(Mis)use of National Assessment Results in Slovenia

Darko Zupanc, MSc, National Examinations Centre

Abstract

National assessment of students' knowledge is frequently carried out in school systems throughout the world and different countries participate in international comparative studies such as TIMSS as well. The elementary reporting on results is a demonstration of the achievement of multi-level benchmarks. This article deals with the distribution of results in mathematics of 13 and 14 year-old students in systems of national assessment in Sweden, England, the USA, Canada, Australia and Slovenia, and aims at comparing these with the results of the international study TIMSS 2003. In national assessment Slovenia shows a pronounced discrepancy: the majority of students are very good at mathematics and the number of those that do not attain basic knowledge is negligible. But is the fact that more than one half of the students in Slovenia achieve excellent or very good grades in mathematics in national assessment good or bad? What is the case in other countries? How do other countries compare in their distribution of grades, or levels of achievement in their systems of national assessment? Do international measurements of knowledge confirm this anomaly revealed by school grades and national assessment in Slovenia? Isn't this embellished demonstration misleading and in reality all about lowering standards of knowledge and inflation of grades? The function and use of national assessment results in the last few years have been questionable.

Key words: achievement levels, (inter)national assessment, TIMSS, use of results

1. Introduction

From 1989 onwards, governments in former socialist countries in Central and Eastern Europe gave high priority to reforming their systems of education (West, Crighton, 1999). Previously, most countries in this part of Europe had had a uniform national system with a controlled input: the curriculum, timetables, learning content and methods were prescribed and a single textbook approved for separate subjects in each grade. The output, however, was only rarely assessed by any kind of uniform system. At the conference for the Organization for Economic Co-operation and Development (OECD) on secondary education in Vilnius (1995), countries placed assessment reform at the top of their list for future tasks, putting it before other important tasks such as teacher training, textbook regulations, and education funding.

The source of change to the educational system in Slovenia can be traced back to the development of examinations at the conclusion of secondary education – the *matura*. This includes the implementation of a new, external, centrally administered *matura*, formerly unknown in this educational system. In 1995 this was a new approach, the first in any of the central or eastern European countries (Gabršček, 2001). In addition, in Slovenia during the 2001/02 school year, national assessment (NA) was carried out for the first time after the 3rd, 6th and 9th grades, along with the trial introduction of the new 9 year primary school programme (Devetletna osnovna šola [Nine year primary school], 2001).

The tradition of an educational system is powerful and any major change affecting the existing system causes strong reaction. According to Crighton (1993), changes to the systems of assessment as well as examination, cause dismay and uncertainty in the society: parents are worried that their children will be deprived because of the changes; teachers are worried that they will not be able to prepare their students well for the new conditions; politicians are afraid that standards will go down or that an independent assessment would reveal some sobering and unpleasant truths about what students really learn in schools.

According to several numerical indicators in the field of education, Slovenia has made giant strides internationally (Tessaring, Wannan, 2004). A high proportion of Slovenia's gross national income is intended for education; a high percentage of the appropriate generation is involved in secondary and different types of higher education; there has been a significant increase in the number of secondary school and university graduates. The *World Declaration on Education for All* (UNESCO, 1990: 9) emphasised years ago that – in addition to statistical indicators of growth – it is important that increased opportunities for education are converted to the development of the individual as well as the society: the decisive factor being how much, and what, the individuals have learned because of increased educational opportunities. Therefore, in Slovenia too, the question as to what kind of knowledge hides behind truly encouraging numerical indicators is brought up. Does the enormous growth in the number of graduates, a high proportion of them deemed excellent students with a multitude of excellent grades, truly reflect a high degree of knowledge; or is it perhaps all about the 'lowering' of criteria and inflation of grades, certificates, degrees and qualifications? What is the distribution of grades and the reaching of achievement benchmarks in systems of national assessment of other countries? Do international measurements of knowledge confirm this exceptionality revealed by school grades and national assessment in Slovenia?

Within Slovenia, a confrontation of two perspectives and approaches in the development of the educational system has been brought about: on one hand, standards of knowledge ought to be set through the curriculum and its achievement be assessed by external tests (Zupanc, 2005b); on the other hand, due to the introduction of new, for Slovenia and this part of Europe, non-traditional approaches, some experts, as well as the public, talk about school and in particular external assessment being stressful. They talk about students being overtaxed, about learning for the test etc. If in an extensive educational reform, external testing showed poor results, opponents of the reform and external testing would claim that the poor results were caused by the reform and the external testing.

2. Methodology

National assessment has been carried out in some Slovene primary schools from the school year 2001/02 onwards, and data for the four years have been collected. Basic data and results are recorded in annual reports (Slavec Gornik, 2002; Štraus, 2003). By studying relevant literature and statistical data on national assessment in Sweden, England, the USA, Australia, Canada and Slovenia, similarities and differences were identified.

At the end of the 2004 calendar year (Mullis et al., 2004) results of the international study of trends in mathematics and science (TIMSS 2003) were published. Countries included in the analysis of national systems of assessment were chosen for the study because all of them participated in TIMSS 2003 with their older primary-school students. A joint international study, which included all the above-mentioned countries, was the basis for a comparison, an analysis and an interpretation of the findings. The study focused on measured achievement in mathematics in the higher primary school grades with the emphasis on reaching certain benchmarks. National and international distributions of benchmarks of Slovene students essentially differ from distributions in other countries.

3. International and national assessment

Comparison of results of national assessment between countries is problematic, especially if we fail to take into proper account what students learn up to a certain age or in a certain number of years of schooling. We can compare results and analyses of national assessment which have been scrutinized by experts, the Ministry and the public. Reports of national, as well as international assessment, distribute students according to achivement on several levels. Approaches differ from country to country, and the number of levels reported as well as the interpretation of achievement on each level may vary. The emphasis in the article is on the comparison of distributions, not on comparing the achievement of individual groups of students in the absolute criterial sense. The issue discussed is, what kind of information on achievement do students, parents, teachers, other experts and the public get from national assessment and how valid is this on an international scale? Is feed-back on a proportion of students with above average, average and below average achievement embellished or realistic?

3.1. International assessment

International assessment includes measuring achievement, usually simultaneously, in educational systems in several countries. Assessment is undertaken on a sample of students of a certain age or in a certain grade in each participating country and comparative analyses of data are produced (Kellaghan, Grisay, 1995). TIMSS 2003 (the Trends in International Mathematics and Science Study) is the third in a continuing cycle of the international mathematics and science assessments of 9 and 13 year-old students in primary schools (Japelj Pavešić et al., 2004). It was conducted from 2001 to 2004 as a joint project between 51 countries and benchmarking participants – members of the IEA (International Association for the Evaluation of the Educational Achievement).

Students' achievement in TIMSS is shown on a scale from 0 to 800 score points (Mullis et al., 2004). On this scale four benchmarks are identified (at 400, 475, 550 and 625 score points), which distribute groups of students according to achievement in five benchmarks of achievement:

- does not reach Low benchmark of achievement (below 400 score points),
- reaches Low benchmark of achievement (400 score points or above),
- reaches Intermediate benchmark of achievement (475 score points or above),
- reaches High benchmark of achievement (550 score points or above) and
- reaches Advanced benchmark of achievement (625 score points or above).

The actual scored points do not reveal what kind of knowledge and skills students having scored a certain number of points have. In TIMSS, one can compare countries according to the percentages of students reaching the various benchmarks. Benchmarks determine knowledge and skills that students reaching a certain benchmark are likely to possess. Students reaching a certain benchmark also display knowledge and skills of all lower benchmarks.

Table 1 shows the distribution of older students (8 years of schooling, or 14 year-old students) in countries, or school systems in provinces, according to the international benchmarks in mathematics.

