2008/ED/EFA/MRT/PI/9

United Nations
Educational, Scientific and
Cultural Organization

Background paper prepared for the

*Education for All Global Monitoring Report 2008*

*Education for All by 2015: will we make it?*

# Results of the National Assessment of Basic Competencies in Hungary

Ildikó Balázsi

2007

# Results of The National Assessment of Basic Competencies in Hungary

Ildikó Balázsi,
Hungarian Center for Evaluation Studies

Assessing Student Achievement in the Hungarian Educational System

## Beginnings

Empirical pedagogical research has a close to four-decade history in Hungary. Our nation has joined the international student assessments relatively early, has participated in various studies from the beginning of the 70's; therefore policymakers and the professionals have continually had information about the performance of the Hungarian students in an international context.

Table 1 summarizes all IEA and OECD organized international student assessment projects that Hungary has participated in the past 40 years.

**Table 1.**
**International student assessments Hungary participated in**

As these studies had large enough student samples, policymakers were able not only to get data about the country's performance on the international scale, but analyses of the various factors influencing student achievement were also possible.

## The First National Studies

Methods and experiences collected through participation in the international studies have helped planning and implementing the first National representative sample based study, named TOF-80 (Báthory, 1997; Báthory and co, 1983) organized by the Department of Curriculum Theory of the National Pedagogical Institute for grades 4 and 8 to assess student achievement in various subject areas. This study was not followed by another assessment cycle, but its experiences later benefited those implementing the regularly repeated Monitor Studies.

The Monitor Studies, first implemented in 1986, were created with the goal of regularly providing data on student performance and the factors affecting the National educational system in grades that are important milestones in the Hungarian educational system. The Monitor Studies did not assess curricular knowledge, but aspects of reading, mathematics, science, IT and cognitive skills; students had to perform using skills and knowledge. A drawback of the tests was the fact that it contained multiple choice questions only. The second assessment cycle took place in 1991, then every two years till 2005. (Vári (1997a), Vári (1997b), Vári et al. (1999), Vári et al. (2000), Vári (2000).) The grades assessed have changed during the various cycles, but most featured grades 4 and 8 because of their importance in the Hungarian educational system. In addition to this, grades 6, 10 and 12 recurred as target populations throughout the years.

## *National Assessment of Basic Competencies (National ABC)*

### Reasons for the implementation and goals of the National ABC

A change of paradigm in the concepts of knowledge became accepted in the Hungarian educational system, which resulted in placing more emphasis on developing competencies next to domain specific knowledge. However changes in school practices, implementation of new methodologies cannot be realized from one day to another. One of the goals of the National ABC study, first implemented in 2001 by the Center for Evaluation Studies (CES) and Education, Assessment and Examination Center (EAEC), is to ease the transition as well as speed it through the assessment of competencies – a process that would otherwise take considerably longer time to reach schools. The study also informs schools of competencies that each student should posses at a given grade level; therefore it's also a tool of presenting the schools with a set of outgoing requirements.

Another policy matter of the Hungarian educational system is the fact that the accountability of schools and teachers has become a dominant issue that policy makers have created numerous orders and laws to enforce. The institutes of professional school inspectors' have ceased to exist in the beginning of the 1990's, since then schools and school maintainers have not received any feedback on the quality of schooling apart from the law enforced reviews and the results of the school leaving exam. The National ABC helps schools' internal evaluation processes by providing them data about their performance, and it is amongst its mid-term goals to disseminate the results to the general public either through the schools or with regular printed media.

The goal of the National ABC studies is to assess whether students are able to use their knowledge and skills in real life situations and gain further information – and not to assess the level of student success in internalizing factual knowledge set by the curriculum. The assessment tests[1] therefore do not measure curricular requirements, but rather how well students can activate what they have learned and can use them in problems relating to everyday situations (Balázsi et al. (2006)). The study currently focuses on reading comprehension and mathematics; long-term goal is to include science in the domains assessed as well. The first assessment included grades 5 and 9, then in 2003 grades 6 and 10. From 2004 the assessment covers grade 8 as well, the target populations of the National ABC study today are grades 6, 8 and 10.

Table 2 shows the dates, grades and domains assessed in the National ABC studies to date.

