

Summary of doctoral dissertation

**TEACHERS' BELIEFS ON SYSTEM LEVEL ASSESSMENTS AND
THEIR IMPACT ON THE TEACHING AND LEARNING PROCESS**

EDIT TÓTH

Supervisor:

**Benő Csapó
DSc**

**University of Szeged
Faculty of Arts
Doctoral School of Educational Sciences
Teaching and Learning Doctoral Program**

2014

Subject and structure of the dissertation

In the past 15 years studies designed and conducted by international organizations became frequent and many countries developed their own national educational assessment program in order to improve the education system. The emerging reforms brought along a governmental demand to interpret the success and efficiency of the educational systems in terms of students' knowledge and to hold schools accountable for the efficient use of allocated resources based on the educational outcomes. In addition, a major role was dedicated to support the institution level improvement of education, the teaching and learning process (*Looney, 2011*).

The assessment programs may contribute to the improvement of the efficiency and success of the education system by providing stakeholders with assessment feedback. The assessment can only fulfil its role in development if the knowledge and information yielded by and gained from the assessments become accessible to the stakeholders and if the stakeholders are aware how they can utilize the information in diagnosing and tackling problems. However, further research and development providing information on the effects of the assessment feedback is needed in order to facilitate the efficiency of the assessments.

Hungary has been taking part in the IEA and OECD studies from their conception. In parallel with these efforts, Hungary has been developing its national educational assessment program since the 2000's. As a fundamental pillar to the assessment system, the National Assessment of Basic competencies received a decisive role in institutional accountability in the 2007/2008 academic year.

We have had the opportunity to benefit from the international – especially Anglo-Saxon – experience accumulated during the past few decades on the functioning of the assessment programs and accountability systems based on them. However, we do not have sufficient empirical evidence on the efficiency and impact of these initiatives. Relying on the framework and current conceptualization of accountability our dissertation aims to examine (1) if the information yielded by system-level assessments and subsequent knowledge appears in schools and (2) how teachers interpret the assessments and the feedback provided by them. Our objective is to collect data which may contribute to the optimization of the national assessment system.

The first section of the thesis entails a literature review and is divided into three distinct parts (parts 1-3). In the first part we introduce the most important factors in establishing a system level student achievement assessment program in certain countries. We also present the conceptualization and framework of accountability as one of the objectives of the assessment programs. In addition, we study the incentive nature of assessments through the principal-agent dilemma and we point out the distortive effects of the dysfunctional incentive systems. The major components of the national assessment system and their impact analyses are detailed in part 4. Part 5 entails the conceptualization, research questions, and hypotheses. The methods section is included in part 6, and the results form the 7th and 8th part of the work. The final section of the thesis includes the summary and implications for further research.

The theoretical background and conceptualization of the study

The theoretical background of the empirical study is provided by the well-known principal-agent dilemma in economics (see *Levačič*, 2009). We use this perspective to introduce the theory and limitations of the test-based accountability systems. We integrate the results of the models describing the feedback mechanisms of the system level measurement-assessment programs in order to be able to study their effects in learning and instruction with special attention to the incentives.

According to *Frink* and *Klimoski* (2004) the definitions of accountability may be organized around two well-defined themes. One of the themes is the context, which refer to the participants of the situation, whereas the other one is the evaluation or feedback which may take any forms. The first theme concerns the interpersonal context with having two participants in focus: the principal and the agent. The principal assigns a task to the agent for they are incapable of delivering it by themselves. Both parties seek to maximize their own profit. The other theme concerns the control and incentivization of the agent. These activities serve the purposes of risk minimization arising from contra selection (i.e. the competency/eligibility of the agent) and moral issues (i.e. the agent acts driven by their own objectives without taking into consideration the aims ad purposes of the principal). A complex hierarchy and a huge number of principals characterize the public sector and consequently the education itself (*Levačič*, 2009).

The governments running the student assessment based accountability systems use students' test achievement to define the efficiency of the education; the results provide a solid foundation for creating incentive schemes differentiating between good and poor performing institutes and teachers to facilitate a better quality of learning and instruction. The two most common forms of incentives are the implicit (publicity of results) and explicit incentives (concrete rewards, sanctions), which can affect students, teachers, heads of institutions and districts.

