

Comparing results of TIMSS and the Hungarian National Assessment of Basic Competencies

Ildikó Balázs, Ildikó Szepesi

Hungarian Educational Authority, Department for Analyses of Public Education /
Department of Assessment and Evaluation

Abstract: In our paper, we compared some characteristics of TIMSS 2015 and the National Assessment of Basic Competencies (NABC) 2015. The NABC assesses all students' reading and mathematics performance in Grades 6, 8 and 10. Both studies assessed Hungarian Grade 8 students' mathematics abilities in the spring of 2015. We linked data of the two studies on the student level using Student Measurement IDs.

We compared TIMSS and NABC mathematics scales based on the Assessment Framework of the two studies along with the results of students in the two assessments. The comparison of the Frameworks revealed that although the two tests use similar content and cognitive categorizations, there are crucial differences between the two constructs. While the basis of TIMSS's mathematics construct is the common part of mathematics curricula of participating countries, NABC intends to measure mathematical literacy, the ability of students to use their mathematical knowledge and competencies in real life situations. The correlation between the TIMSS and NABC mathematics test results (0.79) also confirms that the two tests measure related, but not identical abilities.

To evaluate the representativeness of the TIMSS sample we used school- and class-level weight factors of TIMSS and the student-level weights of NABC combined. The mean performances of the TIMSS sample are only slightly lower than the full NABC cohort's, the effect size of the difference is 0.042 and 0.046 in mathematics and reading respectively. The differences in the standard deviations are somewhat but not considerably larger. The SES-index shows a very good match with no statistically significant differences in the mean and standard deviation of the sample and the full cohort. Our analysis confirms that estimations of population parameters based on TIMSS samples are of a good quality.

Keywords: TIMSS 2015; Hungarian NABC; mathematics test; representativeness

The Hungarian Evaluation and Assessment Framework uses both international and national student assessments to evaluate the performance and other characteristics of the school system (Sinka, 2010). TIMSS, as one of the international large scale student assessments, describes the school systems' characteristics and quality by measuring their students' performance in an international context every four years. The International Reports of TIMSS compare countries based on their student achievement in mathematics and science along with student, teacher and school characteristics in Grades 4 and 8 (Mullis, Martin, Foy, & Hooper, 2016).

In contrast, the Hungarian National Assessment of Basic Competencies (NABC) evaluates individual schools' results, reporting their students' performance in reading and mathematics in Grades 6, 8 and 10 annually (Balázs, Lak, Ostorics, Szabó, & Vadász, 2016). The main aim of NABC is to empower schools and the wider public

<https://doi.org/10.14712/23363177.2018.293>

www.orbisscholae.cz

© 2018 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>).

66 with objective, reliable and comparable data on students' performance in areas crucial for the well-being and prosperity of students in their later life. The reports show students' results by various background characteristics, like settlement type and size, school type and size, socio-economic background and baseline performance from two years earlier. Therefore, schools, school maintainers and parents can evaluate the schools' results taking into account these background characteristics.

The two studies serve different purposes and complement each other. From policy perspective, TIMSS is used to put the results and features of the Hungarian school system in international context, to evaluate strengths and weaknesses in comparison with similar countries or with countries which can be seen as a role model for some reason (see for example Szalay, Szepesi, & Vadász, 2016, pp. 270–271). NABC is mainly used on school level, although detailed analyses of within country structure of the education system are also available (Balázsi et al., 2016), as well as secondary analyses of the data used alone or complementing other primary data collections (see for example Horn, 2013; Kertesi & Kézdi, 2016).

Transparency, validity and reliability of the data presented in these assessments is crucial from educational policy perspective. To ensure these, both TIMSS and NABC published their methods and processes followed during test development, sampling, data collection and reporting (Martin, Mullis, & Hooper, 2016; Aux-Bánfi et al., 2015). In our research, we attempt to analyze retrospectively the validity of the results, based on crosschecking the data in the two databases.

