9%	-3,6%	3,6%	1,8%
Malta	-0,1%	0,0%	-0,2%	-0,8%	-28,3%	-31,4%	-20,7%	-12,0%	-0,8%	-2,9%	-0,7%	0,1%
Morocco	0,3%	0,0%	3,2%	3,6%	-7,3%	-1,6%	-19,1%	-7,4%	-1,2%	-6,5%	-12,5%	1,4%
Norway	-0,2%	0,0%	-0,6%	0,3%	-3,0%	-3,2%	1,5%	2,3%	-1,2%	-4,9%	9,1%	11,8%
Oman	0,7%	0,0%	5,2%	5,0%	-15,5%	-14,7%	-39,1%	-32,6%	-1,5%	-5,5%	-21,0%	-8,7%
Palestinian National Authority	0,3%	0,0%	6,2%	5,5%	-4,3%	-2,9%	-32,1%	-23,6%	-0,8%	-5,3%	-22,9%	-11,3%
Qatar	0,5%	0,0%	14,8%	12,5%	-47,1%	-50,8%	-47,3%	-30,5%	-0,4%	-8,8%	-36,0%	-17,1%
Romania	0,2%	0,0%	0,5%	-0,6%	-1,5%	-0,8%	-9,5%	-11,8%	-2,3%	-3,5%	-6,0%	-7,1%
Russian Federation	-0,2%	0,0%	-0,7%	-2,1%	-1,4%	-5,2%	-6,9%	-19,9%	-0,4%	-3,4%	3,5%	-4,3%
Saudi Arabia	0,5%	0,0%	8,9%	7,9%	-23,8%	-26,3%	-53,4%	-38,9%	-1,1%	-9,2%	-25,0%	-3,1%
Scotland	-0,1%	0,0%	-0,3%	-0,8%	-2,7%	-2,0%	-3,6%	-12,5%	-1,4%	-3,3%	4,5%	-0,4%
Serbia	0,0%	0,0%	-0,4%	-1,2%	-7,2%	-4,4%	-8,4%	-13,3%	-0,9%	-3,4%	-0,4%	-4,3%
Singapore	-0,2%	0,0%	-1,3%	-5,1%	-2,1%	-2,3%	-4,6%	2,9%	-0,6%	-2,7%	-0,4%	3,1%
Slovenia	-0,1%	0,0%	-0,6%	-0,7%	-5,4%	-3,2%	-1,2%	-11,7%	-0,4%	-4,0%	6,8%	-0,5%
Sweden	0,0%	0,0%	-0,7%	-0,5%	-7,4%	-7,8%	-5,5%	-9,3%	-1,5%	-4,3%	8,6%	6,7%
Syria, Arab Republic of	0,3%	0,0%	2,3%	2,7%	-9,5%	-10,5%	-28,0%	-23,6%	-0,4%	-6,4%	-9,9%	0,4%
Thailand	0,0%	0,0%	0,5%	0,1%	-1,1%	-0,7%	-6,7%	-18,5%	-0,1%	-4,0%	-3,4%	-4,3%
Tunisia	0,0%	0,0%	0,3%	1,5%	-3,7%	-3,2%	-11,3%	-5,1%	-0,6%	-7,1%	2,1%	9,2%
Turkey	0,0%	0,0%	1,5%	-0,5%	-1,5%	-2,6%	-15,1%	-25,0%	-0,9%	-3,4%	-8,6%	-10,9%
Ukraine	-0,1%	0,0%	0,1%	-0,2%	-0,6%	-2,0%	-12,4%	-13,5%	0,0%	-3,6%	-2,5%	-3,3%
United States	0,0%	0,0%	-0,7%	-1,6%	-2,0%	-1,6%	-1,8%	-14,3%	-1,3%	-3,5%	6,7%	0,6%
Low	-0,3%	0,0%	-1,7%	-6,0%	-47,1%	-50,8%	-53,4%	-38,9%	-2,6%	-11,0%	-36,0%	-17,1%
High	0,7%	0,0%	14,8%	12,5%	-0,5%	1,4%	1,5%	4,4%	0,8%	-2,5%	9,1%	17,1%
Average	0,0%	0,0%	1,6%	0,9%	-6,6%	-6,6%	-17,0%	-16,9%	-0,9%	-4,9%	-5,4%	-1,4%
Min (absolute)	0,0%	0,0%	0,1%	0,1%	0,5%	0,0%	0,6%	0,6%	0,0%	2,5%	0,1%	0,1%
Max (absolute)	0,7%	0,0%	14,8%	12,5%	47,1%	50,8%	53,4%	38,9%	2,6%	11,0%	36,0%	17,1%
Average (absolute)	0,2%	0,0%	2,2%	2,8%	6,6%	6,7%	17,1%	17,3%	1,0%	4,9%	8,7%	4,3%

Table 4. Sums of squared differences to PV-R deciles (overall mathematics) by country

Sorted by country name

Country	PV-1	PV-W	EAP	WML
Algeria	10	249	365	2353
Armenia	16	38	178	636
Australia	12	22	119	462
Bahrain	13	118	762	1304
Bosnia and Herzegovina	8	51	66	298
Botswana	7	168	2337	2676
Bulgaria	38	31	5	369
Chinese Taipei	24	66	906	13874
Colombia	4	153	976	1594
Cyprus	6	39	27	164
Czech Republic	34	18	207	311
Egypt	22	141	2841	2116
El Salvador	13	234	4943	4300
England	8	24	58	868
Georgia	17	109	943	934
Ghana	33	437	20233	12205
Hong Kong, SAR	35	44	195	8648
Hungary	15	10	131	1630
Indonesia	5	125	821	1408
Iran, Islamic Republic of	8	130	563	1282
Israel	9	43	19	382
Italy	10	39	138	82
Japan	33	15	194	5384
Jordan	8	43	692	551
Korea, Republic of	33	35	401	11781
Kuwait	4	252	3776	3674
Lebanon	32	77	29	347
Lithuania	7	32	195	916
Malaysia	6	35	253	179
Malta	9	12	60	258
Morocco	30	150	1259	2335
Norway	8	41	253	314
Oman	89	140	4140	2925
Palestinian National Authority	31	161	5876	3621
Qatar	25	460	22456	13151
Romania	46	38	40	459
Russian Federation	14	30	173	1388
Saudi Arabia	29	293	8102	6120
Scotland	12	36	74	238
Serbia	4	31	157	530
Singapore	17	30	348	10321
Slovenia	13	45	210	136
Sweden	6	30	222	108
Syria, Arab Republic of	18	170	626	1426
Thailand	13	59	21	362
Tunisia	5	114	73	894
Turkey	12	92	547	1304
Ukraine	6	52	35	245
United States	15	19	183	615
Min	4	10	5	82
Max	89	460	22456	13874
Average	18	94	1810	2607