The majority of information about the effects and consequences of accountability systems is provided by literature giving account on the experiences gathered in Anglo-Saxon countries. Numerous studies pointed out that the introduction of test based accountability programs had an impact on students' improvement (e.g. *Jacob*, 2002; *Carnoy & Loeb*, 2002, *Hanushek & Raymond*, 2005). In addition, data seem to underscore that explicit incentivization had a more significant effect on students' achievement than implicit incentivization (*Hanushek & Raymond*, 2005; *Dee & Jacob*, 2011). However, *Cannel* (1987) reported – as early as in the second half of the 80's – that in many US states a disproportionate number of districts published reports on the results of student assessments claiming that students perform well above the national average. Research carried out to unveil the reasons underlying this phenomenon (see *Linn & Dunbar*, 1990, *Koretz & Barron*, 1998, *Fuller, Gesicki, Kang & Wright*, 2006) showed that an increase in students' performance in high-stakes tests is much more significant than that measured by other measures (e.g. NAEP). Studies trying to identify the factors that may account for this difference (see *Jacob*, 2007, *Jacob & Levitt*, 2003,

Zen, 2007, Figlio & Gletzer, 2002). In what follows, we will summarize teachers' reactions to assessments using the data accumulated by international classroom observations, interviews and questionnaires.

Research showed that teachers perceive pressure exerted by high stakes and low stakes testing procedures to a different extent and incentives have a different role on the classroom work. Data suggest that system level assessments have an impact both on the content and methods of teaching and on classroom assessment. In addition, studies highlight the positive and adverse effects of accountability systems on the teaching process. The incentives with inadequate design and unclear communication may lead to narrowing the curriculum and to the underrepresentation of non-assessed contents and domains. Teachers perceiving pressure exerted by the different stakeholders spend a considerable amount of time on teaching for testing and therefore constructivist teaching methods and the acquisition of higher order thinking skills are overshadowed. The format of high-stakes tests influences classroom assessment, which is mainly reflected in the format of classroom tests.

The background and aims of the empirical study

The components of the national measurement-assessment system established in the 2000s and the assessment cycles of the international (TIMSS, PIRLS, PISA) and national surveys affect every public education institution. The national educational assessment system maintains three elements: the DIFER, the Diagnostic Skills and Ability Assessment and the National Assessment of Basic Competencies. At the time of our research all 4th graders of each and every school completed the tests of the diagnostic assessment in grade 4 and every 6th, 8th and 10th graders took part in the National Assessment of Basic Competencies. The diagnostic and criterion-referenced diagnostic assessment in grade 4 primarily aims at supporting teachers' work and students' development. Beyond the institution and classroom level improvement of education the National Assessment of Basic Competencies is also a major tool of the institution level accountability of education. The results of the National Assessment of Basic Competencies are publicly available. In Hungary due to the "free" school choice the publicized assessment data have a remarkable effect on schools as they may influence the number of students enrolled into the institution and the income as well. In addition, schools have to prepare an action plan if the results of the NABC show that in any seat of the institution half of the students do not succeed in achieving the minimum criterion set forth by the law. The action plan must be approved by the administration. Based on the literature review and the characteristics of national (diagnostic assessment in grade 4, National Assessment of Basic Competencies) and international (TIMSS, PIRLS, PISA) assessment programs the aim of our research is to examine teachers' beliefs, affective reactions and experiences towards system level assessment programs, and to examine how these factors affect their teaching process – the content and methods of their teaching and the classroom level assessment.

The research questions of the empirical study

Based on teachers' emic perspectives the objective of the present study is to examine teachers' attitude towards system level assessments, their perceptions on assessments and their views on the impact of system level assessments on the teaching and learning process.

(1) Teachers' attitudes towards the acceptance and usefulness of the system level assessments

What beliefs and attitudes do teachers have towards system level assessments? What is teachers' opinion on the assessment feedback?

(2) Educational stakeholders' perceived pressure

To what extent do teachers feel pressurized by different assessments?