In our paper, two research questions are addressed. Since both TIMSS and NABC assessed the mathematics performance of students in Grade 8, our first question is whether the two constructs are the same and if not, what the main differences are. Mathematics is a complex construct, and different assessments might define it somewhat differently according to their aims. Besides, during test item development and test item selection, different content or cognitive areas of mathematics might get different emphasis in each study, and item types can differ as well. According to a research conducted by the U.S. Department of Education Institute of Education Sciences: “although the NAEP, TIMSS, and PISA 2003 mathematics frameworks address many similar topics and require students to use a range of cognitive skills and processes, it cannot be assumed that they measure the same content in the same way” (Neidorf, Binkley, Gattis, & Nohara, 2006, p. iv.). To interpret the mathematics results from TIMSS and NABC correctly, we have to take into account any differences of the two studies' mathematics scale constructs. Besides, with linking the TIMSS and NABC databases on student level, we can analyze the correlation between the results in the two studies, giving empirical support for our findings based on the content comparison of the tests.

Our second research question addresses the representativeness of the TIMSS 2015 sample. TIMSS has rigorous procedures for sampling and participation, makes a great effort to ensure that its findings are valid for the whole educational system of the participating countries and regions (LaRoche, Joncas, & Foy, 2016). However, linking TIMSS and NABC data on student level, we can evaluate the representativeness of

the TIMSS sample retrospectively and independently from the databases and procedures used for sampling in TIMSS. Although scientifically might seem unnecessary to evaluate the representativeness of the TIMSS sample due to the scientific rigor of sampling in TIMSS, skeptical views on international large scale studies among educators, policy makers or the general public from time to time question the quality of the sample or the relevance of results coming from a sample to the whole population. For example, László Mendrey, the president of the Hungarian Democratic Trade Union of Teachers at that time stated to an online newspaper that "... only 150 Hungarian schools' 4th Grade students participated in PIRLS. [...] The problem is that PIRLS represents only the results of a fragment of schools, as there are more than four thousand public education institutions in Hungary."¹ Therefore, having data of the whole population on which TIMSS sampling is based gives an excellent opportunity to prove the relevance of the results of international large scale assessments for the whole educational system. Proving representativeness of the sample independently from the study itself makes a strong argument easily comprehensible for a nonprofessional audience as well.

1 Data and Methods

In our research, we chose to link and analyze TIMSS 2015 and NABC 2015 Grade 8 databases, since these assessed the same student population, and have mathematics as one of their cognitive domain in common. Student measurement identification (SMID), which, introduced in 2008, is used in every international and national large-scale assessment in Hungary, and allows us to link the data from the two studies on student level. The Hungarian TIMSS data was collected between 30th March and 28th April, NABC 2015 was administered on 27th May in the same school year. As NABC is a census, the TIMSS sample is approximately² a sample from the NABC population. Indeed, all but 8 students of the TIMSS sample are present in the NABC data file (Table 1). These 8 students either dropped out from the school system or moved abroad between the two data collections, or some database error in one of the studies prevented linking the SMIDs.

We compared the mathematics test contents based on the frameworks of the two studies (Mullis & Martin, 2013; Balázsi et al., 2014). Simultaneously, we evaluated the similarities of the two results using correlations. We have used TIMSS plausible

¹ Article published on December 6th, 2017 on <https://24.hu/belfold/2017/12/06/a-magyar-diakok-minden-eddiginel-jobb-eredmenye-nem-pont-az-aminek-latszok>, the quoted sentences were translated by the author.

² Not exactly, as exclusion policies differ in the two studies. TIMSS, trying to be as inclusive as possible, asks schools to exclude students with special education needs (SEN) only if their disabilities would seriously affect their test writing and results. Schools with only such SEN students were excluded before sampling. In contrast, in NABC all SEN students were excluded to ensure school comparability. However, SEN students learning in inclusive schools are included in the database of NABC, although they are marked as non-eligible. The overall exclusion rate is 5.4% in TIMSS (Mullis et al., 2016, Appendix C.2) and 6.4% in NABC (own calculation).