**Table 2**
**Main aspects of the National ABC studies to date**

While planning the National ABC studies we have focused on the following criteria. The assessment should:

- help create and form the internal evaluation practices of schools,
- help teachers familiarize themselves with the tools and methodology of the competency-based assessments,
- provide schools with data and practices to carry out objective local, institute-level evaluations,
- provide maintainers of schools with data that is reliably comparable with the National performance data.

[1] Test booklets and background questionnaires are available at the www.kompetenciameres.hu website (in Hungarian).

## Characteristics of the assessment

In order to ensure these goals, the Ministry of Education has decided to implement the assessment in all schools in Hungary with the participation of all students in the populations measured. The assessment aims to provide schools not with individual data, but rather a data pool about the student body. So the central analyzes are carried out based on a given sample of students from every school. Naturally, schools have the opportunity to evaluate student data not included in the nationally representative sample using the software the CES has provided them with.

In the first three studies 20 students per school were selected using random stratified sampling and were entered into the central database, however, the sampling process has changed in 2006. In order to provide schools with classroom level data, we have decided to process the entire 8th grade student data centrally. We have also increased the 10th grade sample to 30 students per school, but for the grade 6 data from only a representative sample of schools were collected for central processing and analysis. Through this, we can carry out the national analyses based on the student data of 6th graders of 200 schools, the rest of the schools can analyze their own data with the CES provided software. Since most schools that have 6th grade classes also have 8th graders as well, all schools receive a report based on at least one grade's performance.

In order for schools to get an accurate picture of student performance and rank themselves in the group of schools with similar attributes, participants were asked to fill in the School Principal, School Site[2] and the Student Questionnaires. The questionnaires collect basic information necessary to get a picture of characteristics of the students' family background and the schools' attributes. Using their information, we can conduct more in depth analyses of student performance.

The study assesses student populations every second year from grade 6 to grade 10. However, linking performance data has not been resolved; the added value cannot be measured. Tracking student performance is problematic at best, because of Hungarian data protection laws central data cannot include personal information. Students are identified by numbers, only the schools are able to link IDs on test booklets with the names of students. The need and intent for development of these issues exists, it is possible that the innovations of upcoming years will move the current system along this line.

The other possible way for improvement in the upcoming cycles is to include science in the assessment after creating a framework and an item pool for assessment for the domain. Currently we have two versions of the test booklets with overall about 60-60 mathematics and reading items, hence measuring achievement on subtests is not feasible. A revised assessment system based on blocks would allow for measuring science competency as well as a generally better coverage on the other two domains.

[2] A school site is a unit of a school that has a separate mailing address. A lot of small village schools have been united into one institute due to financial considerations in Hungary in 2005, turning the former school buildings of the villages into a site for a larger institute.

## Reporting the results

The National ABC takes place in May at the end of the school year. Each year we provide feedback based on school site, schools and school maintainers as well (ÉK (2007a)) by the end of following February. In addition of the school site, school or maintainer related results, this report contains data about the distribution of student performance by settlement and school type, and on a regional basis.

From the report schools learn their average performance, the student distribution on the scale and levels of competencies from mathematics and reading. It is also noted whether the school shows a significant change compared to the pervious assessment. In addition, the report shows how the supposed performance relates to the students' socio-economic index. The student background index is formed based on the Student Questionnaire. Its effect on the school site average is estimated using linear regression in all cases where the school site collects the sufficient number of Student Questionnaires, so it can be noted where the school site appears relative to the nationally representative line based on the results of all school sites.

All schools receive all their school sites' reports and a school level report. Maintainers get a maintainer level report and gain access to all their school and school site results.

Booklets that remain at the schools are available for analyses by the Center's software that calculates performance scores after entering the appropriate codes for the responses and presents the data in figures and tables.

# *Results of the 2006 NATIONAL ABC Study*

## The Scale and Levels of Competencies

We used a two-parameter model to calculate student competency in reading and mathematics and the difficulty and slope of the items.[3] In the 2003 assessment student scores were analyzed after standardization: the national average and standard deviation after transformations was 500 and 100 standard point in mathematics and reading for grade 6 and 10. The same process was done with the parameters of grade 8 in 2004 (Balázsi 2006, ÉK 2007c).