How do teachers perceive stakeholders' pressure towards improving students' achievement?

(3) Teachers' beliefs and opinion on the changes system level assessment programs evoke in their teaching practices

How do teachers reflect on the shift in their teaching practices influenced by system level assessments? Based on teachers' emic perspectives what are the most frequently used strategies in the preparation for DSAA and NABA?

What are the relationships between teachers' opinion on system level assessments and shift in teaching practices and preparation strategies for DSAA and NABC?

What is the relationship between the pressure of stakeholders interested in the efficiency of the education system and the perceived change in teachers' teaching practices and the preparation strategies for the DSAA and NBAC?

(4) What is the relationship between teachers' beliefs, the institution level results of the National Assessment of Basic Competencies and the background variables (e.g. taught subject, school type, school (ISCED) level)?

Sample and instruments of the empirical study

We conducted the first empirical survey in spring 2010, in which 486 primary school teachers, 726 upper elementary school teachers and 1191 upper secondary school teachers took part. The sampling unit was the school (N=350). Each upper elementary school questionnaire was filled out by teacher of mathematics, Hungarian grammar and literature and some natural sciences. In addition, in each secondary foreign language

teachers and art teachers were also included in the sample. The two state-wide samples are representative for region and settlement type. Furthermore, the upper secondary sample is representative for school type/program. The second survey was conducted in spring 2012 in ten elementary schools and five upper secondary schools in a town. In our survey teachers of mathematics, Hungarian grammar and literature and some natural sciences, or foreign language or arts and principals took part.

The instruments of the survey were one questionnaire and an interview. The teachers' opinions were mainly assessed on a four level Likert scale (1=disagree; 4=agree). The battery was based on numerous international questionnaires – the subscales were collected and linked in the International Project for the Study of Educational Accountability Project. At the adaptation we had to take into consideration the differences in the education systems. These differences limited the literal translation; we transformed the items to a Hungarian context.

The questions in the study focus on teachers' view on (1) assessments in general and (2) their reports, (3) perceived pressure for different type of measurements/assessments, (4) perceived pressure by different stakeholders, (5) time spent on preparing for the NABC, (6) changes in instruction due to large scale assessments and the NABC. The questionnaire consists of 79 items. Background variables were derived from a background questionnaire, the database of the NABC and the KIR-stat database and the documentation of schools. The interview with its six subjects targeted the further interpretation and the resolution of the contradictions in the answers of the questionnaire, and the collection of information that is too difficult to standardize by means of a questionnaire.

Results of the empirical study

In our teacher study, we analyzed the characteristics of the beliefs concerning the assessments broken down into an institutional, subject and school type level. Based on previous studies conducted among principals and heads of institutions we hypothesized that there are differences between teachers' answers accordingly. In case of the school level we expected a difference between elementary and secondary school teachers (*Pendulla et al., 2003; Moore & Waltman, 2006*), whereas in case of the school types a difference in vocational school teachers was anticipated from the grammar school teachers and vocational secondary school teachers (e.g. *Sinka, 2006*). Based on relevant international research (e.g. *Hamilton, Berend & Stecher, 2005*) and the characteristics of national assessment programs we hypothesized that the beliefs of teachers of different subjects will differ.

(1)When examining the acceptance of system level assessments we assumed that teachers acknowledge the importance of system level assessments and their impact on their work and practices and they consider the assessments to be less adequate tools in the objective evaluation of the performance of schools and education systems. Based on the answers, teachers tend to accept system level assessments and find them useful.

Secondary school teachers accept the assessments better than their colleagues teaching in elementary schools. Nevertheless, we can identify a group of teachers (elementary school: 32%, secondary school: 27%), who consider the assessments as sources of stress and feel that they have more drawbacks than benefits. Similarly, according to a considerable proportion of teachers (elementary school: 44%, secondary school: 53%) system level assessments cannot remarkably contribute to the objective judgement of the performance of schools and school systems. Nevertheless, secondary school teachers do not really feel the impact of assessments on their work. The answers provided by lower and upper grade teachers do not differ in the measured domains. Grammar school teachers proved to accept the assessments better than vocational school teachers; however, its impact is more significant on the work of vocational teachers. Answers show that only 15% of the respondents have difficulty interpreting the NABC reports. The proportion of teachers assuming that the assessment feedback yielded by the NABC mirror students' skills and knowledge and considering NABC as reliable measures and source of feedback on the quality of school and classroom work is low. Nearly 50% of elementary and 80% of secondary school teachers took a look at the results of international assessments. The results are considered reliable only by a small number of respondents.