Table 5a. Multiple comparison of country means (overall mathematics): PV-R vs. PV-1 (above diagonal) and PV-R vs. PV-W (below diagonal)
Sorted by PV-R country mean (largest to smallest)

Country	PV-R	PV-1	PV-W	Chinese Taipei	Korea, Republic of	Singapore	Hong Kong, SAR	Japan	Hungary	England	Russian Federation	United States	Lithuania	Czech Republic	Slovenia	Armenia	Australia
Chinese Taipei	598,3 (4,53)	597,0 (4,34)	598,3 (4,36)		o	o	+	+	+	+	+	+	+	+	+	+	+
Korea, Republic of	597,3 (2,71)	595,5 (2,47)	597,3 (2,38)	o		o	+	+	+	+	+	+	+	+	+	+	+
Singapore	592,8 (3,81)	591,8 (3,73)	592,8 (3,73)	o	o		+	+	+	+	+	+	+	+	+	+	+
Hong Kong, SAR	572,5 (5,79)	570,9 (5,68)	572,5 (5,77)	+	+	+		o	+	+	+	+	+	+	+	+	+
Japan	569,8 (2,41)	568,4 (2,06)	569,8 (2,06)	+	+	+	o		+	+	+	+	+	+	+	+	+
Hungary	516,9 (3,47)	516,1 (3,42)	516,9 (3,39)	+	+	+	+	+		o	o	o	+	+	+	+	+
England	513,4 (4,82)	513,3 (4,79)	513,4 (4,83)	+	+	+	+	+	o		o	o	o	o+	+	+	+
Russian Federation	511,7 (4,10)	510,7 (4,04)	511,7 (3,89)	+	+	+	+	+	o	o		o	o	o	+	+	+
United States	508,5 (2,83)	508,4 (2,77)	508,5 (2,79)	+	+	+	+	+	o	o	o		o	o	+	+	+
Lithuania	505,8 (2,32)	505,3 (2,22)	505,8 (2,32)	+	+	+	+	+	+	o	o	o		o	o	o	+
Czech Republic	503,8 (2,39)	502,8 (2,31)	503,8 (2,41)	+	+	+	+	+	+	o	o	o	o		o	o	o
Slovenia	501,5 (2,11)	500,8 (2,00)	501,5 (2,04)	+	+	+	+	+	+	+	+	+	o	o		o	o
Armenia	498,7 (3,51)	498,0 (3,44)	498,7 (3,45)	+	+	+	+	+	+	+	+	+	o	o	o		o
Australia	496,2 (3,93)	495,7 (3,87)	496,2 (3,70)	+	+	+	+	+	+	+	+	+	+	o	o	o	
Sweden	491,3 (2,26)	491,1 (2,09)	491,3 (2,08)	+	+	+	+	+	+	+	+	+	+	+	+	o	o
Malta	487,8 (1,21)	487,3 (0,87)	487,8 (0,83)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Scotland	487,4 (3,70)	487,1 (3,61)	487,4 (3,63)	+	+	+	+	+	+	+	+	+	+	+	+	+	o
Serbia	485,8 (3,32)	485,6 (3,08)	485,8 (3,17)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Italy	479,6 (3,04)	478,7 (2,93)	479,6 (2,95)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Malaysia	473,9 (5,03)	473,8 (5,01)	473,9 (5,10)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Norway	469,2 (1,98)	468,5 (1,92)	469,2 (1,91)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Cyprus	465,5 (1,65)	465,0 (1,57)	465,5 (1,60)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bulgaria	463,6 (4,97)	463,2 (4,86)	463,6 (4,78)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Israel	463,3 (3,95)	462,8 (3,87)	463,3 (3,88)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ukraine	462,2 (3,62)	461,9 (3,60)	462,2 (3,55)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Romania	461,3 (4,10)	462,3 (4,04)	461,3 (4,07)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bosnia and Herzegovina	455,9 (2,70)	455,9 (2,68)	455,9 (2,63)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Lebanon	449,1 (3,98)	450,5 (3,83)	449,1 (3,91)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Thailand	441,4 (4,95)	441,6 (4,90)	441,4 (4,92)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Turkey	431,8 (4,75)	431,9 (4,68)	431,8 (4,63)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Jordan	426,9 (4,12)	427,3 (4,04)	426,9 (4,12)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Tunisia	420,4 (2,43)	420,5 (2,34)	420,4 (2,36)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Georgia	409,6 (5,95)	410,0 (5,89)	409,6 (5,65)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Iran, Islamic Republic of	403,4 (4,12)	403,2 (3,97)	403,4 (3,92)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bahrain	398,1 (1,57)	398,9 (1,32)	398,1 (1,33)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Indonesia	397,1 (3,81)	396,6 (3,69)	397,1 (3,70)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Syria, Arab Republic of	394,8 (3,76)	396,2 (3,41)	394,8 (3,37)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Egypt	390,6 (3,57)	391,7 (3,41)	390,6 (3,42)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Algeria	386,8 (2,14)	387,6 (1,49)	386,8 (1,44)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Morocco	380,8 (2,97)	381,9 (2,75)	380,8 (2,92)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Colombia	379,6 (3,63)	379,6 (3,60)	379,6 (3,46)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Oman	372,4 (3,37)	375,0 (2,85)	372,4 (2,88)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Palestinian National Authority	367,2 (3,55)	368,1 (3,40)	367,2 (3,45)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Botswana	363,5 (2,27)	363,4 (1,98)	363,5 (1,94)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Kuwait	353,7 (2,32)	353,3 (2,20)	353,7 (2,27)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
El Salvador	340,4 (2,76)	340,2 (2,66)	340,4 (2,64)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Saudi Arabia	329,3 (2,85)	330,9 (2,17)	329,3 (2,10)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ghana	309,4 (4,36)	310,3 (4,15)	309,4 (4,25)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Qatar	306,8 (1,37)	308,3 (0,73)	306,8 (0,68)	+	+	+	+	+	+	+	+	+	+	+	+	+	+