TIMSS 2003	**Does not reach Low benchmark**	**Low benchmark**	**Intermediate benchmark**	**High benchmark**	**Advanced benchmark**
Quebec – Canada	1%	11%	43%	37%	8%
Ontario – Canada	3%	22%	41%	28%	6%
the USA	10%	26%	35%	22%	7%
Australia	10%	25%	36%	22%	7%
England	10%	29%	35%	21%	5%
Sweden	10%	30%	39%	18%	3%
Slovenia 8th grade/9	15%	33%	37%	14%	2%
Slovenia 7th grade/8	10%	31%	38%	18%	3%

Table 1: Distribution of students reaching mathematics benchmarks (older population) in TIMSS 2003. (Mullis et al., 2004; Japelj Pavešić et al., 2004)

A gradual reform of primary education has been carried out in Slovenia within the past few years. The number of years of compulsory primary education has extended from eight to nine years due to the earlier admission of children into schools at six years of age. In 2003, the majority of students were still enrolled in the old system of 8 year primary education, and only a minority – taking national assessment tests – in the new system of 9 years of primary schooling. The sample of older students, who took part in TIMSS within Slovenia, comprised 3,274 seventh-graders of 8 years of primary schooling and 609 eighth-graders of 9 years of primary schooling. In Slovenia, Advanced and High benchmarks in mathematics are reached by 3% and 18% of students respectively. Percentages of Slovene students reaching these two benchmarks are considerably lower than the corresponding percentages of students in other countries of the European Union. In addition, the achievement of students from the new system of primary education proved to be lower.

3.2. National assessment

Greaney and Kellaghan (1996: 5–10) provide reasons for the introduction and use of national assessment in educational systems. In defining school policy, relevant data are of utmost importance. It is important to know what results students in different curricular areas achieve and how the achievement changes with time. It is important to have data on realistically achieved standards, not only on the desired ones. Often, the achievement correlates with external factors which (school) policy can influence as well as with other means: this is why it is important to be aware of correlations between various factors prior to directing activities, making decisions and taking measures. It is easier for a school policy to focus the attention of professionals in schools on being more effective in different areas of school life – which can have an indirect influence on better achievement in education – by national assessment rather than public examination. Results of national assessment bring relevant data for professional and public debate on the educational system within a country. Sometimes these data can cause problems for politicians and civil servants in the short term; in the long term, however, open professional debates can lead to reflection, incentive and confirmation of the reform initiative.

When reporting on national results in international assessment, different kinds of approaches are used. Some include the average benchmarks of students' achievement in curricular areas. Others include reporting on the percentage of students reaching certain benchmarks. Some use established benchmarks of achievement determined by statistical methods and experts' evaluations in presenting results of national assessment. Sometimes a benchmark is named for a certain characteristic verbal designation.

Results of national assessment have the potential to influence the carrying through of a curricular reform and the writing of textbooks and therefore provide relevant information for in-service teacher training.

3.3. Sweden

The system of assessment in Sweden is based on a three-level, or four-level assessment scale (Guide for Foreign Researchers, 2004). Possible grades are:

IG	'Icke Godkänd'	Fail
G	'Godkänd'	Pass
VG	'Väl godkänd'	Pass with Distinction
MVG	'Mycket väl godkänd'	Pass with Special Distinction

If a student in any subject fails to achieve a grade **G** – Pass, no grade is given.

The aim of national testing is to help teachers standardise their grading within the classroom (Wolf, 2000). In 2001, the distribution of 9th grade students in the national assessment of mathematics achievement was as follows (Guide for Foreign Researchers, 2004):

	Proportions [%] of students with a certain grade (and without one)			
Subject	**IG**	**G**	**VG**	**MVG**
Mathematics	13	51	27	9

Table 2: Reached benchmarks of achievement in the 2001 Swedish national assessment

In 2001, 13% of students at the completion of primary school who failed to achieve knowledge and skills in mathematics failed to be awarded a 'Pass' grade.

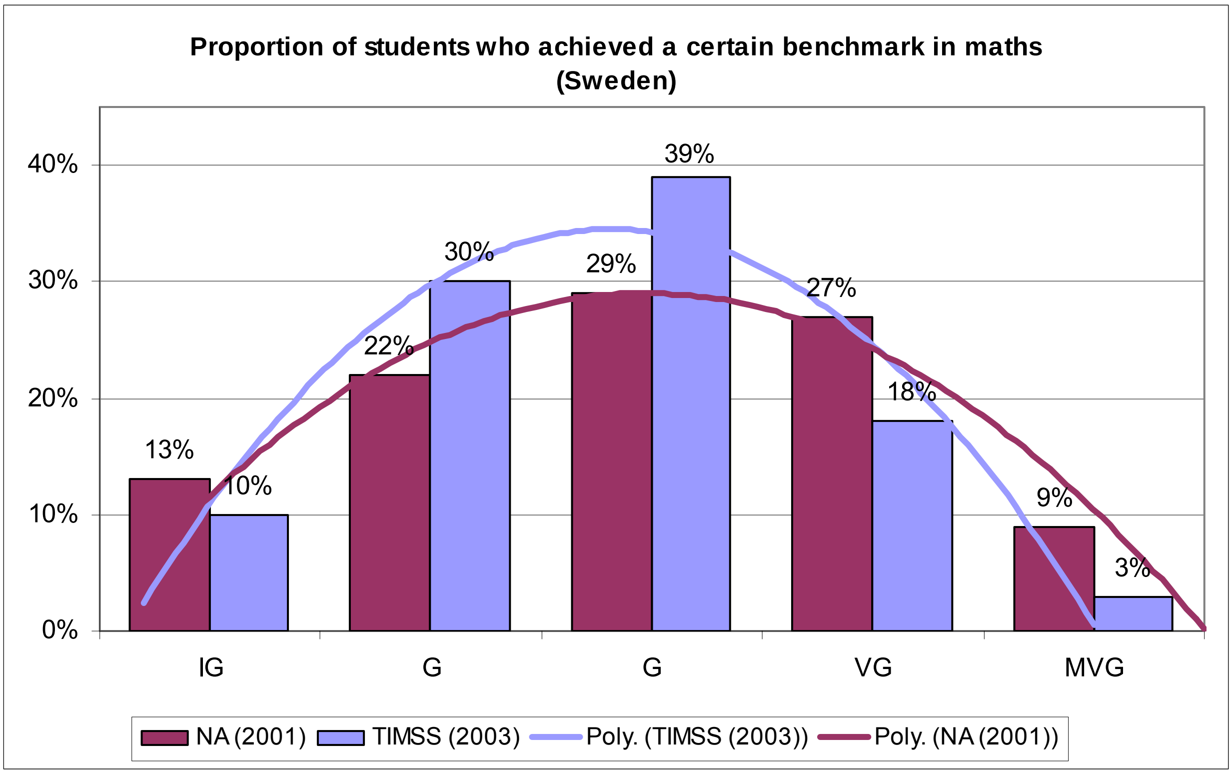

Diagram 1: Comparison of student distribution according to reached benchmarks in mathematics in national assessment (2001) in Sweden and the international TIMSS (2003). (IG – 'Fail'; G – 'Pass'; VG – 'Pass with Distinction'; MVG – 'Pass with Special Distinction')

In TIMSS, benchmarks of achievement are labelled Low, Intermediate, High and Advanced as well as 'Does not reach Low benchmark'. This scale, then, has one benchmark more. If we try to represent both distributions on the same diagram, we can show the lowest and the highest two benchmarks (Fail = 'Does not reach Low benchmark' and Pass with Distinction = High and Pass with Special Distinction = Advanced) at the same spot, while benchmark G – Pass is proportionately distributed between two TIMSS benchmarks, Low and Intermediate.
The overlapping of distribution is not perfect, but the trend curve reveals that the information obtained in Sweden from their national assessment does not significantly differ from the international comparison TIMSS. It turns out that less than 10% of students at this age achieve outstanding knowledge and skills, while 10% or more fail to achieve the set goals.