(2) Based on previous international studies and incentive theories, we accept the hypothesis that teacher feel pressurized by external achievement assessment programs (e.g. *Aydeniz & Southerland, 2012*). Starting out from the characteristic features of the assessments we assumed that teachers feel pressurized to a different extent by assessments carried out by different organizations. The responses show that lower grade teachers mostly feel pressurized by the diagnostic assessment in grade 4, whereas the upper grade teachers tend to feel pressurized by the NABC. However, this tendency does not occur among teachers in secondary education, in secondary education matura is the most frustrating assessment. Elementary school teachers reported on a higher level of stress triggered by internal school assessments and the assessments carried out by county level educational institutes, research groups than secondary school teachers. The NABC is a more significant source of stress for teachers of Hungarian grammar and literature and mathematics than for science teachers. In secondary schools teachers of Hungarian grammar and literature, mathematics and foreign language mostly feel overwhelmed by the expectations set by the NABC, whereas the responses of science teachers and art teachers did not support this tendency. The results of the interview show that teachers in general hold the mathematics and Hungarian grammar and literature teachers responsible for the good performance of students. Results also reveal that many institutions have the same attitude, that is, the leadership holds these teachers accountable for better results. The interviews suggest that teachers' frustration is mainly caused by the publicity of assessment results as teachers believe that they can influence public opinion and the number of enrolled students. On the other hand, local institution level practices of using the assessment data in the evaluation of primarily school teachers', Hungarian grammar and literature and mathematics teachers' work is also a major source of frustration.

(2) According to our results the teachers are afraid of the pressure caused by the school management. Teachers' own expectations about their work also provide a driving force. The answers show that the expectations of teachers and students hardly play an incentive role. Secondary school teachers feel less pressurized by the stakeholders than upper grade elementary school teachers. Due to the diagnostic assessment in grade 4 teachers feel the same amount of pressure caused by all agents than upper grade teachers due to the NABC. The responses of teachers in different school types differ from each other: vocational school teachers feel more significantly pressurized by the colleagues, school management and the municipality, than grammar school teachers. However, parents' pressure is the strongest in grammar schools. In secondary schools the Hungarian grammar and literature teachers perceive bigger pressure by the municipality and the school management than their other colleagues. The interviews highlighted what teachers think about the consequences of the test results. They also showed their perceptions on how and what the stakeholders use the results of the assessments for. According to the answers the NABC has a significant role in the professional judgement and evaluation of teachers. There are schools where the school management use the results of NABC and to evaluate a teacher's work, in spite of the fact that the results of the assessment program do not serve these purposes, because both of the assessed domains are considered as key competencies and the regular and continuous development of these areas must be carried out during each lesson. Most of all the mathematics and Hungarian grammar and literature teachers feel responsible for students' NABC test results. Although every teacher think of development as an important task, in practice great emphasis is laid on development by mathematics and Hungarian grammar and literature teachers. A significant group of teachers attributes importance to the environment's, the local community's and especially parents' conclusions drawn from the test results. They think that the results of the NABC and DSAA have an impact on how parents judge the school. Furthermore, one third of the respondents think that NABC results influence parents' school choice. Thus, due to the free school choice a high stake is attributed to the NABC in elementary schools. In secondary schools concern for the enrollment is hardly perceived.

