FIELD OF STUDY AS A FACTOR INFLUENCING THE MODEL OF VALUE-ADDED ASSESSMENT

Radek Krpec, Michal Burda

University of Ostrava
radek.krpec@osu.cz

ARTICLE INFO

Article type
Full research paper

Article history
Received: May 15, 2011
Received in revised form: June 28, 2011
Accepted: June 29, 2011
Available on-line: June 30, 2011

Abstract

This contribution deals with the possibilities of schools' results evaluation and unbiased assessment of the so called education value-added. Value-added models in education express school contribution to the progress of a pupil in relation to predetermined educational goals. The article is a comparison of two methods of the value-added assessment: method of relative shift and relative gain of knowledge method. The focus is laid on the school's field of study as a factor which could, to a considerable extent, affect the measurement results. Both methods are used for relatively wide range of data drawn from results of secondary school pupils value-added assessments and are compared in respect to the schools' classification according to their field of study. The results show that the field of study is a significant factor influencing the value-added assessment outcomes and have to be taken into account.

Key Words

value-added model, relative gain of knowledge, on-line educational testing, socio-economic facor

Introduction

Value-added assessment and relative gain of knowledge

In present times of changes in the educational system and regarding the pressure put on elementary and secondary school students' progress in the level of education, the necessity of schools' results evaluation is relevant.

One of the possible measurement forms is the value-added assessment based on finding out what part a certain school plays in pupil's progress (Lissitz, 2005; Malach, Malcik, 2010). Here, results from two different time segments at given school are compared. In this respect, test results from separate key phases of the education process may serve to determine the value-added level. Although it is clear that even the value-added assessment does not take into account the whole spectrum of factors affecting pupils' outcomes, and thus it will not solve the problems in measuring a particular contribution of the given school to the pupil's progress, it is still notable improvement against using mere test results.

The OECD definition by Educational policy institute (2008): "Value-added models in education express school contribution to the progress of a pupil in relation to predetermined educational goals. Contribution is a value rid of the other factors instrumental to progress in pupil's education." The value-added assessment models could be divided in two basic groups:

- Simple – assessment is realized in two different time segments
- Contextual – takes into account also factors not influenced by the school

The value-added score of a school is affected by information of contextual character on three levels.

1. Students enrolling at school have been prepared variously in the tested subjects along with other contextual characteristics as socioeconomic status, Income Deprivation Affecting Children Index (IDACI), special educational needs, and so on.
2. The information on their contextual situation is presented during the whole period of school attendance in terms of improvement possibilities.
3. Schools have their educational programmes built up variously regarding the study plans and curricula.

Since the only contextual information available was the field of study, we could not have applied the multiple regression method for the relative shift calculation. One of the possibilities then was to calculate the relative shift, in accordance with so called fields, when each secondary school class partaking in these tests was assigned one of the nine fields of study according to a unified dial. Separate fields were then assigned to "similar" schools with similar educational programme frames and their value-added score is to a certain extent comparable. We have distinguished the following fields: 1. Grammar schools, 2. Lyceum schools, 3. Technical schools, 4. Scientific schools, 5. Economic schools, 6. Services, 7. Pedagogical, social and health-care oriented schools, 8. Social science oriented schools, and 9. Art schools.

Main goal of the article is to compare how the results differ in the sequence of schools with regard to the both methods mentioned – "Total relative shift" and "Total relative gain". Fields of study as a socio-economic factor and it's impact to the value-added school results is also investigated.

Material and Methods

Relative gain of knowledge

One of the models used in practice is the model of relative gain of knowledge (Malčík and Krpec, 2010). For obtaining the student's value-added score, we need to know to what extent the student's outcomes worsened or improved compared with possible presuppositions. By "possible presuppositions" are meant results coincident with results of similar students from different schools. The similarity of students should be considered from the viewpoint of previous results as the best presupposition for the results in future (Malčík, 2007).