	Sweden	Malta	Scotland	Serbia	Italy	Malaysia	Norway	Cyprus	Bulgaria	Israel	Ukraine	Romania	Bosnia and Herzegovina	Lebanon	Thailand	Turkey	Jordan	Tunisia	Georgia	Iran, Islamic Republic of	Bahrain	Indonesia	Syria, Arab Republic of	Egypt	Algeria	Morocco	Colombia	Oman	Palestinian National Authority	Botswana	Kuwait	El Salvador	Saudi Arabia	Ghana	Qatar
Sweden	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Malta	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Scotland	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Serbia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Italy	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Malaysia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Norway	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Cyprus	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bulgaria	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Israel	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ukraine	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Romania	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bosnia and Herzegovina	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Lebanon	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Thailand	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Turkey	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Jordan	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Tunisia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Georgia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Iran, Islamic Republic of	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bahrain	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Indonesia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Syria, Arab Republic of	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Egypt	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Algeria	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Morocco	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Colombia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Oman	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Palestinian National Authority	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Botswana	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Kuwait	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
El Salvador	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Saudi Arabia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ghana	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Qatar	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	o	+	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
		o	o	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	o		o	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	o	o		o	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	o	o	o		o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	o	o		o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	o		o	o	o	o	o	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	o		o	o	o	o	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	o	o		o	o	o	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	o	o	o		o	o	o	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	o	o	o	o		o	o	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	o	o	o	o	o		o	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	o	o	o	o	o	o		o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	o	o	o	o		o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	o		o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	o		o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o		o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o		o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o		o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o		o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o		o	o	o	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o		o	o	o	+	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o+	o		o	o	o	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o	o	o		o	o	+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o	o	o		o	o+	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o+	o	o		o	+	+	+	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o+	o	o		o	o	+	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o		o	o	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o	o		o	+	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o+	o		o	+	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o	o		o	+	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o		o	+	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o		o	+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o		+	+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+
	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	o	

Table 5b. Multiple comparison of country means (overall mathematics): PV-R vs. EAP (above diagonal) and PV-R vs. WML (below diagonal)
Sorted by PV-R country mean (largest to smallest)

Country	PV-R	EAP	WML	Chinese Taipei	Korea, Republic of	Singapore	Hong Kong, SAR	Japan	Hungary	England	Russian Federation	United States	Lithuania	Czech Republic	Slovenia	Armenia	Australia
Chinese Taipei	598,3 (4,53)	588,1 (4,02)	562,6 (4,73)		o	o	+	+	+	+	+	+	+	+	+	+	+
Korea, Republic of	597,3 (2,71)	588,9 (2,17)	564,8 (2,61)	o		o	+	+	+	+	+	+	+	+	+	+	+
Singapore	592,8 (3,81)	585,1 (3,64)	562,4 (3,92)	o	o		+	+	+	+	+	+	+	+	+	+	+
Hong Kong, SAR	572,5 (5,79)	566,8 (5,48)	543,2 (5,76)	+	+	+		o	+	+	+	+	+	+	+	+	+
Japan	569,8 (2,41)	564,1 (1,92)	547,6 (1,81)	+	+	+	o		+	+	+	+	+	+	+	+	+
Hungary	516,9 (3,47)	513,8 (3,30)	505,4 (2,73)	+	+	+	+	+		o	o	o+	+	+	+	+	+
England	513,4 (4,82)	511,4 (4,70)	504,7 (4,10)	+	+	+	+	+	o		o	o	o	o+	+	+	+
Russian Federation	511,7 (4,10)	508,3 (3,82)	501,0 (3,29)	+	+	+	+	+	o	o		o	o	o	+	+	+
United States	508,5 (2,83)	504,9 (2,78)	500,5 (2,43)	+	+	+	+	+	o	o	o		o	o	+o	+	+
Lithuania	505,8 (2,32)	502,7 (2,31)	497,0 (1,93)	+	+	+	+	+	+	o	o	o		o	o	o+	+
Czech Republic	503,8 (2,39)	500,4 (2,41)	498,2 (2,10)	+	+	+	+	+	+	o	o	o	o		o	o	o
Slovenia	501,5 (2,11)	498,6 (2,08)	497,9 (1,86)	+	+	+	+	+	+	+o	+o	+o	o	o		o	o
Armenia	498,7 (3,51)	494,2 (3,15)	488,8 (2,64)	+	+	+	+	+	+	+	+	+	o+	o+	o+		o
Australia	496,2 (3,93)	493,6 (3,65)	489,5 (2,64)	+	+	+	+	+	+	+	+	+	+	o+	o+	o	
Sweden	491,3 (2,26)	488,0 (2,14)	488,7 (2,05)	+	+	+	+	+	+	+	+	+	+	+	+	o	o
Malta	487,8 (1,21)	487,0 (0,96)	483,7 (1,06)	+	+	+	+	+	+	+	+	+	+	+	+	+o	+
Scotland	487,4 (3,70)	485,7 (3,57)	483,5 (3,24)	+	+	+	+	+	+	+	+	+	+	+	+	+o	o
Serbia	485,8 (3,32)	483,9 (3,04)	479,8 (2,88)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Italy	479,6 (3,04)	477,8 (2,94)	478,6 (2,60)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Malaysia	473,9 (5,03)	471,7 (4,78)	471,7 (4,47)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Norway	469,2 (1,98)	466,5 (2,01)	470,8 (2,02)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Cyprus	465,5 (1,65)	465,9 (1,50)	465,1 (1,38)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bulgaria	463,6 (4,97)	465,6 (4,34)	459,9 (4,31)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Israel	463,3 (3,95)	465,1 (3,47)	460,1 (3,43)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ukraine	462,2 (3,62)	462,7 (3,17)	461,3 (3,13)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Romania	461,3 (4,10)	463,4 (3,71)	458,3 (3,62)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bosnia and Herzegovina	455,9 (2,70)	455,3 (2,46)	457,5 (2,46)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Lebanon	449,1 (3,98)	448,4 (3,63)	451,4 (3,43)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Thailand	441,4 (4,95)	443,6 (4,62)	441,8 (4,03)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Turkey	431,8 (4,75)	438,2 (4,04)	429,8 (3,57)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Jordan	426,9 (4,12)	434,6 (3,26)	432,4 (3,31)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Tunisia	420,4 (2,43)	421,5 (2,16)	426,6 (2,31)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Georgia	409,6 (5,95)	419,1 (4,45)	418,2 (4,90)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Iran, Islamic Republic of	403,4 (4,12)	411,3 (3,37)	411,7 (3,33)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bahrain	398,1 (1,57)	407,0 (1,21)	408,8 (1,36)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Indonesia	397,1 (3,81)	407,2 (3,02)	408,1 (3,24)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Syria, Arab Republic of	394,8 (3,76)	404,1 (2,71)	405,5 (2,88)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Egypt	390,6 (3,57)	405,9 (2,44)	405,1 (2,73)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Algeria	386,8 (2,14)	393,6 (1,21)	397,6 (1,36)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Morocco	380,8 (2,97)	392,9 (2,40)	394,7 (2,75)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Colombia	379,6 (3,63)	391,2 (2,62)	392,0 (2,86)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Oman	372,4 (3,37)	391,7 (2,05)	391,2 (2,27)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Palestinian National Authority	367,2 (3,55)	389,7 (2,41)	387,2 (2,71)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Botswana	363,5 (2,27)	379,8 (1,44)	380,5 (1,69)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Kuwait	353,7 (2,32)	374,3 (1,50)	373,5 (1,79)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
El Salvador	340,4 (2,76)	363,7 (1,76)	361,8 (2,12)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Saudi Arabia	329,3 (2,85)	358,5 (1,33)	355,4 (1,74)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ghana	309,4 (4,36)	352,5 (2,37)	346,7 (2,88)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Qatar	306,8 (1,37)	352,1 (0,72)	345,1 (0,96)	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Sweden
Malta
Scotland
Serbia
Italy
Malaysia
Norway
Cyprus
Bulgaria
Israel
Ukraine
Romania
Bosnia and Herzegovina
Lebanon
Thailand
Turkey
Jordan
Tunisia
Georgia
Iran, Islamic Republic of
Bahrain
Indonesia
Syria, Arab Republic of
Egypt
Algeria
Morocco
Colombia
Oman
Palestinian National Authority
Botswana
Kuwait
El Salvador
Saudi Arabia
Ghana
Qatar