(3) Our results show that system level assessments do have an impact on teaching practices. Findings yield evidence that the assessments have a major effect on the teaching methods. In addition, teachers claim that the assessment programs brought the curriculum into the forefront and training/teaching requirements and as a result their work became more focused. The assessments highlight the necessity of supporting and developing poor performers and as a consequence, teachers pay more attention to these students. In contrast with international research evidence, our data does not support the fact that teachers modify the topics and content of their teaching because of the assessments since no patterns emerged concerning the reduction in the quantity of taught elements or content. The impact of assessments is mirrored in the following practices: teachers pay more considerable attention to practicing generala and universal

test taking strategies and the test formats used in system level assessment also appear in the everyday assessment practices.

The answers of primary school and upper elementary school teachers differ in terms of the shift in teaching content and practicing test taking strategies. The answers suggest that secondary school teachers are less likely to change their teaching practices than elementary school teachers. In lower secondary schools Hungarian grammar and literature teachers differ from the other subgroups (mathematics and science teachers) in three dimensions: changes in teaching methods, alignment in the measured and taught contents and practice test taking strategies. Based on the interviews we hypothesize that reading literacy gets a significant attention in schools.

Teachers tend to use the item and test types/formats used in DSAA and NABC. The answers from the interviews show that these practices do not necessarily aim at increasing the efficiency of instruction; they are rather short term targets with having good performance on the NABC in the focus. Elementary school teachers are more likely to prepare their students for the NABC than secondary school teachers. Based on the reports Hungarian literature and grammar teachers in secondary schools are more likely to prepare students for the NABC than teachers teaching other subjects. Vocational school teachers are more likely to teach to the test than their high school counterparts. In elementary schools more than 95% of Hungarian grammar and literature and mathematics teachers prepare their student and practice for the NABC, whereas in secondary schools, 70% of mathematics teachers and 80% of Hungarian grammar and literature teachers do so. On average 15 lessons are spent on preparing for the assessments in elementary schools and 10 lessons are dedicated to related activities in secondary schools.

(6-7) Variables describing a change in the teaching practice were extracted into 5 factors: better instruction, more homework, reallocation of teacher's focus and attention, practicing test taking strategies and narrowing of instruction. The stakeholders of NABC were grouped into 3 factors: school level, outside school and beneficiaries. We used the factors to examine the relationships between the change in teaching practices, stakeholder pressure and teachers' beliefs towards the assessments. Our results show that the change in teachers's practices is influenced by different stakeholders' expectations. The stakeholders exert pressure in different fields and ways. Teachers' likeliness to look for more efficient methods and their more disciplined attitude towards shaping the content and provide better instruction are influenced by the extent to which they accept the assessments and the pressure they perceive in their school. Teachers' and students' expectations have an influence on homework assignment. The expectations of school level stakeholders and parents affect the reallocation of teachers' focus. All three stakeholders as well as teachers' attitudes towards the assessments have an impact on practicing test taking strategies. Narrowing curriculum is influenced by school level stakeholders. Preparation of students for the NABC is primarily influenced by school level stakeholders.

Educational importance, limitations and implications for further research

The efficiency and usefulness of our measurement-assessment system depends to a great extent on how teachers relate to it and how schools use its results. Primarily, our results may provide educational policy makers and stakeholders in teacher training with valuable information. We have insufficient information on teachers' beliefs and attitudes about the NABC and there is scant research evidence on how these factors influence the teaching and learning process and the shifts in the methods and contents of education. Elaborated empirical research aiming at mapping out the impact of system level, high-stakes assessments – especially the NABC – triggering school accountability on the teaching and learning process is scarce. The educational importance of the empirical research lies in the fact that we explored the characteristic features of numerous previously unknown and unexamined beliefs, experiences, attitudes and affective reactions regarding the NABC and diagnostic assessment in grade 4 broken down into school level, school type and major. The results are in line with some previous findings; however, due to the context dependent nature of the topic, they reveal the underlying mechanisms and effects in a Hungarian context.

Our result may serve as a reference for further comparative analyses. In addition, the hypotheses which may be formulated on the basis of the interviews and the findings may provide a guideline and basis for further research.

Our results may be used to study the effects of different educational assessment trainings and to define the agenda and target audience of lectures and trainings in this field. Furthermore, they highlight the fact that bigger emphasis should be laid upon the integration of the body of knowledge on assessment, cybernetics, the relationship between the assessed and taught domains and constructs, the importance of skill and ability development into the teacher training curriculum because this measure seems to be an efficient tool in the further improvement of education.