For calculation of the relative gain of knowledge we use a linear regression model based on pupil's knowledge measurement in two different time segments between input and output results, see e.g. (McCaffrey, Lockwood, Koretz and Louis, 2004; Liu, 2011; Sanders and Horn, 1994). Figure 1 illustrates the process of delimiting the value-added score in two subjects. The horizontal axis demonstrates input results; output results are illustrated by the vertical axis. The field with pupils' results data is represented by regressive line which is, after subjects and fields of study, calculated using the equation:

$$y_{ij(2)} = a_0 + a_1 y_{ij(1)} + \varepsilon_{ij},$$

while

- i – a label of pupil in terms of j-th school,
- $y_{ij(2)}$ – final test result,
- $y_{ij(1)}$ – previous testing results,
- a_0, a_1 – regressive coefficient,
- ε_{ij} – accidental error normally divided, independent and with identical variance for each student.

The regressive line roughly interprets average outcomes of students who were placed at a certain point of the input information axis by their previous results.

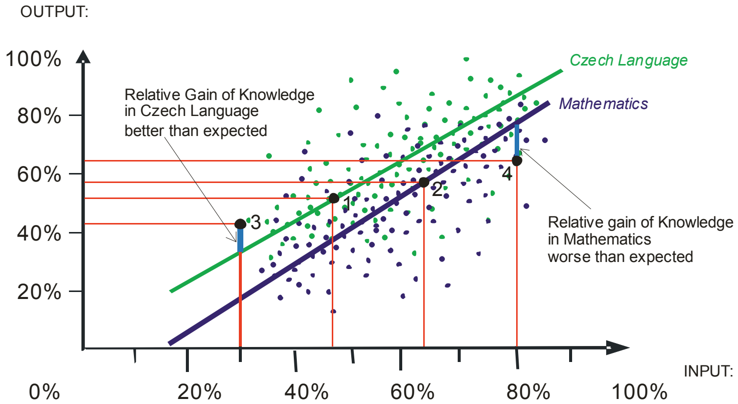


Figure 1 – Method of calculating the Relative gain of knowledge

The Figure 1 shows that student **2**, entering with a success rate of 64%, will probably have an output result reaching 54% in mathematics. This rate is the presumed success rate of the student.

Provided the student will reach better than the presumed results – and in fact half the students always reach better than the presumed results – the student has a "positive residue". Residue is defined as difference between the actually reached success rate and the success rate presumed on the basis of regression line. If the student obtains worse mark than has been presumed, then he or she has a "negative residue", as student **4** in Figure1 has in mathematics. Residues are often referred to as benchmarks of the value-added score. Yet, it will be certainly more accurate if we refer to them as a *relative* value-added. Some of the students reached higher value-added score than others, as suggested by their residues.

Relative shift

Another way to measure the level of education among pupils is to evaluate the so called Relative shift of a pupil (Vector Module 3). It is a ratio expressing the extent of progress or downgrade in a certain test. It is a rate of progress (downgrade) of a pupil against his or her maximum possible progress (downgrade).

Calculation of relative shift for progress:

$$\frac{output\ percentile - input\ percentile}{100 - input\ percentile}\%.$$

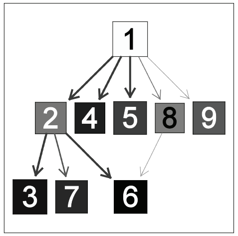
Calculation of relative shift for downgrade:

$$\frac{output\ percentile - input\ percentile}{input\ percentile}\%.$$

Percentile may be interpreted as a ratio of those outperformed by the participant (Chráska, 2007, pp. 202-204). The input percentile is then the one reached by pupil in the input test, the output percentile the one reached in the output test.