3.4. England

At the end of Key Stage 3, 14 year-old students in England take national tests in English, mathematics and science. In accordance with national curricular standards, the majority of students should progress approximately one level every two years. In England, levels of achievement are defined for the entire period of compulsory schooling on a scale from Level W – 'Working towards Level 1', Level 1, Level 2 and all the way to Level 8, as well as the exceptional Level E. At the age of 14, the expected level for the majority of students is 5 or 6; students with the lowest measured achievement reach Level 2, while the best students reach Level 8 or even the exceptional Level E.

Table 3 summarises data on the reached achievement levels in mathematics of 14 year-old students in 2003.

KeyStage3 – 2003	Reached certain level [in %]									
	W	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**E**
Mathematics – 2003			1	7	15	22	25	19	5	

W – 'Working Towards Level 1' E – 'Exceptional Performance'

Table 3: Reached benchmarks of achievement in the English national assessment (National Curriculum Assessments – Provisional, 2003)

It is important to emphasise the range of measured achievement in 14 year olds: the majority reach the national goal for this age – Levels 5 or 6.

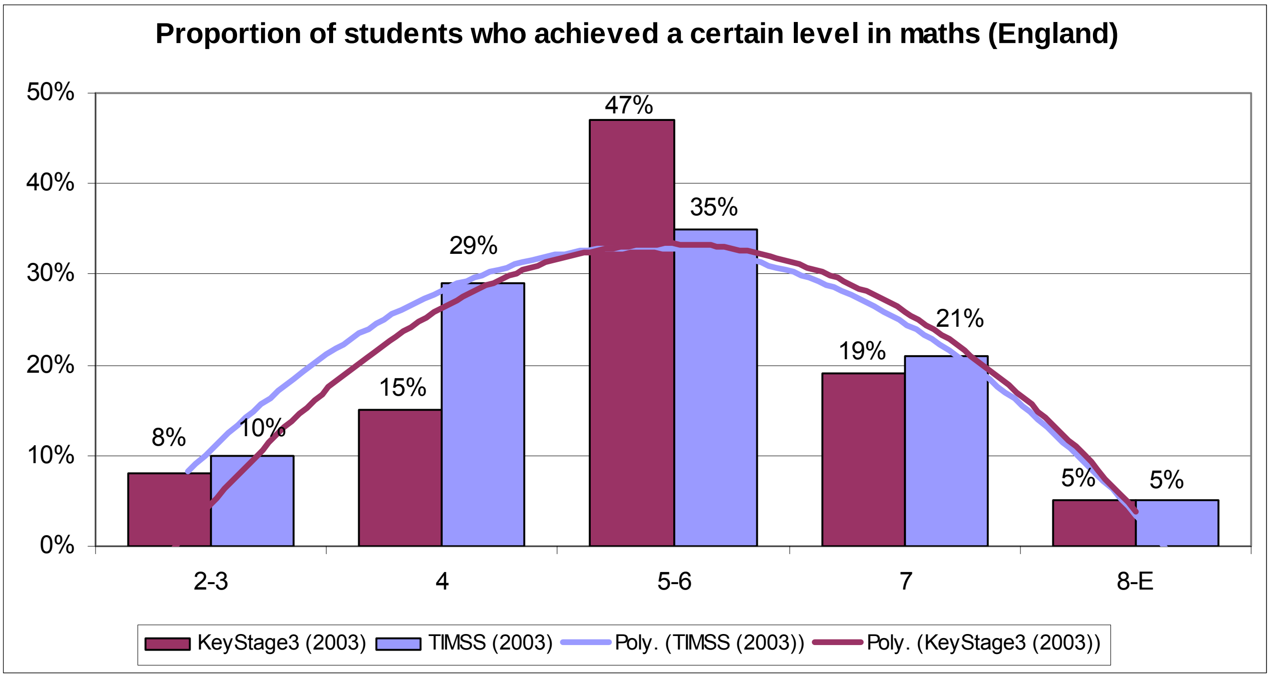

Diagram 2: Comparison of distribution of students according to reached benchmarks in mathematics in national assessment (2003) in England and the international TIMSS (2003)

The number of levels greatly exceeds that of TIMSS. If we try to show both distributions on the same diagram, we can represent the lowest benchmark in TIMSS with national levels which vary from the expected level for the majority for two levels or more downwards, i.e. 'Does not reach Low

benchmark' = Levels 2 and 3, and similarly, the highest TIMSS benchmark with levels which exceed the expected level for the majority for two levels or more, i.e. Advanced = Levels 8 and E. The middles of both distributions overlap: Low benchmark with Level 4, Intermediate with the expected level for the majority, i.e. Levels 5 and 6, and Advanced with Level 7.
Overlapping of distributions in higher benchmarks of achievement is well-defined. Information obtained in England from annual national assessment matches data from the international TIMSS study. The highest and lowest extremes reveal 5% of students with exceptional achievement and up to 10% who fail to reach the set minimal goals.

3.5. The USA

The American system of national assessment NAEP (National Assessment of Educational Progress) has operated since 1969 (Greaney, Kellaghan, 1996). In NAEP, achievement is assessed after 4, 8 and 12 years of schooling. From 1984, reports cite the percentage of students whose achievement is at least at the basic level. Students' results are based on a three-level scale: Basic, Proficient and Advanced; or on a four-level scale, if we take into consideration the group who fails to display even basic knowledge and skills (Below Basic).
The percentage of students at a certain level of achievement in mathematics increases a percent or two from year to year; one third, however, fails to reach the basic level of achievement. For 2003, data for mathematics are as folllows (Percentage of students, by mathematics achievement level and gender, grade 8: 1990–2003 – NCES, 2004):

NAEP	Below **Basic**	At **Basic**	At **Proficient**	At **Advanced**
2003	32%	39%	23%	6%

Table 4: Percentage of students who in Grade 8 in NAEP in mathematics performed at a certain level of achievement (the USA)

Only a few percent of students perform at Advanced level, while the proportion of students who fail to perform at even Basic level is between 30 and 40 percent of Grade 8 students.
If we try to show the distribution on the same diagram as TIMSS, the lowest level in national assessment (i.e. Below Basic) can be divided into two parts on the TIMSS scale ('Does not reach Low benchmark' and Low benchmark). Levels of achievement at higher levels, however, would overlap: Basic with Intermediate, Proficient with High and Advanced with Advanced.
The represented overlapping of distributions is remarkable. Information obtained from national assessment in the USA matches very well the middle and upper parts of the data from the international TIMSS study. The definition of knowledge and skills labelled in NAEP as Below Basic or in TIMSS as Low benchmark and less, i.e. when students fail to reach even lower levels of achievement, however, remains a matter of interpretation. Basic and higher forms of achievement on the TIMSS scale, or Basic, Proficient and Advanced on the NAEP scale, are reached by almost the same percentage of students as measured by the international comparison of knowledge or national assessment.