Since all students in the given populations participate in the testing, all tests become public after the assessments. Therefore trend analyses cannot be done due to the lack of common test items. In order to enable us to compare the data from various years of assessment, a representative sample (of about 3000 students) fills in the so-called Core test following the main assessment – a confidential test that remains the same year after year. Items of the Core test serve as a link between assessment years that help us put previous results on the same scale and enable the creation of trend data.

Levels of competencies created upon professional and statistical guidelines play an important role during our data analyses. These help us sort students into different categories based on their competencies, where we state what the *minimum* of competencies is on a certain level, what is

[3] We used the Parscale 4.2 software to calculate the parameters.

that they are lacking compared to those on a higher level. We set up ***four levels of competencies*** in both domains. Definition of the levels is based on methods used in the PISA 2000 study (Ray & Wu (2002)).

The levels of student performance show us a hierarchy of sorts of the competencies. *Those students who reach a certain level not only possess all skills and knowledge typical to that level, but also those on the levels below it.* Therefore a student whose performance ranks on level three will naturally have completed the requirements of the first and second levels. A student on a given level will probably be able to answer at least half of the items ordered to that given level.

The following two tables briefly summarize the expected processes and procedures on the respective levels. These definitions are valid for all the grades assessed, but naturally we take the age-specific knowledge and skills of the students on the given grade levels into consideration.

**Table 3**
**The skills of students on the performance levels in reading**

**Table 4**
**The skills of students on the performance levels in mathematics**

## Results

Based on the results of the past three assessments we note that the results have not changed considerably, though the differences are significant due to the large sample size. The biggest difference is visible in the case of the 6th grade reading performance, where compared to 2003 even in 2004 there was a 9 point, and then by 2006 an additional 3 point, altogether a 12 point increase in average scores (1. figure). This shift is recognizable not only in the averages, but also in the percentiles (ÉK 2007b). At the same time, we need to consider that the National standard deviation is 100 points, so this increase is not considered overwhelming. Results from upcoming assessments will decide whether this is indeed a positive trend or a random fluctuation.

**Figure 1**
**Student performance distribution in grades 6 and 8**

If we compare the distribution of students on the performance levels, we see a considerable difference in mathematics and reading in terms of the number of students on each level. Overall, reading results are better than those in mathematics. While there were less students to reach the upper 3rd and 4th performance levels in mathematics, the number of students on level 1 and below level 1 is almost twice the size of that in reading. However, we cannot call the reading results good either, considering that about 20-28% of students, about a quarter of the population falls into and under level 1.[4]

**Figure 2**
**Student distribution on performance levels in grades 6 and 8**

[4] In National ABC we have set level 2 as the minimum, where students are still able to use their mathematics and reading skills and competencies to further gather knowledge.

It is a peculiar phenomenon that while the number of students on the upper levels increases in mathematics as the grades progress, reading results get worse. This is not necessarily a decline, since the requirements of each level become higher as the grades progress and the texts and items get harder parallel to the increase of the population's age.

The fact that the 8th grade mathematics are better could be explained by how student rarely meet tasks in grade 6 related to application, modeling, interpretation and mathematization while these are common in upper classes – 8th graders are more familiar with these than their younger schoolmates.

The decrease in reading performance in grades 6 and 8 could be caused by the fact that Hungary does not have classes specifically devoted to reading, teachers hardly ever focus on this. Most often textbooks include questions about concrete facts discussed in previous sections; oral presentations generally focus on data as well. Students are rarely asked to interpret and contextualize what they have read or reflect on themes discussed.

## Gender differences

We get the same result well-known from international studies that boys perform better in mathematics and girls do better in reading.

**Figure 3**
**Student distribution on the competency scale by gender in grades 6 and 8 in the 2006 National ABC Study**

**Figure 3.**
**Student distribution on the performance levels by gender in grades 6 and 8**

While in mathematics this difference is measurable, it is not overwhelming, altogether 10 (standard error: 2.1) and 11 (standard error: 0.6) points in the two grades respectively. This difference in averages is the result of the higher numbers of boys among those who performed exceptionally well. While in the case of the 5 and 25 percentile the difference between boys and girls is small, it is more than 20 points at the 95 percentile. We get a similar picture when observing the distribution by performance level, the number of boys and girls is about the same below level 1, but the number of girls on level 2 is higher, while on levels 3 and 4 there are more boys than girls.