Hasse diagrams

The numbered hubs represent individual fields of study, while the direction of arrows signals statistically significant divergence between the given subjects (i.e. rejection of the zero hypothesis of congruity of mean values). The arrows are directed from the fields with higher mean value to those with lower mean value. Hubs not connected with an arrow are incommensurable (i.e. the zero hypothesis was not rejected). Boldness of an arrow represents an extent of significance to which the zero hypothesis was rejected (thin line for the level of 0.05, then 0.01, 0.005 and the thickest line for the level of 0.001). Colour (shades of grey) of the hub also represents relative information on the mean value of the given subject and the hub size represents variance of values.



Description of a tested set

For the comparison we have made use of a set of results from testing which excluded pupils who had results of input or output test between 2007 and 2010 years from some of the three subjects either missing or the result was zero. This set comprised 4,714 pupils of the first year from 83 secondary schools in the Moravian-Silesian region who were examined in three subjects: Czech language (CL), mathematics (MA), English language (EL). The original tests also included the German language, but due to low participation of pupils from various fields of study it was decided only for the learners of English. Pupils' division into separate fields of study was as follows:

Field	No. of pupils
1	1,603
2	485
3	1,205
4	103
5	365
6	432
7	404
8	76
9	41
total	4,714

Table 1: Number of pupils divided into separate fields of study

Results

Comparison of outcomes of "Total relative gain" and "Total relative shift"

Firstly, we will have a look at a comparison of the averages of "Total relative gain" and "Total relative shift" in each individual field. We will see a list of individual fields ordered by the "Total relative gain" and "Total relative shift" in individual subjects.

The order of Czech language is identical in the first three positions, differences occur between the fourth and fifth position and between the sixth and seventh position. Since the differences in "Total relative shift" averages between the fourth and fifth position and also in "Total relative shift" between the sixth and seventh position are not statistically significant, the reversed order could have been coincidental. We may presume that the order shows no significant variances.

In the rest of subjects (mathematics, English language) the sequence of fields does not vary at all. Thus it can be presumed that there is no cardinal difference between both methods.

Field	Average "Total relative shift" in CL
1	0.10
8	0.01
2	-0.01
5	-0.07
9	-0.08
4	-0.09
7	-0.09
3	-0.12
6	-0.14

Field	Average "Total relative gain" in CL
1	4.23
8	0.33
2	0.08
9	-0.84
5	-1.64
7	-2.29
4	-2.62
3	-2.89
6	-3.53

Field	**Average "Total relative shift" in MA**
1	0.13
2	0.01
3	-0.05
5	-0.06
8	-0.08
4	-0.14
6	-0.18
7	-0.24
9	-0.31

Field	**Average "Total relative gain" in MA**
1	7.31
2	0.81
3	-1.58
5	-2.80
8	-5.27
4	-6.84
6	-8.06
7	-10.09
9	-12.62

Field	Average "Total relative shift" in EL
1	0.06
9	0.05
5	0.02
2	-0.01
3	-0.07
7	-0.10
8	-0.10
6	-0.13
4	-0.22

Field	**Average "Total relative gain" in EL**
1	3.12
9	1.97
5	0.22
2	-0.52
3	-1.73
7	-2.16
8	-2.47
6	-2.95
4	-4.78

Table 2 to 7: Total relative shift and total relative gain in Czech language, mathematics and English language

Let us see now how the results differ in the sequence of schools with regard to the both methods mentioned – "Total relative shift" and "Total relative gain". "Total relative shift" or "Total relative gain" of a given school is delimited as an average value of "Total relative shift" or "Total relative gain" of all pupils from the relevant school (Raudenbush and Willms, 1995). If we sequence individual schools in accordance with "Total relative shift" and "Total relative gain", we will find out that the differences are more remarkable. We will rank the first ten schools by the average of "Total relative gain" and associate them with general ranking according to the average of "Total relative shift" in tests from Czech language. We will proceed identically with the last ten schools according to "Total relative gain" in Czech language.