Table 6. Gender differences (overall mathematics) for each score method and country (reference: girls)
Sorted by country name

Country	PV-R		PV-1		PV-W		EAP		WML	
	Difference	t	Difference	t	Difference	t	Difference	t	Difference	t
Algeria	5,3 (1,80)	2,94	5,4 (1,75)	3,10	5,3 (1,58)	3,35	4,3 (1,80)	2,39	4,5 (2,10)	2,17
Armenia	-4,0 (3,73)	-1,08	-2,1 (3,33)	-0,63	-4,0 (3,50)	-1,15	-4,4 (3,45)	-1,29	-2,8 (3,46)	-0,82
Australia	15,4 (7,71)	2,00	16,4 (7,63)	2,15	15,4 (7,67)	2,00	16,2 (7,80)	2,07	11,8 (6,39)	1,85
Bahrain	-31,9 (3,63)	-8,78	-30,7 (2,99)	-10,24	-31,9 (2,90)	-10,97	-24,0 (2,69)	-8,90	-30,0 (2,96)	-10,11
Bosnia and Herzegovina	-1,2 (2,49)	-0,48	-0,6 (2,38)	-0,24	-1,2 (2,17)	-0,55	-2,2 (2,07)	-1,06	-3,7 (2,20)	-1,67
Botswana	-15,4 (3,33)	-4,63	-15,6 (2,47)	-6,31	-15,4 (2,27)	-6,81	-9,1 (2,25)	-4,06	-10,9 (2,68)	-4,08
Bulgaria	-14,6 (4,95)	-2,94	-13,9 (4,86)	-2,85	-14,6 (4,64)	-3,14	-12,6 (4,63)	-2,72	-14,9 (4,81)	-3,10
Chinese Taipei	-0,8 (4,16)	-0,19	-1,9 (3,91)	-0,48	-0,8 (3,90)	-0,20	-2,6 (3,71)	-0,70	-6,1 (5,01)	-1,22
Colombia	31,8 (4,34)	7,31	32,6 (4,25)	7,66	31,8 (4,09)	7,76	24,5 (3,42)	7,18	24,8 (3,85)	6,43
Cyprus	-20,4 (3,24)	-6,32	-22,3 (2,84)	-7,86	-20,4 (2,67)	-7,65	-17,7 (2,88)	-6,15	-19,9 (2,84)	-7,00
Czech Republic	-2,4 (2,42)	-1,00	-2,8 (2,28)	-1,22	-2,4 (2,23)	-1,08	-2,4 (2,43)	-0,97	-2,1 (2,47)	-0,87
Egypt	-13,3 (6,45)	-2,06	-15,3 (6,01)	-2,55	-13,3 (6,01)	-2,21	-9,3 (4,57)	-2,03	-12,4 (4,92)	-2,52
El Salvador	20,6 (4,89)	4,22	21,4 (4,82)	4,44	20,6 (4,57)	4,51	11,3 (3,81)	2,96	12,5 (4,40)	2,84
England	5,8 (5,69)	1,02	5,5 (5,58)	0,99	5,8 (5,54)	1,05	6,6 (5,71)	1,15	3,2 (5,16)	0,63
Georgia	-4,2 (4,27)	-0,98	-0,4 (3,21)	-0,12	-4,2 (3,50)	-1,20	-2,3 (3,14)	-0,73	-4,5 (3,85)	-1,16
Ghana	21,9 (3,56)	6,16	21,2 (3,27)	6,50	21,9 (3,08)	7,13	10,7 (1,92)	5,54	13,1 (2,60)	5,04
Hong Kong, SAR	-11,5 (6,74)	-1,70	-12,8 (6,61)	-1,94	-11,5 (6,55)	-1,75	-11,9 (6,35)	-1,88	-18,7 (6,42)	-2,91
Hungary	0,6 (3,57)	0,17	1,3 (3,43)	0,37	0,6 (3,25)	0,19	0,9 (3,47)	0,27	-0,4 (3,41)	-0,11
Indonesia	-3,9 (3,95)	-1,00	-3,9 (3,68)	-1,06	-3,9 (3,56)	-1,11	-4,5 (3,26)	-1,39	-5,9 (3,49)	-1,68
Iran, Islamic Republic of	-7,0 (8,06)	-0,87	-7,1 (7,91)	-0,90	-7,0 (7,83)	-0,89	-5,7 (6,67)	-0,86	-8,8 (6,46)	-1,36
Israel	-2,6 (5,40)	-0,48	-2,3 (5,25)	-0,44	-2,6 (5,03)	-0,52	-2,6 (4,81)	-0,54	-8,3 (4,72)	-1,76
Italy	6,0 (3,20)	1,87	5,9 (2,83)	2,09	6,0 (2,76)	2,16	4,6 (2,98)	1,54	3,3 (2,84)	1,16
Japan	4,2 (4,33)	0,96	4,4 (4,05)	1,09	4,2 (4,06)	1,02	3,3 (3,83)	0,87	-3,2 (3,43)	-0,93
Jordan	-20,5 (8,83)	-2,32	-19,9 (8,55)	-2,33	-20,5 (8,73)	-2,34	-15,7 (7,07)	-2,23	-16,7 (7,11)	-2,35