68 Table 1 Number of students in TIMSS 2015 and NABC 2015 Grade 8 databases.

Number of students (unweighted)	Total	In TIMSS sample (eligible students)	With performance data (in TIMSS database)	With data on parents highest education
Total		5,058	4,893	4,857
In NABC database	88,967	5,050	4,891	4,855
NABC eligible	84,113	4,887	4,738	4,705
With NABC performance data	78,985	4,615	4,492	4,463
With NABC SES-index	62,317	3,836	3,736	3,715

values and NABC IRT ability scores to calculate correlations along with the weighting variable and jackknife error calculation methods of TIMSS (Martin et al., 2016, Aux-Bánfi et al., 2015).

TIMSS and NABC use slightly different methods and software for calculating the performance of students (Yamamoto & Kulick, 2016; Aux-Bánfi et al., 2015, pp. 91–105). Also, TIMSS item parameters are calculated based on data from every participating country using equal weights for every country (Foy & Yin, 2016). These differences might affect Hungarian students' TIMSS 2015 performance scores and hence the correlation between the two tests' results. In order to exclude these effects from our comparison, we have also compared NABC scores with a performance score of the TIMSS mathematics test calculated from item response level data of the Hungarian students with methods and software used in NABC.

To evaluate the representativeness of the TIMSS sample we compared NABC mathematics and reading results of students in the TIMSS sample to the results of the overall NABC cohort. We have also used students' SES-indices to compare the TIMSS sample to the whole population. The SES-index is based on some questions of the non-compulsory student background questionnaire of NABC, so due to a large number of missing values we should interpret results based on SES carefully. However, the response rate is high, 79% of non-missing students have SES data as well. The index consist variables related to highest education of parents, number of books at home and educational and economical resources possessed by the students' family. It was anchored so the average SES is 0 and the standard deviation is 1 for the overall student population of the three grades involved in the assessment (Aux-Bánfi et al., 2015).

For the analyses of the representativeness of the TIMSS sample, using NABC student weights alone is not appropriate. TIMSS weights consist six different factors, school-, class- and student-level sampling weight factors³ are supplemented with school-, class- and student-level adjustment weight factors to adjust for non-re-

³ Student level sampling weight factors are 1 for all students, as all students of a selected class are added to the sample.

sponse on school-, class- and student-level (Martin et al., 2016). In NABC, as it is a census and for schools, classes and students it is compulsory to participate except of students missing from school on the day of assessment, the only weight factor not 1 by definition is the student adjustment factor for non-response (Aux-Bánfi et al., 2015). If we had analyzed the results of TIMSS students (i.e. students selected for participation in TIMSS) unweighted or using NABC weights alone, we would have neglected the correction effects of weighting used in TIMSS to clear estimates of population parameters from biases rising from unequal sampling probabilities and different response rates.

The TIMSS-sample file used in our analysis contains every student in responding schools and classes, however, non-responding schools and classes are not included. Hence, to analyze correctly the characteristics of students in the TIMSS sample we used school- and class-level weight and adjustment factors of TIMSS and the student-level weight factor for non-response of NABC combined, along with the block bootstrap method of NABC for error calculations. We have also calculated the effect sizes of differences, the difference divided by the standard deviation using the estimation method described in Hedges (2007) for nested data with unequal school sizes.

2 Results

2.1 Similarities and differences in the TIMSS and NABC Grade 8 mathematics constructs and results

Both TIMSS and NABC declare their scope and content in their Assessment Frameworks (Mullis & Martin, 2013; Balázsi et al., 2014). The TIMSS framework derives its mathematics scale and test items primarily from the mathematics curricula of the participating countries. In contrast, NABC intends to measure mathematical literacy: “the ability of an individual to understand and analyze the role of mathematics in the real world; the skillful use of mathematical tools; the willingness and ability to use the acquired mathematical knowledge in real life situations; the use of mathematical tools in communication and cooperation during social interactions – on a level adequate for the age of the individual”. During test development, the mathematics Core Curriculum is taken into account to ensure students do not face problems involving mathematical tools and knowledge they did not learn up to that grade. However, according to the definition of mathematical literacy, in NABC mathematics test items are usually not purely mathematical and do not resemble simple textbook examples, but the mathematical problem students need to solve is embedded in some situation similar to the situations in which students should use mathematical tools and knowledge in their everyday life.