School	**Ranking by "Total relative gain" in CL**	**Ranking in "Total relative shift" in CL**
X	1	5
XXV	2	17
XXII	3	3
LXI	4	11
LXXX	5	41
LX	6	1
XII	7	10
III	8	9
XL	9	13
XLIX	10	12
.	.	.
.	.	.
.	.	.
XIX	74	26
XXX	75	81
XLIII	76	51
LXXV	77	78
LV	78	54
XV	79	60
XXXIII	80	71
VIII	81	56
L	82	61
XVII	83	82

Table 8: Variances in Total relative gain and Total relative shift

The table proves large variances. For example, school LXXX is, regarding the average of "Total relative gain", on the fifth place, while in terms of "Total relative shift" it is as far as on the 41st place. Similarly, school XIX is with its average of "Total relative gain" back on 74th position, its average of "Total relative shift" is on 26th position. If we determine the difference between the averages of "Total relative gain" and "Total relative shift", we will find out that the most significant gap is 60 positions in the case of school on the 23th position in its average of "Total relative gain" and on 83th position, which is the last one, in its average of "Total relative shift".

The differences in ranking are not so significant in mathematics and English language tests, as they are in the case of Czech language. The largest ranking difference in mathematics between the averages of "Total relative gain" and "Total relative shift" is 14 positions and the same gap is 18 positions in the case of English language.

Let us now examine in what manner the school rankings correlate the averages of "Total relative gain" and "Total relative shift".

The Spearman's rank correlation coefficient between the school rankings in accordance to the averages of "Total relative gain" and "Total relative shift" in the case of Czech language is 0.604. The value of correlation coefficient proves that, even if some of the schools vary significantly in their ratings, the difference in total school rankings is not that remarkable. The correlation dependence in mathematics is very high. The Spearman's rank correlation coefficient between the school rankings in accordance to the averages of "Total relative gain" and "Total relative shift" in the subject of mathematics is 0.981, so the variation in rankings is minimal. And so is the coefficient between the school rankings in accordance to the averages of "Total relative gain" and "Total relative shift" in English language which is 0.975, thus the correlation dependence is again significant.

As the correlation analysis shows, in the subject of Czech language more remarkable variations occur. The variations in mathematics or English language are almost negligible.

Another way of confronting both methods is to compare the final ranking of individual pupils in "Total relative gain" and "Total relative shift" in the subject of Czech language. The largest gap in ratings is 2,012 positions. These considerable differences appear with pupils who had relatively high input ranking, so their relative shift is not very significant. Considering the first 15 pupils with the largest gaps in ranking, we find out that these pupils had excellent outcomes in both input and output tests. Their relative shift ranks them among the average, but in regard to their relative gain they rank among the best 500 out of the total of 4,714 pupils. No significant differences in pupil rankings occur in either method in tests from mathematics and English language.

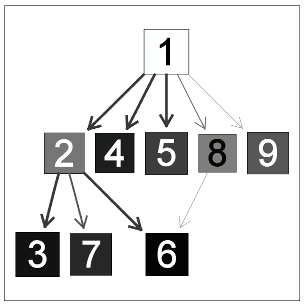
Analysis of value-added assessment outcomes variance among fields of study

We also carried out an analysis of variance in "Total relative gain" and "Total relative shift" of the fields of study. In all cases the statistic tests very strongly reject the zero hypothesis about non-existent differences among the fields of study (each with the p-value lower than 2.2×10^{-16}, see the table for values of the F-statistics).