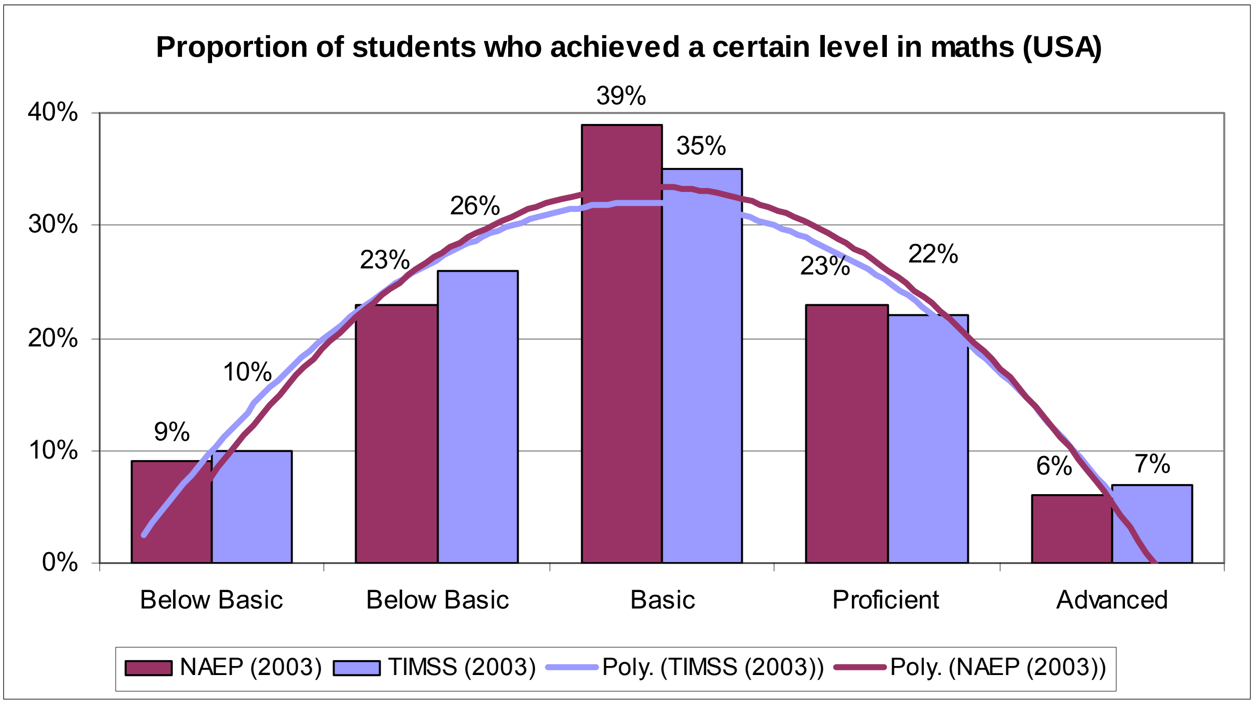

Diagram 3: Comparison of distribution of students according to reached benchmarks in mathematics in NAEP (2003) in the USA and the international TIMSS (2003)

3.6. Canada – Ontario and Quebec

In Canada each province or terrritory has its own system of education. The study deals with data from the two largest central provinces, Ontario with mainly an English-speaking population and Quebec with mainly a French-speaking population. Data from the TIMSS study are available only for these two provinces.

After having found out that students' achievements in school subjects are an important indicator for the operation of the school system, educational ministeries in provinces and territories (CMEC – Council of Ministers of Education, Canada) in 1989 decided to carry out national assessment, the SAIP (School Achievement Indicators Program) in reading, writing and mathematics (Canada – Council of Ministers of Education, 1996) for 13 and 16 year-old students. Achievement in individual subjects is shown on five levels representing a continuum of knowledge and skills reached by students in primary and secondary school. Criteria for Level 1 represent knowledge and skills that students normally reach in the lower grades of primary school, criteria for Level 5, however, are reached by the most able students at the end of their secondary education. Development groups prepare tests whereby the majority of 13 year-old students ought to reach Level 2 and the majority of 16 year-old students Level 3.

This is the achievement of 13 year-old students in the assessment of mathematics in 2001[i]:

[i] Data for schools with lessons in English and for schools with lessons in French are combined so that they are comparable with data from TIMSS 2003 in continuation.

SAIP – 13 years	Proportion of students according to reached levels of acheievement in mathematics – [in %]					
	Below Level 1	**Level 1**	**Level 2**	**Level 3**	**Level 4**	**Level 5**
Ontario	11.8%	27.5%	35.6%	23.7%	1.3%	0.1%
Quebec	11.0%	18.2%	33.6%	35.8%	1.3%	0.0%

Table 5: Percentage of 13 year-old Canadian students who in SAIP in mathematics in 2001 performed at a certain level of achievement (SAIP Mathematics III Assessment 2001: The Public Report, 2002: 119)

The expected level is reached by one third of students. A significant proportion (between 24% and 36%) reaches a higher level, Level 3, and slightly over 1% of students even Levels 4 or 5. Another characteristic is that in Canada too it is reported that at 13 years of age, one fourth or one fifth of students (18% to 27%) show merely elementary achievement in mathematics in the lower grades of primary school, i.e. Level 1, and that there are more than 10% of students who fail to achieve even that.

The expected level for the majority of students at this age is Level 2; this is why in representation on a common diagram this level overlaps with Intermediate benchmark in TIMSS. Two lower levels ('Does not reach Low benchmark' = Below Level 1 and Low Benchmark = Level 1) and two higher levels (High benchmark = Level 4 and Advanced benchmark = Level 5) adequately overlap. In addition, the highest level, Level 5, which in SAIP is reserved for better students at the end of secondary education, is joined to Level 4 and thus the highest benchmark in TIMSS.

Overlapping of distribution in neither of the provinces is ideal, but it is nevertheless evident that the information obtained in Ontario and Quebec in Canada in their national assessment does not significantly vary from the international comparison TIMSS.

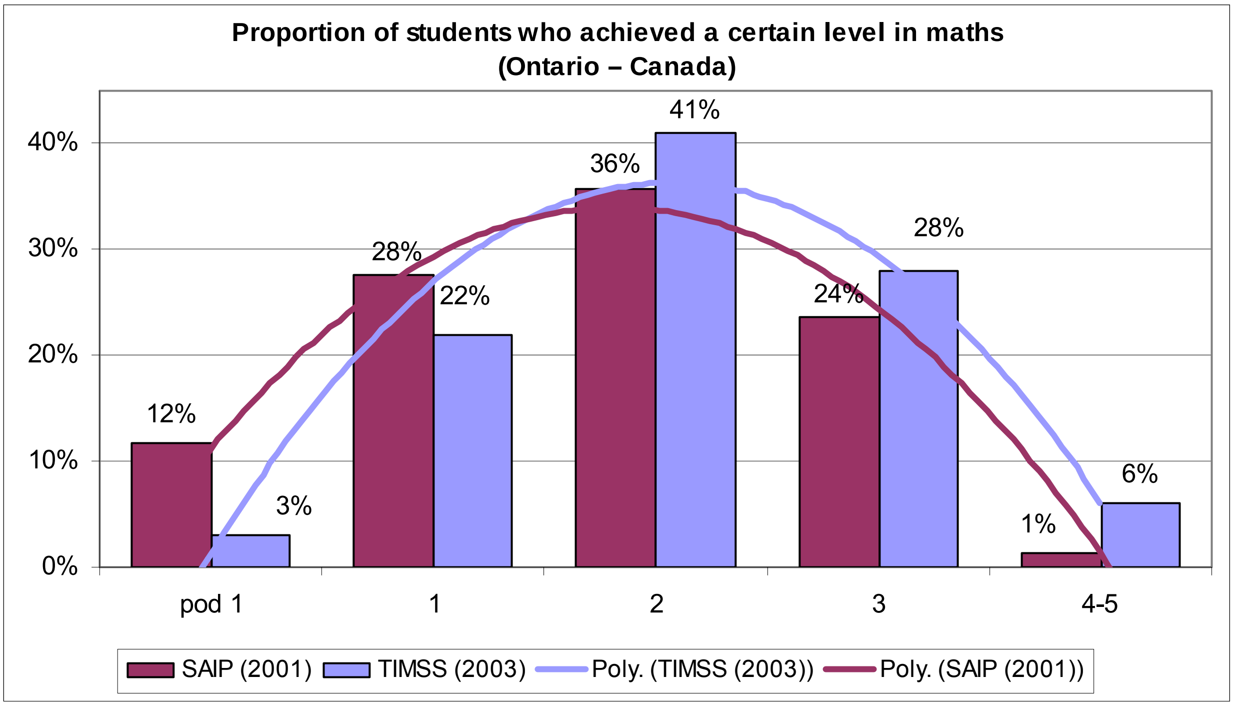

Diagram 4: Comparison of distribution of students according to reached benchmarks in mathematics in SAIP (2001) in Ontario (Canada) and the international TIMSS (2003)