Korea, Republic of	4,0 (3,37)	1,18	4,3 (3,32)	1,29	4,0 (3,26)	1,22	3,0 (3,08)	0,99	3,6 (4,28)	0,85
Kuwait	-22,0 (4,82)	-4,56	-24,6 (4,31)	-5,71	-22,0 (4,43)	-4,97	-11,5 (2,88)	-3,98	-16,3 (3,41)	-4,78
Lebanon	13,4 (3,55)	3,76	12,6 (3,19)	3,94	13,4 (3,36)	3,98	12,9 (3,30)	3,91	11,9 (3,35)	3,57
Lithuania	-6,7 (2,61)	-2,56	-8,0 (2,44)	-3,27	-6,7 (2,58)	-2,59	-7,6 (2,72)	-2,78	-7,0 (2,45)	-2,85
Malaysia	-10,6 (4,43)	-2,40	-9,7 (4,37)	-2,22	-10,6 (4,31)	-2,47	-10,5 (4,24)	-2,47	-11,9 (4,01)	-2,96
Malta	0,1 (2,16)	0,05	-0,7 (2,02)	-0,33	0,1 (1,97)	0,05	0,2 (2,22)	0,11	-1,7 (2,51)	-0,69
Morocco	8,5 (4,82)	1,77	10,5 (4,60)	2,27	8,5 (4,48)	1,90	6,4 (4,24)	1,50	6,2 (4,87)	1,28
Norway	-3,9 (2,48)	-1,58	-2,7 (2,35)	-1,15	-3,9 (2,14)	-1,83	-4,7 (2,50)	-1,87	-5,9 (2,58)	-2,29
Oman	-54,3 (5,61)	-9,67	-53,8 (5,52)	-9,75	-54,3 (5,59)	-9,71	-35,3 (3,84)	-9,19	-42,0 (4,30)	-9,77
Palestinian National Authority	-36,2 (6,52)	-5,55	-37,0 (6,18)	-5,99	-36,2 (6,18)	-5,86	-24,7 (4,40)	-5,61	-29,6 (5,00)	-5,92
Qatar	-37,5 (2,92)	-12,83	-35,7 (1,70)	-20,97	-37,5 (1,57)	-23,95	-15,5 (1,35)	-11,50	-21,4 (1,79)	-11,95
Romania	-18,0 (3,33)	-5,42	-17,8 (3,10)	-5,73	-18,0 (3,20)	-5,63	-15,4 (3,17)	-4,84	-16,7 (3,64)	-4,59
Russian Federation	-4,6 (3,75)	-1,24	-5,4 (3,62)	-1,48	-4,6 (3,57)	-1,30	-3,8 (4,04)	-0,95	-7,1 (4,12)	-1,72
Saudi Arabia	-22,6 (4,99)	-4,52	-22,1 (4,38)	-5,05	-22,6 (4,43)	-5,10	-12,0 (2,78)	-4,31	-16,7 (3,58)	-4,67
Scotland	2,6 (3,51)	0,75	2,6 (3,27)	0,81	2,6 (3,29)	0,80	2,5 (3,33)	0,75	-0,7 (2,94)	-0,25
Serbia	-6,3 (3,90)	-1,62	-7,4 (3,63)	-2,03	-6,3 (3,64)	-1,74	-6,6 (3,70)	-1,78	-7,7 (3,70)	-2,07
Singapore	-14,6 (4,36)	-3,36	-16,2 (4,09)	-3,97	-14,6 (4,01)	-3,66	-15,7 (4,06)	-3,88	-17,5 (4,58)	-3,83
Slovenia	2,4 (3,24)	0,74	3,8 (2,72)	1,40	2,4 (2,48)	0,97	3,2 (2,75)	1,18	1,1 (2,53)	0,45
Sweden	-3,6 (2,47)	-1,46	-5,0 (2,28)	-2,20	-3,6 (2,18)	-1,65	-3,8 (2,58)	-1,49	-4,5 (2,47)	-1,81
Syria, Arab Republic of	16,2 (5,61)	2,89	14,9 (5,46)	2,72	16,2 (5,28)	3,08	12,4 (4,19)	2,96	12,0 (4,50)	2,66
Thailand	-22,8 (4,67)	-4,87	-23,9 (4,40)	-5,43	-22,8 (4,21)	-5,40	-19,9 (4,01)	-4,95	-19,9 (3,84)	-5,19
Tunisia	21,0 (2,41)	8,71	22,7 (2,11)	10,78	21,0 (1,98)	10,58	19,0 (2,22)	8,55	20,3 (2,52)	8,04
Turkey	-0,5 (3,89)	-0,14	1,0 (3,67)	0,26	-0,5 (3,62)	-0,15	-0,8 (3,37)	-0,24	-2,0 (3,47)	-0,59
Ukraine	-5,2 (2,94)	-1,78	-5,8 (2,89)	-1,99	-5,2 (2,96)	-1,77	-4,8 (3,19)	-1,50	-8,4 (3,34)	-2,50
United States	3,6 (2,23)	1,62	3,7 (2,15)	1,71	3,6 (2,11)	1,71	3,4 (2,20)	1,56	0,6 (2,16)	0,26

Table 7. t-test values for mean differences (overall mathematics) between adjacent levels on 'books at home' (backward difference coded) by method and country
Sorted by country name