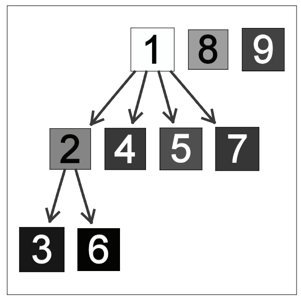
Subject	F-statistic value for	
	relative shift	relative gain
Czech language	30.307	61.702
Mathematics	43.025	76.910
English language	17.907	26.778

Table 9: F-statistic value for relative gain and relative shift

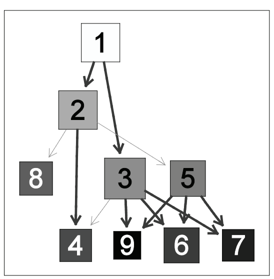
We calculated also a set of two selective Student t-tests (*non-pooled SD*) for each subject to prove statistically significant differences between individual field-of-study couples. To prevent the simultaneous statistic interference we have used the Holm's scheme. The results of this testing are demonstrated graphically using the Hasse diagrams (Burda, 2006).



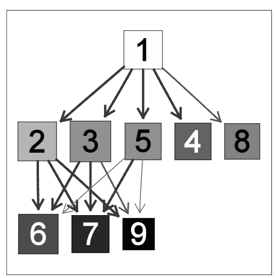
a) CL relative gain



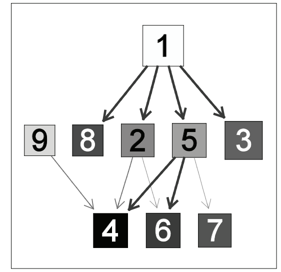
b) CL relative shift



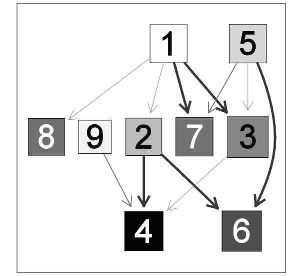
c) MA relative gain



d) MA relative shift



e) EL relative gain



f) EL relative shift

Drawing comparison between "Relative gain of the field of study" and "Relative shift of the field of study"

The analysis of variance in the previous part tells us that some fields differ significantly in their results. Pupils ordinarily select subjects which are suitable for them and which correspond with their skills and capabilities. This fact made us decide for the results comparison with the field of study as a factor affecting the value-added assessment to eliminate the pupils' input qualities and assess the value-added score only. We delimited the relative gain for each pupil calculated always under the terms of the field of study, hereafter as "Relative gain of the field of study", and we also delimited the relative shift calculated only under the terms of the field of study, hereafter as "Relative shift of the field of study".

Comparison of school rankings in "Total relative gain" and "Relative gain of the field of study" in each subject will be dealt with in the next chapter. Now we will have a look at a comparison of school rankings in "Relative gain of the field of study" and "Relative shift of the field of study" in individual subjects.

Considering the model where "field" means a factor affecting the value-added score, no significant differences can be found between the two methods. Even in the subject of Czech language the Spearman's ranking correlation coefficient is 0.928, i.e. very high correlation dependence in ranking. The largest gap is 32 positions. Sources of such differences are suggestion for further research and they will not be covered more in this article.

Regarding the school rankings in accordance to "Relative gain of the field of study" and "Relative shift of the field of study" in mathematics, the Spearman's rank correlation coefficient is 0.942. The following chart shows that differences between both models are comparatively small in the first ten places.

School	**Ranking by "Relative gain of the field of study" in MA**	**Ranking by "Relative shift of the field of study" in MA**
XXVIII	1	1
X	2	2
LXVIII	3	3
LXVII	4	4
LIII	5	11
XIV	6	9
LVII	7	15
XLII	8	19
XIII	9	7
XXIV	10	6
.	.	.
.	.	.
.	.	.
IV	74	79
LX	75	77
LXXX	76	75
L	77	74
LXXV	78	76
LXX	79	82
LXVI	80	78
XLVII	81	80
XXXIII	82	81
LXXI	83	83

Table 10: Differences between both models

The Spearman's rank correlation coefficient in the English language is 0.902 between "Relative gain of the field of study" and "Relative shift of the field of study", thus it is quite high as well.

Interesting thing is that the Spearman's rank correlation coefficient between "Relative gain" and "Relative shift" in Czech language noticeably drew near 1by adding the field as a value-added influential factor, while in mathematics and English language this value slightly lowered.