With regard to the international comparison TIMSS, distributions from national assessment are in both provinces slightly out of line. The differences are defined at both ends of the scale where data from national assessment are 'stricter' as they indicate that over 10% of 13 year-old students fail to perform at the lowest level of achievement, while TIMSS measures only 1% of such students in Quebec and 3% in Ontario.

Similarly 'strict' is the situation with the highest achievement whereby national assessment measures only 1% of students who, for two levels or more, exceed the expected achievement for this age, while TIMSS measures 6% of students who have reached the highest achievement benchmark in Ontario and 8% in Quebec.

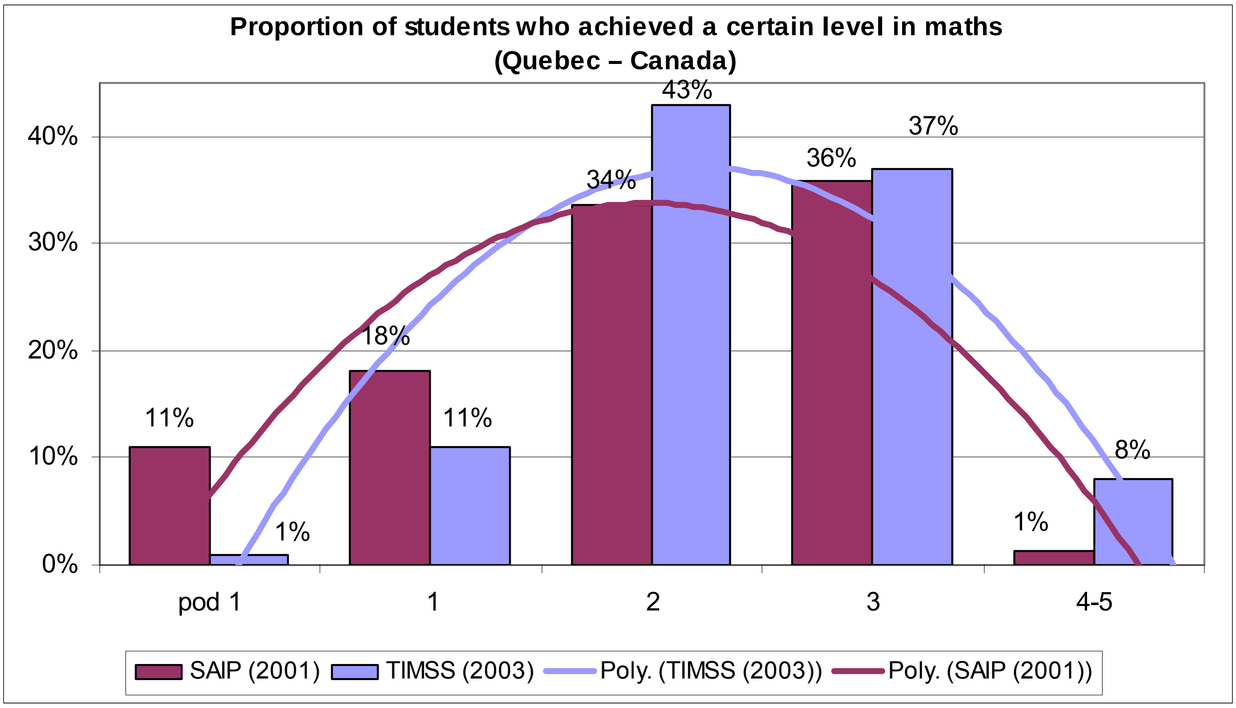

Diagram 5: Comparison of distribution of students according to reached benchmarks in mathematics in SAIP (2001) in Quebec (Canada) and the international TIMSS (2003)

3.7. Australia

Australia is divided into six States and two Territories, and each has jurisdiction over its educational system. The Australian Capital Territory (ACT) is one of the most developed States or Territories (Literacy and Numeracy, 2000) with the highest percentage of students reaching the national achievement level (Edwards, 2002: 115). Students' achievement is measured in comparison to the National Profile Level. In all schools in the ACT, students' achievement is described for each key area of learning using 8 Profiles, from kindergarten to the 12th year of schooling (Literacy and Numeracy, 2000). Levels of achievement and profiles are calibrated in such a way as to enable students (mainly in the early years of schooling) to progress from one level of achievement to the next every 18 to 24 months. Profile levels are set with the expectation that the majority of students will reach Profiles 5 or 6 in their 9th year of schooling (at 14 years of age). Outstanding students can reach Profile 8 by the end of their 10th year of schooling.

ACTAP	Profile 4	Profile 5	Profile 6	Profile 7
Mathematical literacy	11.2%	43.0%	38.6%	7.7%

Table 6: Percentage of students who in 9 years of schooling in Canberra and its surrounding area reached a certain level of achievement in ACTAP in mathematical literacy in 1999

The national report for mathematics achievement in Australia (National Report on Schooling in Australia, 2003: 27) revealed in 2003 that after seven years of schooling 81.3% of students reach the set minimal benchmark of achievement. In 2003, 18.7% of students of this age failed to reach basic benchmarks of achievement in mathematics; these students would have problems in further schooling. Differences between States and Territories in Australia are big; the best are students in the ACT with 86.4% achievement, which means that in seven years of schooling 13.6% of children fail to reach the minimal desired standard in mathematics at the end of their primary education.
In comparing distribution achievement in mathematics in ACTAP and TIMSS, students who reach only Profile 4 compare to those who in TIMSS do not reach Low benchmark. The best students, however, those that reach Profile 7, are compared to the group who in TIMSS reaches Advanced benchmark. The two middle profile levels on the ACTAP scale (Profiles 5 and 6) are proportionately distributed between the three middle benchmarks on the five-level TIMSS scale (i.e. Low, Intermediate and High benchmarks).