The girls' lead is much bigger over the boys in reading. In grade 6 the difference is 37 points (standard error: 1.7), in grade 8 it is 43 points (standard error: 0.5), it is almost half a standard deviation. This difference is present on the overall distribution, there is no considerable lag in the performance of boys in terms of the percentiles. Furthermore, boys fall back even more by grade 8 in reading. (It should be noted that by grade 10 the difference becomes even smaller, than in grade 6; this can be caused by the bigger dropout rate of boys.) This tendency is well visible throughout the distribution of the performance levels; there are almost twice as many boys below level 1 as girls, while there are only half as many boys on level 4 as girls.

## The Results by Geographic Location

Hungary is divided up to 19 counties and the capital city, these are grouped into 7 geographical regions. The North-Western region, Northern Transdanubia and Central Hungary are the most developed. Central Hungary owes most of its significantly higher per capita GDP to Budapest that was 3 210 000 HUF (1 USD ≈ 200 HUF). It is also a fact that the Eastern regions of the country are generally more backward economically; the state of infrastructure and the number of unemployed are the least favorable here, also. Figure 5 show the location of the regions and the per capita GDP in HUF. Central Hungary's per capita GDP is shown without the inclusion of Budapest.

**Figure 5**
**The regions of Hungary and the per capita GDP in 2004 in 1000 HUF units (source: www.ksh.hu) 5**

Figure 6 shows data from the 2006 National ABC study by region, where we can see that in accordance with the economical indices the Northern Hungarian and the Northern Great Plain regions have the lowest average performance, student in the Western Transdanubian region have the highest average performance on both domains and both grades as well. In most cases significant differences in the data of the other regions cannot be found. The difference between the two extremes is about 30 points in all cases, about a third of the standard deviation. The National ABC results more or less reflect the differences that the economic indices do; they follow the highs and lows of socio-economic development. Naturally, this relationship is not universal, as performance of the Southern Great Plain does not fall back compared to the other regions while the per capita GDP ratio is almost as low as that of the Northern Great Plain region.

**Figure 6**
**Student distribution on the competency scale in the regions in grades 6 and 8 in the 2006 National ABC study**

## The results by type of settlement

The previous national Monitor Studies have reported about the differences in performance between the various settlement types, the study cycles showing bigger and bigger differences between the performance of schools in the capital city and villages. We differentiate between four settlement types: capital city, county capitals, cities and villages. In the 2006 assessment students studying in the capital city have performed best followed by a small difference of 3-10 points the students of county capitals. The gap before the city schools' students is considerably larger, about 30 points, about a third of the standard deviation. The lowest average performers are the village schools' students with another circa 30 point drop.

[5] Central Hungary's per capita GDP is shown without the inclusion of Budapest.

**Figure 6.**
**Student distribution on the competency scale by settlement type in grades 6 and 8 in the 2006 National ABC Study**

The performance differences between settlement types are similar to the differences in socio-economic status between the settlement types. The gap becomes even greater as it is a general trend for an intellectual, well-off village family to send their children to a nearby city school. In Hungary, parents have the right to choose the school for their children freely.

The following section shows that these differences between settlement types disappear, as we take into consideration the students' socio-economic and cultural status. Even more so, the effects of capital city Budapest turn for the adverse.

## The effect of the socio-economic and cultural background on performance

Numerous questions of the Student Questionnaire map the social, economic and cultural capital of students' and their families. To examine the effects of the socio-economic and cultural background we used hierarchical regression analysis, a two-level linear hierarchical model (Raudenbush & Bryk (2001)) where students are on the first level, classes are on the second (ÉK 2007c).

Table 5 shows the intercept and the fractions of variance in student performance originating from within-class differences and between-class differences in the two-level linear hierarchical base model. The last row shows that the differences between classes are responsible for 24-30% of the overall variance. This is a relatively high rate that shows how the average performance of students in various classes can be considerably different – it is a deciding factor what school and what class the student attends, the class limits performance to a certain extent. The PISA study examined the proportion of variance caused by the between and within school differences, and showed much less values of 10% or less in certain countries.