If we compare rankings of individual pupils in "Relative gain of the field of study" and "Relative shift of the field of study" in Czech language, the Spearman's rank correlation coefficient is 0.915, i.e. high. The situation is similar in the case of mathematics with the Spearman's rank correlation coefficient of 0.910 and English language with 0.933, i.e. high in both cases.

For drawing any conclusion, a deeper analysis of data causing this divergence is necessary. At the moment we could state that both methods assess the value-added score in different ways, but taking into account the value-added influential factors, the differences in both methods' results are comparatively slight. On selecting the model, it is important to consider what we are really about to assess – whether we are interested in shift under the terms of a group or gain against the presumed gain under the terms of a group of tested individuals.

Drawing comparison between "Total relative gain" and "Relative gain of the field of study"

In this part we will see what changes take part in the model of "Relative gain", provided we regard the "field" as an influential factor. That means we will divide all the tested individuals into 9 groups according to their field of study and we will delimit "Relative gains" only under the terms of a group.

If we compare the school rankings according to "Total relative gain" and "Relative gain of the field of study" in Czech language, we will logically deduce that shifts in ranking correspond with the Hasse diagrams created for the variance analysis of "Total relative gain" and "Total relative shift" of the fields of study. Some grammar schools met downgrade in school rankings, while schools with the fields of study no. 3, 6 and 7 reached quite considerable progress. The deepest fall is 50 positions, from the 21st place to the 71st place. The highest leap is 37 positions from the 44th place to the 7th place. As for the schools at the top or in the end of the chart, no significant divergence occurred as was presumed (see following figures). Negative figures mean a shift upwards in ranking, i.e. progress, and positive figures signal a shift downwards in ranking, i.e. downgrade, taking into account the field of study as an influential factor.

School	Change in ranking	Field of study
LVII	-37	6
XIV	-33	3
LIII	-28	3,9
VII	-28	3
LII	-27	6
LXVII	-24	3
LXXIII	-22	6
XXVIII	-20	2,3
LXXII	-20	3,6,7
XLV	-19	6

School	Change in ranking	Field of study
XLVII	50	1
XXIII	47	1
I	46	1
II	39	1
LXXIX	33	1
IX	32	1,8
XXIV	29	1
XXXIX	28	1
XI	27	1
LXVI	27	1

Table 11 and 12: School rankings according to "Total relative gain" and "Relative gain of the field of study" in Czech language

The situation in mathematics is similar. The highest progress can be seen in schools with the fields of study no. 6, 7, 9, and 4. The deepest downgrade occurs in schools with the field of study no. 1 which, again, corresponds with our variance analysis.

School	Change in ranking	Field of study
LVII	-45	6
LXXVI	-44	9
LXXIII	-38	6

VIII	-37	6
XLIII	-33	6
XLII	-31	4
LII	-30	6
XX	-27	7
LXXVIII	-26	2,7
XXXVI	-24	3,6,7

School	Change in ranking	Field of study
II	39	1
LIX	38	1
LXXXIII	38	1
XI	36	1
LX	35	1
LXXX	35	1
XXIII	33	1
XXV	31	1
LXVI	29	1
LXXXII	29	1,2

Table 13 and 14: School rankings according to "Total relative gain" and "Relative gain of the field of study" in mathematics

Finally, we will have a look at changes in school ranking in the English language subject. Here we can also see progress in schools in accord with the above mentioned variance analysis. Downgrade in rating also mostly occurs in schools with the field of study no. 1.