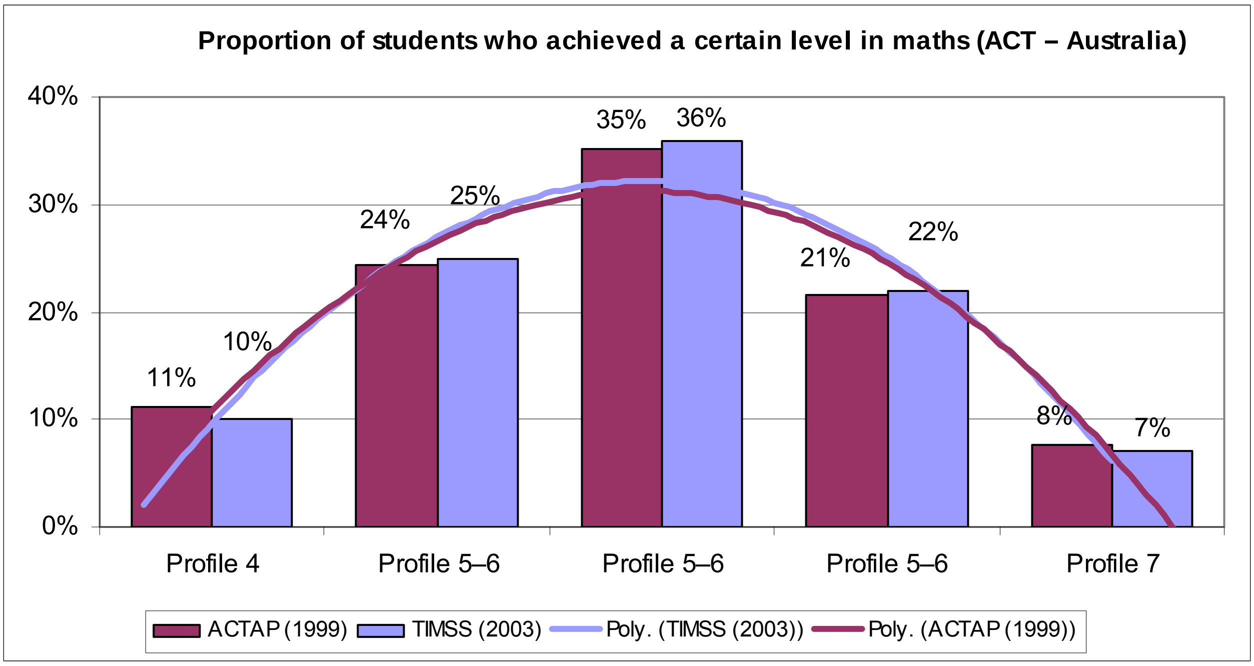

Diagram 6: Comparison of distribution of students according to reached benchmarks in mathematics in ACTAP (1999) in the Capital Territory in Australia and the international TIMSS (2003)

The diagram reveals a remarkable overlapping of distributions. One of the reasons is that the middle proportion of students, who in ACTAP reach Profile 5 (43%) and Profile 6 (39%), is proportionaly distributed over three categories of TIMSS. Matching of distributions is, however, more than obvious at both ends of the scale. National tests measure that 11% of students at Profile

4 lag behind the expected mathematics achievement for 14 year olds; TIMSS, on the other hand, measures 10% of students failing to reach Low benchmark. There is only one percent of difference between the best of students: 8% of students at Profile 7 significantly exceed the expected mathematics achievement for 14 year olds; TIMSS, on the other hand, measures 7% of such students at Advanced benchmark.

3.8. Slovenia

According to the Elementary School Act (Zakon o osnovni šoli [Elementary School Act], 1996 and 2001), each of the three three-year cycles in Slovene 9 year primary schools should be completed by national assessment (NA). At the end of Year 9, the compulsory NA for students would include tests in the Slovene language, mathematics and one other selected subject. The scale of grades would have five levels, from Insufficient (1) to Excellent (5), where grade 1 is negative and other grades are positive. A characteristic of NA in Slovenia was this legislative provision: in giving the final grade in Year 9, the teacher's grade and the grade from the NA are combined according to a special principle. This gave the otherwise declarative 'national assessment' a distinctive element of high-stake 'public examination' becoming very competitive for students, teachers and schools. Should a student get a negative grade in the NA,[ii] (s)he would fail to complete primary school. From 2002 onwards, when the first students in pilot schools completed 9 years of primary school, to 2005, national tests had the status of a trial[iii].

Mathematics	Proportion of students according to reached grades – [in %]				
	Insufficient (1)	**Sufficient (2)**	**Good (3)**	**Very good (4)**	**Excellent (5)**
NA (2003)	2.7	14.7	29.5	24.2	28.9

Table 7: Proportions of students in 9 year primary school who in mathematics reached a certain grade in NA in Slovenia (Štraus, 2003: 40)

Minimal, core and advanced standards of knowledge, as written in course syllabi, were a matter of discussion among experts (Zupanc, 2005). Subject experts who prepared tests for NA dealt with these issues intensely (Slavec Gornik, 2002: 75) as they faced problems after tests had been taken to set benchmarks: from Insufficient (1) to Excellent (5). According to a criterion-referecned judgement, national tests in individual subjects were prepared in such a way that more than one half of tasks would assess the so-called minimal standards that every student should reach. With regard to the written standards of knowledge in course syllabi and the content, number and selection of tasks, the positive benchmark (Sufficient (2)) in a test ought to be not far from 50% of the total score of test points; in mathematics, in none of the years was this the case. Positive benchmarks dropped as low as 20% of the total score of points. The committee who had the authority of deciding on this, adopted the criterion (Slavec Gornik, 2002: 50–51) to allow for the

[ii] The Act anticipated an exception, however: should a student in one or two subjects fail to obtain a positive grade, but had good grades in other subjects, (s)he would get the so-called 'near pass' grade with the otherwise negatively assessed subject and would complete his/her primary school education.
[iii] Combining the teacher's grade and the grade from NA was not compulsory in the transitional period. If a student's final grade should go down because of his/her achievement in NA, this principle would be ignored.

lowering of the benchmark even well below the absolute criterion, should more than 2 to 3% of students be unsuccessful. They explained their decision by stating that the percent of students with a negative grade at the end of primary school with only the teachers' internal assessment was approximately the same. With such setting of benchmarks, in the last years ca. 3% of students in Year 9 have obtained negative grades; on the other hand, more than one half of students were assessed with more than good: Very good (4) and Excellent (5) – more than one quarter of them achieved outstanding results: they got an excellent grade (5).

Even at first sight, Slovene national tests show a significant deviation when compared with data from other countries: it is as if their distribution is reversed. In other countries, 10% or more of students fail to reach Low benchmark of achievement in mathematics, while on the other hand only a few percent of the best students reach above average and exceptional achievement. In Slovenia, it is the other way around: insufficient achievement in national assessment is reported only in a few percent, while above average grade is reached by more than one half of students, with over one fourth with the highest grade.

Grades in national assessment in mathematics in Slovenia can be represented on a traditional five-level scale, i.e. a scale with the same number of levels as the TIMSS scale. For comparative distributions TIMSS data for students in Year 8 of the new 9 year primary school have been used[iv].