Our results should be interpreted in light of the following limitations. Although our research exhibits a mixture of qualitative and quantitative approaches, the assessment instruments used in the study may only serve as means to draw indirect conclusions from teachers' opinions regarding the examined phenomena. We did not examine or observe teachers' actual behaviours and practices and no classroom observations were carried out. We do not possess any information on the extent of teachers' conformity or willingness to match their attitudes and behavior to a hypothesized social norm. Our results do not yield direct evidence on how the underlying mechanism of the effects of assessments on the teaching and learning process, thus, in many instances we can only derive hypotheses.

We found several reasons for the further improvement of our assessment instrument. On the one hand, at the time of devising the instrument we found it imperative to provide an opportunity for international comparison. Thus, some themes

did not prove to be relevant in a Hungarian context but they can be complemented with additional items which serve the better understanding of the impact of the domestic system level assessments and accountability on institutional and teacher practices.

The study variables should be complemented with the teachers' age, professional experience, and teachers' knowledge by all means in order to examine the relationship between these factors and teachers' attitude about the measurement-assessment programs. In addition, it would be beneficial to expand the scope of the study to an institutional level. This way the relationship between teachers' attitudes and beliefs and institutional variables (e.g. institutional education program, quality management programs and efforts, prevailing assessment culture) could also be analyzed.

According to the national public education act accepted in 2011 the decentralized maintenance of the educational institutions became centralized/governmental; the local/regional bodies of the Klebelsberg Institution Maintenance Centre perform the maintainer tasks. Along with the central maintenance, the autonomy of the educational institutions weakened, which has an impact on the functioning of institutions and the professional work too. The latter is associated with changes in the curriculum, material and school book regulation. These, in contrast with previous process are steps taken towards the strengthening of the input based control of the schools. In addition to the input based control, the government/the law reinforces the process based accountability by the establishment of the school inspection system. The NABC remains a tool for performance or outcome based accountability. It would be worthwhile to explore teachers' attitudes in the new maintainer and regulatory environment.

The empirical study was carried out with the support and infrastructure of the MTA-SZTE Research Group on the Development of Competencies and the Center for Research on Learning and Instruction, University of Szeged.

The research was also supported by the **European Union** and the **State of Hungary, co-financed by the European Social Fund** in the framework of TÁMOP 4.2.4. A/2-11-1-2012-0001 'National Excellence Program'.