	Few/no books <> one shelf					One shelf <> one bookcase					One bookcase <> two bookcases					two bookcases <> three or more				
Country	**PV-R**	**PV-1**	**PV-W**	**EAP**	**WML**	**PV-R**	**PV-1**	**PV-W**	**EAP**	**WML**	**PV-R**	**PV-1**	**PV-W**	**EAP**	**WML**	**PV-R**	**PV-1**	**PV-W**	**EAP**	**WML**
Algeria	1,53	1,71	1,84	1,37	1,19	3,97	4,40	4,87	3,70	3,44	-0,48	-0,92	-0,65	-0,60	-0,95	-0,93	-1,51	-1,19	-0,90	-1,05
Armenia	0,30	0,22	0,36	0,63	1,75	2,94	3,09	3,18	2,81	2,69	1,67	1,72	2,01	1,69	1,01	0,03	0,22	0,03	0,11	0,04
Australia	4,12	5,90	5,05	4,62	4,62	4,65	4,82	5,86	5,64	5,34	5,59	6,93	6,75	5,55	5,04	3,15	3,45	3,69	3,60	3,12
Bahrain	1,36	0,96	1,47	0,95	0,71	9,47	9,60	10,37	8,16	8,08	3,14	4,25	3,80	3,53	3,46	-2,85	-3,50	-3,48	-2,60	-2,76
Bosnia and Herzegovina	5,80	6,17	6,38	5,18	4,81	5,55	5,64	6,14	5,51	4,97	1,61	2,63	1,89	1,29	0,89	1,24	1,00	1,35	1,35	1,51
Botswana	2,14	2,27	2,53	1,55	1,32	3,61	6,49	5,94	4,74	3,98	-0,87	-1,37	-0,98	-0,59	-0,74	-0,02	-0,18	-0,02	-0,05	-0,08
Bulgaria	3,96	4,01	4,48	4,01	3,68	4,41	4,78	4,83	4,62	4,45	3,78	3,79	4,09	3,47	3,35	1,24	1,21	1,38	1,02	0,49
Chinese Taipei	8,36	8,67	9,03	8,43	6,80	6,02	6,22	6,87	6,18	5,63	5,05	5,52	5,52	5,49	4,55	2,19	2,45	2,35	1,89	2,00
Colombia	7,47	7,94	9,45	8,96	8,86	4,79	5,69	5,60	5,42	4,92	2,67	3,23	3,10	2,00	2,17	1,19	0,43	1,53	1,79	1,00
Cyprus	5,96	6,22	6,87	5,41	5,23	7,87	8,50	8,85	7,84	7,70	6,12	6,56	6,52	6,05	5,62	-1,79	-1,95	-2,09	-2,13	-1,84
Czech Republic	2,70	2,84	3,47	2,98	3,05	11,06	11,39	12,48	10,63	9,17	7,80	8,00	7,36	6,45	5,02	4,14	4,32	4,37	3,80	2,86
Egypt	1,75	2,49	2,09	2,65	2,15	3,99	3,69	4,38	3,99	3,55	0,76	0,77	0,77	1,20	1,00	-3,02	-2,81	-3,32	-3,11	-3,02
El Salvador	6,64	6,83	8,46	6,53	6,39	4,21	5,40	5,00	3,83	3,59	1,03	2,51	1,63	2,12	2,07	-2,42	-3,69	-3,61	-3,32	-3,29
England	6,91	7,28	6,93	6,28	5,76	8,16	8,52	8,31	7,51	7,06	3,75	3,75	4,08	3,11	3,11	5,43	5,55	5,81	5,50	4,14
Georgia	1,85	2,44	2,15	2,34	2,01	2,68	2,76	3,24	2,71	2,73	3,65	3,42	3,57	3,18	2,88	0,83	0,66	0,93	0,77	0,48
Ghana	-0,39	-0,46	-0,44	0,19	0,34	3,37	4,11	4,34	3,89	3,50	-1,56	-1,65	-1,68	-1,16	-1,03	0,04	-0,18	0,05	0,10	0,17
Hong Kong, SAR	6,30	6,54	6,21	5,76	4,48	4,98	5,73	5,58	5,12	4,75	1,07	1,62	1,10	0,88	0,58	1,48	1,09	1,91	1,59	1,23
Hungary	5,62	6,12	5,99	6,12	6,10	8,04	8,27	8,21	7,54	7,09	6,36	6,28	6,81	6,49	5,07	5,20	5,28	5,39	4,52	3,76
Indonesia	-1,00	-0,60	-1,11	-0,89	-0,69	6,22	6,17	6,87	6,28	5,43	0,99	0,90	1,15	1,54	0,92	-0,38	-0,38	-0,37	-0,39	-0,38
Iran, Islamic Republic of	4,21	5,49	4,99	4,30	4,01	6,79	6,68	7,23	6,69	5,60	1,28	1,46	1,43	1,68	1,31	-0,81	-0,83	-0,89	-0,63	-0,45
Israel	2,89	2,98	3,09	2,92	2,86	4,94	5,54	5,44	4,90	4,38	3,30	3,30	3,83	3,50	2,63	1,16	1,04	1,24	1,32	0,42
Italy	3,74	3,72	3,92	3,65	3,83	6,05	7,87	7,69	7,31	7,61	3,70	3,85	4,45	4,61	4,53	1,70	2,15	2,02	1,64	1,11
Japan	4,54	5,09	5,45	4,96	3,71	5,98	6,04	6,40	5,62	4,11	2,70	3,28	3,02	2,94	1,74	3,06	3,10	3,31	2,43	1,79