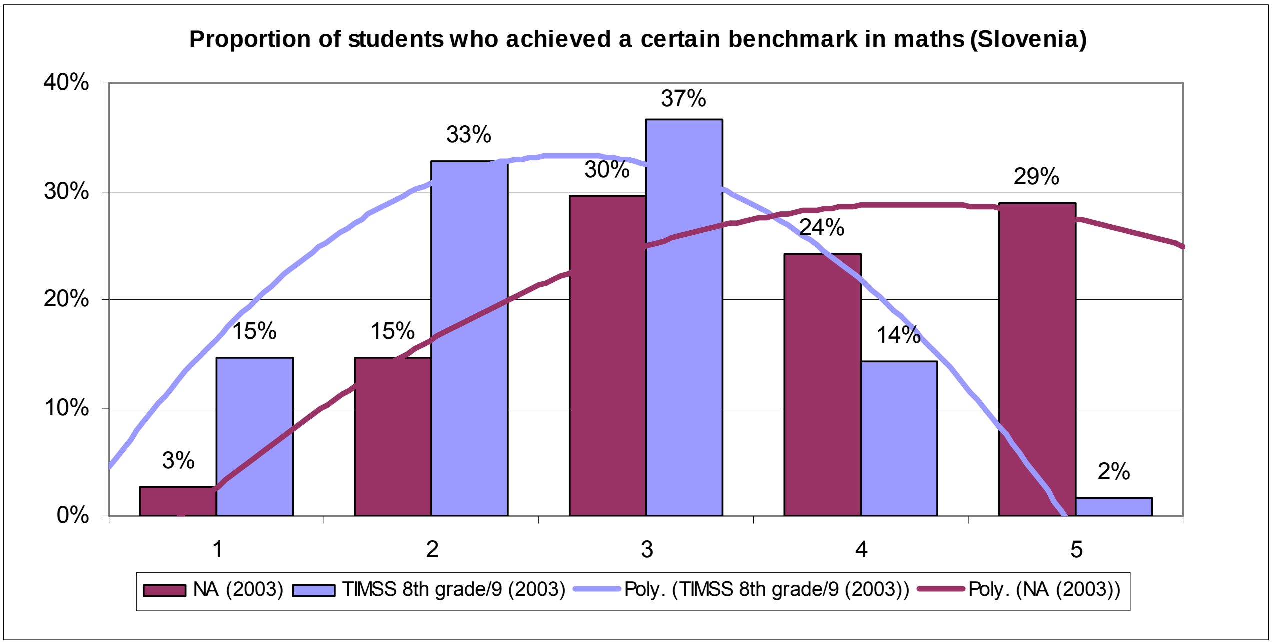

Diagram 7: Comparison of distribution of students in Year 8 of the 9 year primary school according to reached benchmarks in mathematics in NA in Slovenia (2003) and the international TIMSS (2003)

In none of the countries where distributions of levels of achievement in national assessment and in international TIMSS have been compared, such a shift has been established. TIMSS results

[iv] Students of Year 7 of 8 year primary school were not included in NA, and this is why a comparison of distribution of results of students of Year 8 of the new 9 year primary school (although their number was smaller) who were included in TIMSS 2003 as well as national assessment in Slovenia, is represented. Even if data for students in the old programmes were taken into consideration, both distributions would not differ significantly.

themselves, where a well-marked proportion of students fail to reach Low benchmark of achievement (among students in Year 8 of 9 year primary school there are 15%) and very few (2%) reach advanced benchmark, have sparked off debates among experts in Slovenia. If we compare this with the information that students, parents, teachers and other experts get with national assessments at the end of their schooling,[v] the difference has a strong sobering effect. Not only are proportions of students in national assessment who reach Low benchmark of achievement (15%), or even fail to reach it (3%), twice or even five times lower than measured by international assessment TIMSS, but above average grade, i.e. more than Good (3), is obtained by more than one half of the population (53% = 24% + 29%), and even here there are more 'excellent' than 'very good' students. The curve does not have a correct tail at all, as TIMSS measures that at approximately this age there are only a few percent of outstanding students in mathematics in Slovenia (2%).

In Slovenia, a debate opened in the area of education as to whether this distinctly embellished picture on mathematics achievement is a characteristic of other subjects as well. One of the important reasons for such an unusual distribution of grades in national assessment is the legislative provision[vi] of combining the teacher's grade and the grade in national assessment, where a positive grade in national assessment is a requirement for the successful completion of a primary education. If results of national assessment had to be used for that purpose as well, i.e. a successful completion of primary education – then those who made decisions were in an unpleasant situation with regard to ensuring standards of achievement. Lowering of criteria for the lowest benchmarks of achievement probably had as its consequence the lowering of criteria for higher benchmarks as well. This is likely what happened in Slovenia and is an important reason for the abnormal distribution of students on benchmarks of achievement in national assessment.

4. Discussion

School policies aim to improve the educational structure of the population: to provide high-quality primary education for everyone, to enable the majority to complete secondary education, to enrol a high proportion into tertiary education etc. As a developing country Slovenia has endeavoured to catch up to and join the developed countries, and education is a pointer leading towards the desired goal.

Public examinations and national assessment are one of the controlling tools of educational systems (Inbar, 1986: 272), and whenever instruments, or tools, are used in interim systems, experts first have to understand how they function, what their effects are, and how to manipulate them. For example: in order to observe a certain phenomenon or process, mirrors or lenses can be used in everyday life as well as in science. With mirrors and lenses one can get very different pictures of the real world: real or virtual, diminished or enlarged, even inverted; it all depends on

[v] Distribution of teachers' grades in schools at the end of schooling is similar to distribution of grades from NA as well.
[vi] In 2005 the Elementary School Act on this point changed and combining the teacher's grade and achievement in NA has been revoked. The demand that a Pass grade in NA is a requirement for a successful completion of primary education has been shelved.

the purpose and proper use of the tool in question. The use of mirrors and lenses can even be dangerous; one can cut one's fingers with the sharp pieces of the broken tool. A similar situation can occur in any school system with newly introduced external assessment. Here too, tools can be used to very different purposes: when the purpose is clear, expert application and interpretation of results can reveal important details, and they zoom and focus on the true picture; with improper application, however, ignorance or misuse can blur or even distort the picture.

Slovenia has one of the highest percentages of population aged between 20 and 24 in Europe which has completed at least upper secondary school (Tessaring, Wannan, 2004: 35): 85.8% in 1999, and 90.7% in 2003. All developed countries lag behind: the European average is 76.7%. In the European Union, 15.6% of all pupils and students in the education systems are enrolled in tertiary education on average. According to this indicator, Slovenia is in second or third place with 21.6% (Key Data on Education in Europe 2005: 146), i.e. in front of much more developed countries. One ought to enthuse over these very satisfactory Slovene numbers, but one nevertheless cannot help but doubt that the exceptionally high proportion of persons with certificates and the extremely low drop-out rate reveal the inflation of grades, certificates, degrees and qualifications. Are these persons employable, can they compete internationally on an equal footing in further education? If the system is easily surmountable, does it discriminate adequately and separate the good from the bad, the exceptional from the average?

In Slovenia, the deep desire of school policy to improve the educational structure of the population, as revealed by numerical indicators, probably quite unwillingly exploited national assessment as well as the use of its results. Most likely because of the pressure applied by opponents of the newly introduced external assessment as well as the public, among other things, a criterion-referenced interpretation of results among experts was not made and 'strict' standards were not adhered to; criteria were lowered and high grades awarded instead – with inflationory effect.

Other countries too keep asking themselves about the keeping and lowering of standards. In the United Kingdom, the keeping of standards from year to year, in the absolute sense, is strongly emphasised (Wolf, 2000: 28). In the Netherlands in the 1990ies doubt was expressed in the Parliament about the growing number of secondary-school students wishing to continue their studies as it was felt the annual benchmarking for the continuation of studies had the effect of lowering standards of achievement (Alberts, 2001: 353). Another well-known case is when the American public learned from the report (*National Commission on Excellence in Education*, 1983) in 1983 that the entire nation was at risk (Cizek, 2001). Not coincidentally, research on teachers' grading practices revealed the phenomenon of what came to be referred to as 'grade inflation'. Early studies showed that an overall student's grade point average (GPA) did not reveal a simultaneous growth of real achievement measured by the common yardsticks of the time. Cannell (1987) attracted the public's attention by what is now known as the Lake Wobegon effect (Koretz, 1988) when he found out that all states in the USA, as well as the majority of districts, reported that their students achieved above average results compared to the national norm. Based on their data charts he reached the conclusion that students, parents, the school system, legislators and the mass media had been shown inflated and misleading reports on achieved benchmarks in standardised and national norm-referenced tests.

Perhaps this was caused by a widespread misunderstanding about the relationship between positive self-esteem and achievement. It was a teachers' desire (Cizek, 2001: 7) that their students succeed and all to have the personal esteem associated with those accomplishments. Unfortunately, the awarding of higher grades to heighten self-esteem and stimulate accomplishment too often had neither effect.