School	Change in ranking	Field of study
XLII	-58	4
LVII	-43	6
XLIII	-37	6
LXXVII	-36	2,4
LXVII	-27	3
VII	-23	3
LXXIII	-23	6
XXXIV	-22	2,4
LII	-22	6
VIII	-21	6

School	Change in ranking	Field of study
LXXX	35	1
XXIII	34	1
XXXIX	34	1
XLVII	34	1
LXXXI	33	1
I	32	1
II	31	1
XXIV	31	1
LXXVI	31	9
XXV	29	1

Table 14 and 15: School rankings according to "Total relative gain" and "Relative gain of the field of study" in English language

The above figures and analysis of variance (see subsection 3.2) prove that it is more than appropriate for the model to regard the "field of study" a factor affecting the calculation of value-added score. There are undoubtedly other influential factors, but for a lack of relevant information their influence on the above mentioned models was not attestable.

Discussion

The paper claims that the field of study could be a significant factor influencing the value-added assessment models. However, the extent to which it influences the model is heavily determined by its definition. Our division of the schools into nine fields of study is based on an analysis of school educational programmes. It is also based on findings that socio-economic factors significantly determine which school student choose.

We have identified nine types of schools with similar study plans. As can be seen in the results presented in this paper, our distribution of schools into the fields of study works well in the Czech Republic – other countries with different educational programmes may need to develop their own distribution.

It is known that socio-economic factors could also significantly affect the value-added assessment models. However, such research is left for the future.

Conclusion

The article presents two important value-added assessment models: method of relative gain of knowledge and method of relative shift. Thanks to sufficiency of comparatively vast sets of data, it was possible to confront both models from the point of units' position in the set, ranked from the best to the worst. The final results could suggest that in sufficiently large sets there will be no larger differences in final ranking of results, but these differences are quite remarkable for some of the individuals. The article at the same time observes in what manner the values change when information on the field of study is taken into account. The analyses we carried out tell us that it is more than appropriate to regard the field of study as an influential factor for the value-added assessment. When regarding the field of study an influential factor for the value-added assessment, the differences in both models outcomes are mostly slight or comparable. We cannot decide which model is correct, since each of them has its use in certain situation and certain assessment.

In following research we will examine and compare further value-added assessment methods. As very appropriate we find the method of multiple regressions, using other socio-economic factors including the Rasch analysis. It is also advisable to explore more deeply what causes those more significant differences the positions of individuals in final ranking.

References

Burda, M. (2006) Visualization of cosymmetric association rules using Hasse diagrams and concept lattices. In: *Znalosti 2006*. pp. 175-182.

Chráska, M. (2007) *Metody pedagogického výzkumu: základy kvantitativního výzkumu*. Praha: Grada.

Liu, O. (2011) Value-added assessment in higher education: a comparison of two methods. *Higher Education*, vol. 61, no. 4, pp. 445-461. Netherlands: Springer Netherlands.

Lissitz, W. R. (2005) *Value Added Models in Education*. Maple Grove: JAM Press.

Malčík, M., Krpec, R. (2010) Monitoring přidané hodnoty ve vzdělávání v Moravskoslezském kraji. Ostravská univerzita v Ostravě. [2010-09-30].

Malčík, M. (2007) Electronic support of educational results measuring. In: Proceedings Information & Communication Technology in Education 2007, Rožnov pod Radhoštěm.

Malach, J., Malčík, M. (2010*) Value–added assessment in postsecondary schools. Theoretical approaches and research results in the Czech republic.* Kultura i Edukacja 2010 No. 5 (79).

McCaffrey, D., Lockwood, J., Koretz, D., Louis, T., Hamilton, L.,(2004) Models for Value-Added Modeling of Teacher Effects. *Journal of Educational and Behavioral Statistics*, vol. 29, no. 1, pp. 67-101. Washington: American Educational Research Association.

Raudenbush, S., W., Willms, J., D. (1995) The Estimation of School Effects. *Journal of Educational and Behavioral Statistics*, vol. 20, no. 4, pp. 307-335.

Sanders, W., L., Horn, S., P. (1994) The tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, vol 8., issue 3, pp. 299-311. Netherlands: Springer Netherlands.

SCIO (2008). Vektor Modul 3: Manuál. Praha.