The differences in the definition of the two constructs have definite effect on the test booklets. The NABC 2015 mathematics test have a much higher reading load,

70 students need to read approximately two times as many words in NABC than in TIMSS during a 45 minute test period.⁴

Both TIMSS and NABC categorizes test items according to two aspects: their content and their cognitive demand, named content and cognitive dimensions in TIMSS, content areas and thinking processes in NABC. TIMSS assigns target percentages of testing time to different content and cognitive domains. NABC assigns target percentage intervals based on the number of items to every content area – thinking process category pair.

Both TIMSS and NABC have four content categories in Grade 8. TIMSS items can belong into the *Number*, the *Algebra*, the *Geometry* or the *Data and Chance* cognitive category, while in NABC the four categories are named *Quantities, numbers, operations, Assignments, relationships, Shapes, orientation*, and *Statistical characteristics, probability*. TIMSS testing time is divided so 30% of testing time used to solve *Number* items, 30% used for *Algebra* items, 20% used for *Geometry* and also 20% used for *Data and Chance*. In NABC, 35–40% of items belong to *Quantities, numbers, operations*, 25–30% belong to *Assignments, relationships*, 20–25% belong to *Shapes, orientation*, and 12–15% belong to *Statistical characteristics, probability*. The two divisions of mathematical contexts highly overlap (Table 2). For example, most topics from NABC's *Quantities, numbers, operations* category appear in TIMSS's *Number* category, however, NABC puts calculations of specific quantities in relation with geometric shapes, like calculating the volume or area of a geometric shape or using the Pythagorean theorem into the *Quantities, numbers, operations* category, while in TIMSS these items belong to the *Geometry* content category.

On the cognitive dimension, categorizations of the cognitive procedures necessary to solve the test items are even more similar in the two studies. TIMSS uses the *Knowing, Applying* and *Reasoning* categories in a way that 35%, 40% and 25% of testing time is devoted to each. In NABC, 25–30% of items belong to the *Knowledge of facts and simple operations* cognitive category, 45–55% belong to the *Application, integration* category and 20–25% belong to the *Complex solutions and evaluation* category. The three categories used in NABC are almost equivalent to the categories used in TIMSS, both based on Bloom's taxonomy of the cognitive domain. Both TIMSS and NABC uses multiple-choice and constructive response items, in TIMSS at least half of the items, in NABC 55–65% of items are multiple choice according to the Framework.

The correlation coefficient between the TIMSS and NABC mathematics test results is 0.79 (Table 3). While this is a high value, not as high as would be anticipated in two

⁴ To analyze the reading load of the tests, we counted the number of words (with Word's word-counting function) in the 45 minutes long mathematics parts of the fourteen TIMSS 2015 booklets. We have also counted the words in eight 45 minutes long mathematics blocks of the NABC booklets between 2012 and 2015. Pairwise comparison of word counts in the two studies showed that in average the ratio of the number of words in the 45 minutes long blocks in NABC versus TIMSS was 189% (the standard deviation of the ratios was 19 percentage points).

Table 2 Content topics of NABC matched to the content categories of TIMSS (the percent figures in the first column indicate the percent of test items devoted to individual content area).