Welle-Strand and Tjeldvoll (The Situation in Primary and Secondary Education in Norway, 2003: 22) cite a more radical approach for Norwegian schools whose achievement in the international comparison of TIMSS 2003 was poor. They claim that their populist educational policy in the second half of the 20th century is the guilty party for the falling quality of the Norwegian educational system. Typical of such a school policy is the anti-intellectual approach where practical skills and local knowledge become more important than formal knowledge and core skills. Learning through play and practical work becomes more important for students than a systematic building up of knowledge through set goals. The system is directed towards not producing 'losses' and leads, among other things, to neglecting big differences, ignoring competitiveness in favour of average, and as a result, diverts attention away from talented students. In the authors' opinon, all this contributes to the lowering of students' level of achievement.

Experts' and practitioners' opinions in Slovenia also voice the thought that exaggerated positive self-esteem lacks proper foundation and school principals observe (Erčul, Lorenčič, 2005: 61) a lack of endeavour for 'I want to know'. They also note that the goals of education are mostly set too low.

5. Bibliography

Alberts, R. V. J. (2001). Equating Exams as a Prerequisite for Maintaining Standards: experience with Dutch centralised secondary examinations. *Assessment in Education: principles, policy & practice,* 8(3), pp. 353–368.

Canada – Council of Ministers of Education, (1996). *School Achievement Indicators Program. Science Assessment. Framework and Criteria.* Toronto, Ontario.

Cannell, J. J. (1987). *Nationally normed elementary achievement testing in America's public schools: Hoe all 50 states are above the national average* (2nd ed.). In: Daniels, W. V.: *Friends of Education*, p. 3.

Cizek, G. J. (2001). Conjectures on the Rise and Call of Standards Setting: An Introduction to Context and Practice. In: Cizek, G. J. (edit.). *Setting Performance Standards – Concepts, Methods, and Perspectives.* Lawrence Erlbaum Associates, Publishers, pp. 3–18.

Crighton, J. (1993). *Poland – National Standards in Education: the creation of a national assessment and examination service – a proposal*. World Bank. Washington DC, p. 9.

Devetletna osnovna šola: nacionalni preizkusi znanja – informacije za učence in starše [Nine-year primary school: national assessment – Information for pupils and parents]. (2001). Državna komisija za vodenje nacionalnih preizkusov znanja v 9-letni osnovni šoli. Državni izpitni center. Ljubljana, pp. 1–16.

Edwards, R. W. (2002). *Australian Social Trends 2002*. Australian Bureau of Statistics. Commonwealth of Australia. Canberra, pp. 1–226.

Erčul, J., Lorenčič, I. (2005). Povzetek razprav ravnateljev o TIMSS-u [Headmasters discussion summary about TIMSS]. *Vodenje v vzgoji in izobraževanju*, 3(1), Brdo pri Kranju, pp. 59–64.

Gabršček, S. (2001). Access to Higher Education in Slovenia: assessment reform and re-creation of a national examination. *Assessment in Education: principles, policy & practice*, Vol. 8, No. 3, pp. 369–390.

Greaney, V., Kellaghan, T. (1996). *Monitoring the Learning Outcomes of Education Systems. Directions in development.* The World Bank. Washington, DC, pp. 1–88.

Guide for Foreign Researchers (2004). http://www.lu.se/info/Guide_Foreign/A15.Guide.swedu.html

Inbar, D. E. (1986). Educational policy-making and planning in a small centralized democracy. *Comparative Education*, 22(3), pp. 271–281.

Japelj Pavešić, B., Brečko, B. N., Čuček, M., Vidmar M. (2004). *TIMSS 2003 – Mednarodna raziskava trendov znanja matematike in naravoslovja – Povzetek izsledkov* [TIMSS 2003 – Students achievement in Mathematics and Science – International Trends]. Pedagoški inštitut. http://ceps.pef.uni-lj.si/timss-long.pdf

Kellaghan, T., Grisay, A. (1995). *International Comparisons of Student Achievement: Problems and Prospects. Measuring What Students Learn.* Organization for Economic Cooperation and Development – OECD. Paris.

Key Data on Education in Europe 2005. (2005). European Commission, EURYDICE, Eurostat, pp. 1–364.

Koretz, D. (1988). Arriving in Lake Wobegon: Are standardized tests exaggerating achievement and distorting instruction? *American Educator*, 12(2), pp. 8–15.

Literacy and Numeracy (2000). ACT – Australian Capital Territory Government 2000. Education & Community Services, pp. 1–206. http://www.decs.act.gov.au/publicat/pdf/frontsec%20final.pdf

Mullis, I.V.S., Martin, M.O., Gonzales, E.J., Chrostowski, S.J. (2004). *TIMSS 2003 International Mathematics Report*. International Association for the Evaluation of Educational Achievement – IEA. TIMSS & PIRLS International Study Center, pp. 1–437.

National Curriculum Assessments of 7, 11 and 14 year olds in England, 2003 (Provisional). (2003). http://www.dfes.gov.uk/rsgateway/DB/SFR/s000407/index.shtml

National Report on Schooling in Australia (2003). National Benchmark Results Reading, Writing and Numeracy Years 3, 5 and 7 – Preliminary Paper 2003. Ministerial Council on Education, Employment Training and Youth Affairs, pp. 1–40.

Percentage of students, by mathematics achievement level and gender, grade 8: 1990–2003. (2004). NCES – National Center for Education Statistics. http://nces.ed.gov/nationsreportcard/mathematics/results2003/natachieve-gender-g8.asp

SAIP Mathematics III Assessment 2001: The Public Report. (2002). Council of Ministers of Education, Canada (CMEC), pp. 1–137.

The Situation in Primary and Secondary Education in Norway – Report. (2003). Ministry of Education and Research, pp. 1–54.

Slavec Gornik, A. (Ed.) (2002). *Nacionalni preizkusi znanja – Letno poročilo o izvedbi v šolskem letu 2001/2002* [National assessment – annual report 2001/2002]. Državni izpitni center. Ljubljana, pp. 1–102

Štraus, M. (ur.) (2003). *Nacionalni preizkusi znanja – Letno poročilo o izvedbi v šolskem letu 2002/2003* [National assessment – annual report 2002/2003]. Državni izpitni center. Ljubljana, pp. 1–97
Tessaring, M., Wannan, J. (2004). *Vocational education and training – key to the future*. Lisbon-Copenhagen-Maastricht: mobilising for 2010. Cedefop synthesis of the Maastricht Study. European Centre for the Development of Vocational Training, pp. 1–67

UNESCO. (1990). *World Declaration on Education for All*. World Conference on Education for All. Meeting Basic Learning Needs. New York, p. 9.
http://www.cies.ws/PaperDocuments/PDF/WorldDeclarationonEducationForAll.pdf

West, R., Crighton, J. (1999). Examination Reform in Central and Eastern Europe: issues and trends. *Assessment in Education,* 6(2), pp. 271–289.

Wolf, A. (2000). A Comparative Perspective on Educational Standards. V: Golstein, H., Heath, A. (Eds.). *Educational Standards. Proceedings of The British Academy – 102*. The British Academy, pp. 9–30.

Zakon o osnovni šoli [Elementary School Act]. (1996, 2001). Uradni list RS, 12/1996 and Uradni list RS, 59/2001.
http://www.uradni-list.si/1/objava.jsp?urlid=199612&stevilka=570
http://www.uradni-list.si/1/objava.jsp?urlid=200159&stevilka=3147

Zupanc, D. (2005a). Standardi znanja v slovenski šoli [Knowledge Standards in Slovene Schools]. *Sodobna pedagogika*, 56(1), pp. 124–147.

Zupanc, D. (2005b). Standardi znanja v edukaciji [Standards in Eduacation]. *Psihološka obzorja* [Horizonts of Psychology], 14(3), pp. 124–147.