Jordan	2,94	2,80	3,00	2,55	2,51	5,10	5,22	5,96	5,59	5,36	1,15	1,03	1,29	1,34	0,39	1,14	1,33	1,30	0,74	1,57
Korea, Republic of	3,50	3,52	3,91	3,74	3,06	7,32	9,04	8,22	7,75	5,92	8,33	8,48	8,78	8,22	5,46	7,26	8,83	8,89	7,31	5,76
Kuwait	3,11	3,86	4,05	3,71	3,74	2,41	4,02	4,30	3,03	2,66	0,88	1,07	1,14	0,65	0,69	-2,92	-2,94	-2,97	-1,79	-1,88
Lebanon	3,75	4,32	4,59	3,73	3,40	5,54	5,83	5,78	5,84	5,45	1,20	0,98	1,18	1,07	0,65	-1,15	-0,99	-1,21	-1,09	-0,95
Lithuania	3,89	4,00	4,32	4,15	3,89	9,19	9,75	9,94	8,58	7,53	5,53	5,51	5,95	5,89	5,17	0,08	0,31	0,08	0,18	-0,45
Malaysia	4,69	5,19	4,72	4,41	4,06	10,06	10,18	10,55	10,09	9,28	3,57	3,52	4,14	3,18	2,20	2,81	3,15	2,81	2,58	2,19
Malta	10,34	12,75	12,11	10,03	9,44	6,68	6,78	7,32	6,83	7,02	6,29	6,22	6,85	5,94	5,25	0,57	0,43	0,59	0,68	-0,24
Morocco	1,59	1,69	1,78	1,27	0,92	2,80	2,69	2,93	3,17	3,31	1,30	1,63	1,48	1,14	1,10	-0,73	-0,93	-0,86	-0,62	-0,59
Norway	5,69	5,80	5,80	4,47	4,35	7,42	10,03	11,89	9,19	8,82	4,48	4,45	4,68	4,14	4,08	2,94	3,22	3,83	3,24	2,78
Oman	5,78	6,74	7,46	6,20	5,98	5,69	6,18	6,10	5,39	5,21	1,04	1,80	1,14	0,57	0,78	-0,66	-0,96	-0,78	-0,19	-0,46
Palestinian National Authority	3,73	3,99	4,20	4,15	3,72	3,79	4,20	4,13	3,85	3,38	1,47	1,47	1,59	1,93	1,56	-1,56	-2,09	-1,95	-2,01	-1,78
Qatar	6,14	7,08	7,99	5,35	5,27	8,76	9,51	11,01	8,74	8,02	0,81	1,67	1,06	0,80	0,80	-2,39	-3,49	-3,47	-2,10	-2,28
Romania	5,94	6,04	5,62	5,07	4,44	7,04	7,07	7,27	6,83	6,27	3,98	4,71	4,29	4,17	3,47	1,31	1,18	1,29	1,05	0,41
Russian Federation	1,46	1,30	1,68	1,20	0,75	6,53	6,89	7,83	7,54	6,30	4,42	4,28	4,65	4,22	3,80	1,34	1,42	1,47	1,21	0,64
Saudi Arabia	3,74	5,52	6,54	4,78	4,84	4,16	4,84	4,91	5,06	4,92	1,18	0,96	1,75	1,79	1,59	-1,98	-2,14	-2,63	-2,35	-2,19
Scotland	7,46	7,72	7,37	6,76	6,25	7,90	7,90	8,36	7,75	6,36	5,78	6,37	6,52	5,60	4,78	2,22	2,65	2,58	2,33	1,15
Serbia	5,96	6,55	7,08	6,10	5,36	9,93	10,56	10,17	10,62	9,90	1,01	0,87	1,05	0,92	0,25	1,66	1,65	1,77	1,84	1,51
Singapore	6,33	6,78	6,82	6,31	4,94	9,27	9,60	10,25	9,67	8,56	4,40	4,89	4,71	4,05	2,84	2,58	2,59	2,82	2,44	2,15
Slovenia	5,57	5,71	5,79	4,95	4,71	9,65	10,38	10,74	8,89	7,68	4,71	4,62	4,90	4,18	3,45	1,12	1,46	1,18	1,21	0,51
Sweden	5,17	5,38	6,11	5,47	5,69	6,60	7,27	6,96	5,86	5,99	4,85	5,80	5,18	4,53	3,98	5,69	5,54	6,15	5,38	5,12
Syria, Arab Republic of	1,73	2,09	2,01	1,78	2,10	4,11	4,32	5,11	4,52	4,52	0,04	-0,06	0,05	0,42	-0,01	-0,79	-1,06	-1,01	-1,03	-1,35
Thailand	4,54	4,64	4,70	4,12	3,89	6,54	6,39	6,64	6,51	6,49	3,40	3,43	3,70	3,67	3,32	2,97	3,69	3,05	2,49	1,32
Tunisia	2,02	2,32	2,83	2,25	1,82	8,77	9,51	11,05	8,89	8,46	5,53	6,12	6,15	4,70	3,84	-1,69	-1,68	-1,85	-1,78	-1,82
Turkey	9,52	10,86	10,56	9,85	9,19	8,54	8,64	8,64	8,29	6,31	4,14	4,37	4,24	4,22	3,01	-0,25	-0,23	-0,27	-0,13	0,10
Ukraine	4,29	4,23	4,67	3,28	2,85	8,91	10,05	9,96	10,21	8,82	3,88	3,80	4,12	3,69	3,13	1,65	1,60	1,74	1,91	1,13
United States	5,87	6,24	6,40	5,23	5,03	12,18	14,70	16,14	14,45	14,73	7,31	8,22	7,95	7,27	6,14	2,39	2,48	2,70	2,07	1,31