Content areas in NABC	
Quantities, numbers, operations (35–45%)	Covered in the Number category of TIMSS
	<i>Numbers</i> : number line; intervals; place value; fractions and decimals (equivalence, comparison, reduction, visualization) <i>Calculations, operations</i> : multiple operations (e.g. writing, perform, powers, square root, rounding), data needed for a calculation; calculating percentages of a value, conversion between per cents and fractions, their visualization; calculating with ratios; proportion compared to 1 <i>Measurements</i> : scales (reading and representing data) e.g. thermometers, clocks; comparing quantities; conversion of units; computing with time <i>Divisibility</i> : common divisors, greatest common divisor, smallest common multiple, remainders, divisibility rules
	Covered in the Algebra category of TIMSS
	<i>Calculations, operations</i> : substituting a value into an algebraic expression without rearrangement
	Covered in the Geometry category of TIMSS
	<i>Calculations, operations</i> : operations with geometric shapes (e.g. perimeter, area, volume, Pythagorean Theorem)
	Not explicitly covered in the TIMSS framework
<i>Numbers</i> : scientific notation; <i>Measurements</i> : time zones	
Assignments, relationships (25–30%)	Covered in the Number category of TIMSS
	<i>Proportionality</i> (direct and inverse proportionality, examples of proportions where each value is different from 1): ratio of numbers and quantities; scaling compared to other numbers than 1; calculating the total from percentages and the percentage value of a quantity
	Covered in the Algebra category of TIMSS
	<i>Matching quantities</i> (tables, functions, diagrams, graphs, etc. - not statistical data): reading relationships (value, slope, continuation, evaluation, etc.); representation of relationships (e.g. on graphs, diagrams), examination of representations; writing and application of relationship rules, parameterization, general formula, etc., relationship between variables <i>Parametric algebra</i> : operations with algebraic expressions and formulas with rearrangement; equations and inequalities <i>Sequences</i> : finding the next or a given element using the rule, finding the sequence number of an element, finding the sum of elements (without formula)
Shapes, orientation (20–25%)	Covered in the Geometry category of TIMSS
	<i>Two-dimensional shapes</i> : knowledge of geometric characteristics (e.g. diagonal of a square, angles of a triangle, angles and diagonals of regular and irregular polygons, parts of the circle); transformations in two dimensions: congruence (reflection through a line or a point, translation,

Shapes, orientation (20–25%)	<p>rotation), symmetry, similarity (only based on intuition), completing a pattern; perimeters and areas of two-dimensional shapes (estimation, covering, rearrange parts, relation between parameters)</p> <p><i>Three-dimensional shapes, dimensions:</i> representations of three dimensional objects (views, nets, components, etc.), bounding volumes (e.g. choosing the right box for a present); three-dimensional transformations (rotation, translation, similarity, reflection across a plane– recognizing the result of a transformation based on intuition); relationship between parameters of a three dimensional shape and its volume and surface</p> <p><i>Orientation:</i> directions and cardinal directions, angle of view (based on intuition), locations in coordinate systems (e.g. chessboard, the globe, the Cartesian plane, contour maps)</p>
Statistical characteristics, probability (12–15%)	<p>Covered in the Data and Chance category of TIMSS</p> <hr/> <p><i>Collecting statistical data from tables/diagrams:</i> reading data, comparing data (smallest, largest, differences), evaluating and analyzing data</p> <p><i>Statistical representation and data matching:</i> representing and matching data given in different forms (e.g. in written text, in tables, in diagrams)</p> <p><i>Statistical calculations:</i> e.g. mean (average, weighted average), median, range, mode</p> <p><i>Statistical methods:</i> e.g. choosing, interpreting, using, evaluating the appropriate statistics, identifying the data necessary for a statistics, identifying the statistical properties inferable from a statistical representation</p> <p><i>Probabilities:</i> certain, impossible, possible events, chance, more likely, less likely, frequency, relative frequency etc.</p> <p><i>Combinatory:</i> counting</p> <hr/> <p>Not explicitly covered in the TIMSS framework</p> <hr/> <p><i>Event graphs:</i> counting the edges, paths; <i>Sets:</i> basic operations and their properties; <i>Formal logic:</i> logical values, operations</p>

measurements of the same construct assessed two times in a two-month timeframe.⁵ The correlation between the TIMSS mathematics and the NABC reading results is only slightly lower, 0.75. Recalculating TIMSS mathematics scores based on NABC’s methods using only Hungarian students’ data increased the correlation insignificantly (to 0.80), therefore our conclusions are not affected by the methods used for scaling.

Table 3 Correlation between the TIMSS 2015 and the NABC 2015 results.

	NABC mathematics		NABC reading	
	correlation	(SE)	correlation	(SE)
TIMSS mathematics score	0.79	(0.013)	0.75	(0.013)
TIMSS science score	0.74	(0.015)	0.73	(0.014)

⁵ Cronbach’s coefficient alpha, which is a lowerbound estimate of the internal consistency of a test and also can be seen as an estimate of the correlation between two tests measuring the same construct, is 0.91 for both mathematics tests for the Hungarian students (Martin, Mullis & Hooper, 2016 – Exhibit 11.8, pp11.16, Lak et al., 2016 – Table 2, p. 6).