References

- Abrams, L., Pedulla, J., és Madaus, G. (2003): Views from the classroom: Teachers' opinions of statewide testing programs. *Theory into Practice*, **42**. 1. sz. 18–29.
- Aydeniz, M. és Southerland, S. A. (2012): A National Survey of Middle and High School Science Teachers' Responses to Standardized Testing: Is Science Being Devalued in Schools? *Journal of Science Teacher Education*, **23**. 3. sz. 233–257.
- Bourque, M. N. (2004): Leave No Standardized Test Behind. In: Phelps, R. P. (szerk.): *Defending Standardized Testing*. Lawrence Erlbaum Associates, New Jersey. 227–253.
- Cannell, J. J. (1988): Nationally normed elementary achievement testing in America's public schools: How all fifty states are above the national average. *Educational Measurement: Issues and Practice*, **7**. 2. sz. 5–9.
- Carnoy, M. és Loeb, S. (2002): Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, **24**. 4. sz. 305–331.
- Clarke, M., Shore, A., Rhoades, K., Abrams, L., Miao, J. és Li, J. (2003): *Perceived effects of state-mandated testing programs on teaching and learning: findings from interviews with educators in low-, medium-, and high-stakes states*. National Board on Educational Testing and Public Policy. Boston College. <http://www.bc.edu/research/nbetpp/statements/nbr1.pdf>. (2014. május 12-i megtekintés)
- Dee, S. és Jacob, B. (2011): The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, **30**. 3. sz. 418–446.
- Figlio, D. N. és Loeb, S. (2011): School Accountability. In: Hanushek, E. A., Machin, S. és Woessmann, L. (szerk.): *Handbook of the Economics of Education, Volume 3*. Elsevier, North-Holland Publishing Company, Amsterdam. 383–421.
- Frink, D. D. és Klimoski, R. J. (2004): Advancing Accountability Theory and Practice: Introduction to the Human Resource Management Review special edition. *Human Resource Management Review*, **14**. 1. sz. 1–17.
- Fuller, B., Gesicki, K., Kang, E. és Wright, J. (2006): *Is the No Child Left Behind Act Working? The Reliability of How States Track Achievement*. Working Paper 06-1. Policy Analysis for California Education, Berkeley.
- Hamilton, L. S., Berends, M. és Stecher, B. (2005): *Teachers' Responses to Standards-Based Accountability*. Working Paper. RAND Corporation, Santa Monica.
- Hanushek, E. A. és Raymond, M. E. (2005): Does school accountability lead to improved student performance. *Journal of Policy Analysis and Management*, **24**. 2. sz. 297–327.
- Jacob, B. A. (2005): Accountability Incentives and Behavior: The Impact of High Stakes Testing in the Chicago Public Schools. *Journal of Public Economics*, **89**. 5–6. sz. 297–327.
- Jacob, B. A. (2007): *Test-based Accountability and Student Achievement: An Investigation of Differential Performance on NAEP and State Assessments*. Working Paper 12817. National Bureau of Economic Research. <http://www.nber.org/papers/w12817.pdf>. (2014. március 14-ei megtekintés).
- Jones, B. D. (2007): The unintended outcomes of high-stakes testing. *Journal of Applied School Psychology*, **23**. 2. sz. 65–86.
- Jones, B. D. és Egley, R. J. (2004): Voices from the frontlines: Teachers' perceptions of high-stakes testing. *Education Policy Analysis Archives*, **12**. 39. sz. 1–29.
- Koretz, D., Linn, R. L., Dunbar, S. B. és Shepard, L. A. (1991): *The Effects of High Stakes Testing: Preliminary Evidence about Generalization across Tests*. <http://nepc.colorado.edu/files/HighStakesTesting.pdf> (utolsó megtekintés: 2014. április 10.)
- Koretz, D. (2002): Limitation in the Use of achievement tests as measures of educators' productivity. *Journal of Human Resources*, **37**. 4. sz. 752–777.
- Koretz, D. (2008b): *Measuring up: What educational testing really tells us*. Harvard University Press, Cambridge.
- Koretz, D. és Barron, S. I. (1998): *The Validity of Gains on the Kentucky Instructional Results Information System (KIRIS)*. RAND Corporation, Santa Monica.
- Koretz, D., McCaffrey, D. és Hamilton, L. (2001): *Toward a Framework for Validating Gains Under High-Stakes Conditions*. CSE Technical Report. N. 551. Center for the Study of Evaluation, University of California, Los Angeles.
- Levačič, R. (2009): Teacher Incentives and Performance: An Application of Principal-Agent Theory. *Oxford Development Studies*, **37**. 1. sz. 33–46.

- Linn, R. L. és Dunnbar, S. B. (1990): The Nation's Report Card Goes Home: Good News and Bad About Trends in Achievement. *Phi Delta Kappan*, **72**. 10. sz. 127–133.
- Moore, J. és Waltman, K. (2007): *Pressure felt by teachers to increase test scores in reaction to NCLB: An investigation of related factors*. (Paper) Annual meeting of the American Educational Research Association, Chicago, IL.
- Pedulla, J., Abrams L., Madaus, G., Russell, M., Ramos, M., és Miao, J. (2003): *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Boston: National Board on Educational Testing and Public Policy, Lynch School of Education, Boston College.
- Williams, L. (2009): Down and dirty with grammar. *Literacy Learning: the Middle Years*, **17**. 1. sz. 11–20.