Table 8. Multiple comparison of country means (geometry): PV-R vs. PV-1 (above diagonal) and PV-R vs. PV-W (below diagonal)
Sorted by PV-R country mean (largest to smallest)

Country	PV-R	PV-1	PV-W	Chinese Taipei	Korea, Republic of	Singapore	Hong Kong, SAR	Japan	Hungary	England	Russian Federation	United States	Lithuania	Czech Republic	Slovenia	Armenia	Australia
Chinese Taipei	591,97 (4,62)	590,31 (4,08)	591,97 (4,10)	o	+	+	+	+	+	+	+	+	+	+	+	+	+
Korea, Republic of	586,59 (2,33)	586,70 (2,07)	586,59 (2,08)		+	+	+	+	+	+	+	+	+	+	+	+	+
Singapore	578,35 (3,37)	578,36 (3,33)	578,35 (3,41)	+		o	o	+	+	+	+	+	+	+	+	+	+
Japan	572,86 (2,16)	572,86 (1,91)	572,86 (1,75)	+	o		o	+	+	+	+	+	+	+	+	+	+
Hong Kong, SAR	569,90 (5,47)	569,47 (5,37)	569,90 (5,20)	+	o	o		+	+	+	+	+	+	+	+	+	+
England	510,12 (4,41)	508,56 (4,29)	510,12 (4,35)	+	+	+	+		o	o	o	+o	+o	+	+	+	+
Russian Federation	509,63 (4,07)	509,48 (3,67)	509,63 (3,57)	+	+	+	+	o		o	o	+	+	+	+	+	+
Hungary	507,59 (3,63)	507,54 (3,55)	507,59 (3,39)	+	+	+	+	o	o		o	o	+o	+	+	+	+
Lithuania	506,92 (2,65)	506,16 (2,23)	506,92 (2,20)	+	+	+	+	o	o	o		+	+	+	+	+	+
Slovenia	499,46 (2,40)	499,46 (2,14)	499,46 (2,00)	+	+	+	+	+	+	o+	+		o	o	o	+	+
Czech Republic	497,62 (2,73)	499,36 (2,24)	497,62 (2,35)	+	+	+	+	+	+	+	+	o		o	o	+	+
Malta	495,12 (1,12)	495,17 (0,80)	495,12 (0,76)	+	+	+	+	+	+	+	+	o+	o		o	o	+o
Armenia	492,85 (4,12)	492,15 (4,05)	492,85 (3,89)	+	+	+	+	+	+	+	+	o	o	o		o	o
Italy	489,59 (3,05)	489,70 (2,98)	489,59 (3,01)	+	+	+	+	+	+	+	+	+	+	o	o		o
Australia	487,44 (3,63)	489,27 (3,27)	487,44 (3,34)	+	+	+	+	+	+	+	+	+	+	+	o	o	
Serbia	485,74 (3,60)	484,26 (3,42)	485,74 (3,31)	+	+	+	+	+	+	+	+	+	+	+	o	o	o
Scotland	485,45 (3,85)	484,23 (3,27)	485,45 (3,25)	+	+	+	+	+	+	+	+	+	+	+	o	o	o
United States	479,94 (2,52)	479,31 (2,40)	479,94 (2,40)	+	+	+	+	+	+	+	+	+	+	+	+	+	o
Malaysia	476,89 (5,55)	477,45 (5,52)	476,89 (5,45)	+	+	+	+	+	+	+	+	+	+	+	+	+	o
Sweden	471,69 (2,52)	471,32 (2,33)	471,69 (2,28)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bulgaria	468,23 (5,05)	469,67 (4,42)	468,23 (4,37)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ukraine	467,22 (3,57)	465,37 (3,25)	467,22 (3,16)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Romania	466,44 (4,03)	465,12 (3,90)	466,44 (3,82)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Lebanon	462,13 (3,95)	462,85 (3,51)	462,13 (3,46)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Norway	458,71 (2,28)	458,90 (2,26)	458,71 (2,12)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Cyprus	457,68 (2,70)	460,51 (1,99)	457,68 (2,03)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bosnia and Herzegovina	450,90 (3,47)	453,18 (2,81)	450,90 (2,65)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Thailand	441,93 (5,32)	439,97 (5,08)	441,93 (5,02)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Tunisia	436,79 (2,59)	435,32 (2,28)	436,79 (2,15)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Israel	436,04 (4,28)	437,69 (3,90)	436,04 (3,88)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Jordan	435,60 (3,85)	436,15 (3,75)	435,60 (3,68)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Algeria	432,14 (2,10)	431,35 (1,44)	432,14 (1,39)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Iran, Islamic Republic of	422,67 (4,38)	423,04 (4,10)	422,67 (3,93)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Syria, Arab Republic of	417,19 (3,44)	416,39 (3,16)	417,19 (3,15)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bahrain	412,27 (2,11)	411,52 (1,45)	412,27 (1,37)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Turkey	411,11 (5,09)	411,56 (4,81)	411,11 (4,79)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Georgia	408,61 (6,71)	407,46 (6,49)	408,61 (6,16)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Egypt	406,31 (3,41)	406,19 (3,32)	406,31 (3,21)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Morocco	396,37 (3,64)	395,83 (3,30)	396,37 (3,16)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Indonesia	394,58 (4,45)	396,68 (3,93)	394,58 (3,80)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Palestinian National Authority	388,17 (3,78)	389,76 (3,10)	388,17 (3,24)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Oman	387,46 (3,03)	385,17 (2,66)	387,46 (2,78)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Kuwait	384,65 (2,84)	385,85 (2,66)	384,65 (2,30)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Colombia	371,35 (3,30)	370,73 (3,21)	371,35 (3,27)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Saudi Arabia	358,90 (2,61)	360,25 (2,30)	358,90 (2,08)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Botswana	324,55 (3,18)	326,95 (2,66)	324,55 (2,46)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
El Salvador	317,66 (3,68)	317,61 (3,21)	317,66 (3,13)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Qatar	301,46 (1,80)	301,19 (1,08)	301,46 (0,85)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ghana	274,52 (4,86)	274,01 (4,47)	274,52 (4,43)	+	+	+	+	+	+	+	+	+	+	+	+	+	+

	Sweden	Malta	Scotland	Serbia	Italy	Malaysia	Norway	Cyprus	Bulgaria	Israel	Ukraine	Romania	Bosnia and Herzegovina	Lebanon	Thailand	Turkey	Jordan	Tunisia	Georgia	Iran, Islamic Republic of	Bahrain	Indonesia	Syria, Arab Republic of	Egypt	Algeria	Morocco	Colombia	Oman	Palestinian National Authority	Botswana	Kuwait	El Salvador	Saudi Arabia	Ghana	Qatar
Sweden	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Malta	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Scotland	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Serbia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Italy	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Malaysia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Norway	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Cyprus	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bulgaria	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Israel	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ukraine	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Romania	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bosnia and Herzegovina	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Lebanon	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Thailand	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Turkey	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Jordan	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Tunisia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Georgia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Iran, Islamic Republic of	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bahrain	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Indonesia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Syria, Arab Republic of	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Egypt	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Algeria	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Morocco	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Colombia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Oman	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Palestinian National Authority	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Botswana	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Kuwait	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
El Salvador	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Saudi Arabia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ghana	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Qatar	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Sweden	o	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Malta	o	o	+	+o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Scotland	o	o	o+	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Serbia		o	o	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Italy	o		o	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Malaysia	o	o		o	+	+o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Norway	o	o	o		o	o	o	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Cyprus	+	+	+	+	+	o	o	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bulgaria	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Israel	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ukraine	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Romania	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bosnia and Herzegovina	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Lebanon	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Thailand	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Turkey	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Jordan	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Tunisia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Georgia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Iran, Islamic Republic of	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Bahrain	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Indonesia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Syria, Arab Republic of	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Egypt	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Algeria	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Morocco	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Colombia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Oman	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Palestinian National Authority	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Botswana	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Kuwait	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
El Salvador	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Saudi Arabia	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ghana	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Qatar	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+