In the case of Hungary PISA estimates the rate of variance to be much higher resulting from the between school differences, but this does not contradict the findings of the National ABC study, as the PISA assessed 15-year old population attends secondary level educational institutes. At the beginning of the secondary education the school tracking system causes a highly polarized selection mechanism. Students have the opportunity to apply to 8 and 6 year secondary academic school upon completion of grades 4 and 6, in reality however only a small portion of students take this opportunity. About 90% of students finish upper elementary level schooling in the traditional primary schools. Segregation in the public educational system takes place after the 8$^{th}$ grade.

**Tabel 5**
**Dividing the variance of student performance according to within-class differences and between-class differences**

We have included those background variables in the model that best describe the socio-economic and cultural status of the students. The variables used are listed in Table 6. Filling in the Student Questionnaire was not mandatory however 90% of grade 6 and 85% of grade 8

students answered the questions. The number of responses may have been lower on certain variables as the table below shows.

**Tabel 6.**
Variable used in the linear hierarchical model

As the table shows, not only did we use student level values, but with some variables the class averages as well. It is a well known fact of the international literature that the socio-economic and cultural background effects through the average socio-economic and cultural background of the class. The ordinal variables were not dichotomized as the analysis has shown that they have roughly linear effect on the performance.

Apart from background variables we have accounted for the type of settlement as well, since we assumed that the effect of the socio-economic background can be different in various types of settlements.

Variables which did not have a significant effect, were eliminated from the model one by one resulting in the model in Table 7.

**Table 7**
**Variables effecting student competencies and their effects**

The explanatory value of the models is 34-37%, meaning that the background variables included in the model account for a 34-37% of the variance. The fraction of variance caused by the between class differences has decreased by 72-80%, the fraction of variance caused by the within class differences has decreased by 19-23%.

As one can see, in the case of both the grades and assessment areas, the parents' educational level and the number of books have the biggest influence, though they act in different ways in each grade. The class average does not have a significant effect in grade 6; grade 8 shows a slight decrease in the effect of student level parents' education while the effect of class level parents' education becomes significant. It is peculiar that the parents' employment situation has not been a deciding factor when considered alongside the other variables.

Among the variables of social allowances grade 6 showed a significant effect with regards to educational aid, grade 8 with regards to free meal services – resulting in an overall 7-12 point difference in the estimated performance while other variables were the same.

Personal computers in the students' home only effected mathematics performance which in itself is not really surprising, but the class average of the variable in grade 6 shows a strongly significant positive effect on both domain performances. This may be a more complex effect originating from overall economic situation of the family that in the case of grade 8 is triggered by the class average of the parents' educational level and number of books.

The negative value in the table related to the variables internet connection, own books and own desk overall reflect a positive effect, since the 'yes' responses received a score 1 and 'no' responses received a 2. We have marked those rows in italics where the positive response received a smaller code than the negative hence resulting in an inverse sign. The effect of own desk is significant only in the reading performance of 6 graders. Own books have the greatest effect – however, we have to note there have been hardly any 'no' responses to this question.

It is strange how 'parents helping the student' had a negative effect, but causal relationships will most likely become inverse: parents study with those who get behind and need help constantly.

After school lessons also have an inverse value, the effect of mathematics lessons is highly adverse, foreign language and music lessons show a positive effect. It is highly probable that this ties into our previous phenomenon: while foreign language and music lessons are an indicator of the financial status of the family, mathematics lessons are for those who cannot learn the materials taught at schools.

Out of the dummy variables of the settlement type we have the variable of the capital city left in the $8^{th}$ grade model, surprisingly with a negative effect. This could be caused by the fact that students generally have the best socio-economic status in Budapest, and after a point the performance cannot follow the increase of the social status.

## Using the results

Schools are the primary users of the outcomes of the National ABC. In the early years only a few paid attention to the results and even fewer could analyze them at schools but surveys about use of data show that more and more schools interpret and process results, hold meeting more often discussing their standings (Lannert & Nagy (2006)). Regulations force schools to continually evaluate their scores. On one hand, the Public Education Act has an amendment that makes it obligatory for lower achieving schools to make provisions; on the other hand the Act pronounces that among other things the annual quality control documents must be based on the assessment data.

Apart from school level use of the data, the national reports are becoming an important part of policy decision making as well, along with the use of the assessment database for research purposes. A great number of secondary analyses have been prepared that reflect on deeper relationships within the results (Lannert & Nagy (2006)), more and more researchers and PHD students use the database and the national report for their work.