Publications in connections with the dissertataion

- Tóth Edit és Hódi Ágnes (2014): *How Hungarian teachers' misconceptions may lead to the dysfunction of a national large-scale diagnostic assessment program: the implications of a small-scale inquiry*. (Paper) EARLI SIG 23. Bielefield, Germany, November 10–12. (közlésre elfogadva)
- Tóth Edit és Sipos Judit (2014): A rendszerszintű mérések hatása a pedagógusok tanítási gyakorlatára: egy tanári kérdőíves vizsgálat tanulságai. *Iskolakultúra*, **24**. 10. sz. 38–52. (megjelenés alatt)
- Tóth Edit és Hricsovinyi Julianna (2014): Az Országos kompetenciamérés intézményi hasznosítása *pedagógusok körében készített interjúk alapján*. 12. Pedagógiai Értékelési Konferencia, Szeged, 2014. május 1–3. 87.
- Tóth Edit és Tóth Krisztina (2014): *A rendszerszintű mérések elfogadottsága és tanításra gyakorolt hatása közötti összefüggések vizsgálata pedagógusok körében*. 12. Pedagógiai Értékelési Konferencia, Szeged, 2014. május 1-3. 86.
- Tóth Edit és Hricsovinyi Julianna (2013): *Pedagógusok tapasztalatai és tájékozottsága az Országos kompetenciaméréssel kapcsolatban*. XIII. Országos Neveléstudományi Konferencia, Eger, 2013. november 7–9. 520.
- Tóth Edit (2013): A teljesítmény alapú ösztönzés a közoktatásban. In: Molnár Gyöngyvér és Korom Erzsébet (szerk): *Az iskolai sikerességet befolyásoló kognitív és affektív tényezők értékelése*. Nemzedékek Tudása Tankönyvkiadó, Budapest. 221–236.
- Tóth Edit (2012): Pedagógusok nézetei az osztályozásról és a szöveges értékelésről. *Iskolakultúra*, **22**. 1. sz. 73–83.
- Tóth Edit (2011): Pedagógusok nézetei a tanulóiteljesítmény-mérésekről. *Magyar Pedagógia*, **111**. 3.sz. 225–249.
- Tóth Edit és Kinyó László (2011): *Pedagógusok nézetei a mérés-értékelés néhány aspektusáról és ezek kapcsolata a tanulók teljesítményével*. XI. Országos Neveléstudományi Konferencia, Budapest, 2011. november 4–6. 254.
- Tóth Edit és Csapó Benő (2011): *Comparing Primary and High-School Teachers' Attitudes towards Testing and Accountability*. Paper. 14th European Conference for the Research on Learning and Instruction. Exeter, United Kingdom, August 30–September 3, 2011. 57.
- Tóth Edit és Csapó Benő (2011): *Teacher's Opinion on Testing and Accountability*. Paper. In: *Inciting the Social Imagination: Education Research for the Public Good*. New Orleans, United States of America, April 8–12, 2011. 221.

- Tóth Edit és Csapó Benő (2010): *A tanulói teljesítménymérések hatása a tanítási folyamatra az általános iskolában dolgozó pedagógusok nézetei alapján*. X. Országos Neveléstudományi Konferencia, Budapest, 2010. november 4–6. 169.
- Tóth Edit, Csapó Benő és Székely László (2010): Az iskolák és osztályok közötti különbségek alakulása a magyar iskolarendszerben: egy longitudinális vizsgálat eredményei. *Közgazdasági Szemle*, **57**. 9. sz. 798–814.
- Tóth Edit (2010): Teszt alapú elszámoltathatóság a közoktatásban. *Iskolakultúra*, **20**. 1. sz. 60–79.
- Tóth Edit (2009): *A tanárok értékeléssel kapcsolatos attitűdjei, nézetei*. IX. Országos Neveléstudományi Konferencia. Veszprém, 2009. november 19–21.
- Tóth Edit és Csapó Benő (2008): *A teszt alapú elszámoltathatósági rendszereken belüli érdekek és érdekkonfliktusok hatása az oktatási gyakorlatra*. VIII. Országos Neveléstudományi Konferencia, Budapest, 2008. november 13–15. 290.
- Tóth Edit (2008): Koretz, Daniel: Measuring Up: What Educational Testing Really Tells Us. *Magyar Pedagógia*, **108**. 4. sz. 283–285.