2.2 Properties of the TIMSS sample

While the two mathematic tests measure slightly different abilities, NABC still can be used to evaluate the TIMSS sample. As the TIMSS 2015 sample is practically a sample of the NABC cohort, the TIMSS sample should be a representative sample of the students in the NABC. The NABC results, as well as other characteristics of students in the TIMSS sample should be similar to the overall national results when using the TIMSS school- and class-level weights combined with NABC student-level weights.

First, we selected to compare the mathematics results of the TIMSS sample with the NABC cohort's (Table 4). The TIMSS sample's average performance was 8.2 points lower than the overall performance of students in Grade 8, which is a statistically not significant difference. Standard deviation estimated from the TIMSS sample was 14.6 points higher than the population parameter, which is a significant difference on a 0.05 significance level, but not on the 0.001 significance level.

Table 4 NABC 2015 Grade 8 mathematics performance of students.

	Number of students (weighted)	Mean performance (SE)	Standard deviation (SE)
Students in the TIMSS sample	81,836	1609.4 (11.47)	209.0 (6.00)
All students	84,108	1617.6 (2.50)	194.4 (1.14)

To evaluate the relevance of the difference in the mean performances estimated from the TIMSS sample and the NABC participants, we also estimated the effect size as a proportion of the standard deviation of the whole population, which is 0.042. A difference of the same effect size in Hungarian students' average performance on the TIMSS scale would be 3.9 points,⁶ not statistically significant. Although the two tests' contents are not identical, and a difference on the NABC mathematics scale could not be transferred directly to the TIMSS scale, these findings confirm that the national average based on the whole NABC cohort probably would lay in the confidence interval of the mean performance of the Hungarian national average published by TIMSS. The two estimations of the standard deviations differ slightly more, indicating that it is possible that on the TIMSS scale, the standard deviation is somewhat overestimated.

Comparing reading results of the TIMSS sample to the whole NABC cohort's leads to the same results: TIMSS students' reading performances are somewhat, but not statistically significantly lower than the national average, the effect size of the difference is 0.046. The standard deviation of reading performances of TIMSS students is 13.2 points higher than the standard deviation of the whole population, which is also a significant difference on the 0.05 level and not on the 0.001 level.

We also compared TIMSS students based on their socio-economic status, where differences were even smaller (Table 5). The effect size of difference in the average SES was 0.009.

⁶ In TIMSS, the average performance of the Grade 8 population was 514 score points (*SE* 3.8), the standard deviation is 93 (*SE* 2.2) (Mullis et al., 2016).

Table 5 Socio-economic status of students based on their answers to the NABC 2015 Grade 8 Student background questionnaire.

	Number of students (weighted)	Mean (SE)	Standard deviation (SE)
Students in the TIMSS sample	67,435	-0.027 (0.0643)	1.037 (0.0313)
All students	66,277	-0.018 (0.0161)	1.019 (0.0076)

Q-Q plots comparing the percentiles of the distribution of mathematics and reading results and SES-indices of the TIMSS sample to the NABC full cohort's same values also show that the sample represents the full cohort very well (Figure 1). The mathematics and reading results show minor differences on the lower and upper end of the distribution: somewhat more students have low results and slightly less students have high results in the TIMSS sample than in the full cohort. The distribution of the SES-index shows an almost perfect match between the two groups of students.

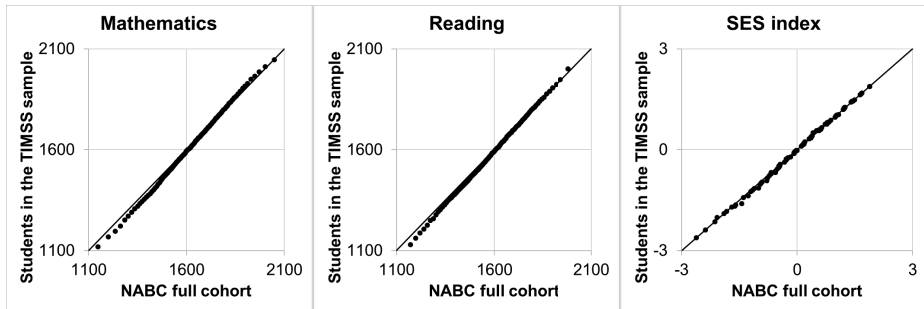


Figure 1 Q-Q plots comparing the distribution of the TIMSS sample to the full cohort's.

3 Discussion

In our research, we compared TIMSS and NABC mathematics scales based on the Framework of the two studies along with the results of students in the two assessments. Although TIMSS and NABC both measure mathematical abilities of students, there are some differences in the two constructs. NABC test items are usually not purely mathematical but mathematical problems embedded in real life situations. While TIMSS also uses problem situations in some of their items, the TIMSS mathematics test mainly contains items more similar to regular examples in a mathematical textbook. Furthermore, while the content and cognitive categorizations and the share of items from the different categories are similar in the two constructs, some minor differences do exist in the frameworks. For example in NABC there is slightly more emphasis given to the *Application, integration* cognitive category, while TIMSS has a little higher percent of testing time for the *Knowing* category. Examining the

correlation coefficient between the two mathematics results (0.79) also confirms that in accordance with the findings from the comparison of the frameworks, the two tests measure related, but not identical abilities.

We evaluated the representativeness of the TIMSS sample using NABC mathematics and reading results and the SES-index of students. Our analysis confirms that the sample of TIMSS represents very well the full NABC cohort, and estimations of population parameters based on TIMSS samples are of a good quality.

4 Further Research

Beside mathematics results, in the continuation of the research presented in this paper, we intend to compare other characteristics of the students measured in the two studies. Student's socio-economic status (SES) is highly correlated with their abilities, and the NABC uses SES as one of the main characteristics in school reports. Hence, its validity is crucial for the study. TIMSS measures the same or similar socio-economic variables, using them to provide international comparisons of the effect of SES on mathematics and science abilities. Therefore, crosschecking the stability of these variables can support the validity and relevance of reports based on SES for both studies.

We also intend to analyze how missing data of non-participating students can distort the results of the studies. On the one hand, we are going to analyze how the NABC achievement of students in the TIMSS sample with missing data compares to the NABC achievement of TIMSS participants and how their participation would have affected the national TIMSS result of Hungary. And, vice versa, we are going to analyze, how students with missing data in the NABC performed in TIMSS. TIMSS uses follow-up sessions for absent students to maximize participation rate, while NABC is written on the same day in every school without any possibility to reach students absent on the day of assessment. Accordingly, the later has a somewhat lower participation rate on student level, 94% compared to 97% in TIMSS. Our research question is whether there are systematic patterns in absent students' characteristics and abilities, and how missing data affects school level and overall results. Similarly, we are going to examine the consequences of non-responding to the student questionnaire in NABC.

References

- Aux-Bánfi, I., Balázs, I., Balkányi, P., Balogh, V. K., Gyapay, J., Lak, Á. R., Ostorics, L. I., Palin-csár, I., Rábai-Szabó, A., Rózsa, Cs., Szabó, Á., Szabó, L. D., Szepesi, I., Szipócs-Krolopp, J., & Vadász, Cs. (2015). *Országos kompetenciamérés. Technikai leírás*. Retrieved from https://www.oktatas.hu/koznevelas/meresek/kompetenciameres/tanulmanyok_publikaciok.
- Balázs, I., Balkányi, P., Bánfi, I., Szalay, B., & Szepesi, I. (2012). *A PIRLS és TIMSS 2011 tartalmai és technikai jellemzői*. Retrieved from https://www.oktatas.hu/koznevelas/meresek/timss/timss_2011_meres.

- 76 Balázsi, I., Balkányi, P., Ostorics, L., Palincsár, I., Rábai-Szabó, A., Szepesi, I., Szipőcs-Krolopp, J., & Vadász, Cs. (2014). *Az Országos kompetenciamérés tartalmi keretei – Szövegértés, matematika, háttérkérdőívek*. Retrieved from https://www.oktatas.hu/koznevel/meresek/kompetenciameres/tanulmanyok_publikaciok.
- Balázsi, I., Lak, Á. R., Ostorics, L., Szabó, L. D., & Vadász, Cs. (2016). *Országos kompetenciamérés 2015 – Országos jelentés*. Retrieved from <https://www.oktatas.hu/koznevel/meresek/kompetenciameres/eredmenyek>.
- Foy, P. & Yin, L. (2016). Scaling the TIMSS 2015 Achievement Data. In M. O. Martin, I. V. S. Mullis, & M. Hooper, (Eds.), *Methods and procedures in TIMSS 2015*. Retrieved from <http://timssandpirls.bc.edu/publications/timss/2015-methods.html>.
- Hedges, L. V. (2007). Effect Sizes in Cluster-Randomized Designs. *Journal of Educational and Behavioral Statistics*, 32(4), 341–370.
- Horn, D. (2013). Diverging performances: The detrimental effects of early educational selection on equality of opportunity in Hungary. *Research in Social Stratification and Mobility*, 32(June), 25–43.
- Kertesi, G. & Kézdi, G. (2016). On the Test Score Gap between Roma and non-Roma Students in Hungary and its Potential Causes. *Economics of Transition*, 24(1), 135–162.
- Lak, Á. R., Palincsár, I., Szabó, L. D., Szepesi, I., & Szipőcs-Krolopp, J. (2016). *Országos kompetenciamérés 2015 – Feladatok és jellemzőik. Matematika, 8. évfolyam*. Retrieved from <https://www.oktatas.hu/koznevel/meresek/kompetenciameres/feladatsorok>.
- LaRoche, S., Joncas, M., & Foy, P. (2016). Sampling design in TIMSS 2015. In M. O. Martin, I. V. S. Mullis, & M. Hooper, (Eds.), *Methods and procedures in TIMSS 2015*. Retrieved from <http://timssandpirls.bc.edu/publications/timss/2015-methods.html>.
- Martin, M. O., Mullis, I. V. S., and Hooper, M. (Eds.). (2016). *Methods and procedures in TIMSS 2015*. Retrieved from <http://timssandpirls.bc.edu/publications/timss/2015-methods.html>.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 International results in mathematics*. Retrieved from <http://timssandpirls.bc.edu/timss2015/international-results>.
- Mullis, I.V.S. & Martin, M.O. (Eds.) (2013). *TIMSS 2015 assessment frameworks*. Retrieved from <http://timssandpirls.bc.edu/timss2015/frameworks.html>.
- Neidorf, T.S., Binkley, M., Gattis, K., & Nohara, D. (2006). *Comparing mathematics content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 assessments* (NCES 2006-029). Retrieved from <http://nces.ed.gov/pubsearch>.
- Sinka, E. (2010). *OECD review on evaluation and assessment frameworks for improving school outcomes. Country background report. Hungary*. Retrieved from <http://www.oecd.org/edu/school/50484774.pdf>.
- Szalay, B., Szepesi, I., & Vadász, Cs. (2016). *TIMSS 2015 Összefoglaló jelentés*. Retrieved from https://www.oktatas.hu/koznevel/meresek/timss/timss_2015_meres.
- Yamamoto, K. & Kulick, E. (2016). TIMSS 2015 achievement scaling methodology. In M. O. Martin, I. V. S. Mullis, & M. Hooper, (Eds.), *Methods and procedures in TIMSS 2015*. Retrieved from <http://timssandpirls.bc.edu/publications/timss/2015-methods.html>.

Ildikó Balázsi
Department for Analyses of Public Education
Hungarian Educational Authority
Szalay utca 10-14, 1055 Budapest
balazsi.ildiko@oh.gov.hu

Ildikó Szepesi
Department for Assessment and Evaluation
Hungarian Educational Authority
Szalay utca 10-14, 1055 Budapest
szepesi.ildiko@oh.gov.hu