How do Teachers in Different Stages of Formal Education Respond to Accountability Pressure?

© Edit TÓTH, © Ágnes HÓDI
MTA-SZTE Research Group on the Development of Competencies, Szeged, Hungary

In line with international trends Hungary has been developing its own national assessment system since 2001. The system consisted of three major pillars until 2012: (1) a school-readiness test battery administered when entering formal education (2) the Diagnostic Skills and Ability Assessment (DSAA) assessing 4th graders, and (3) the National Assessment of Basic Competencies (NABC) carried out in grades 6, 8 and 10. Introduced in 2006, NABC has a twofold function: in addition to playing an important role in supporting learning and instruction, it is also a powerful tool of institutional accountability. The implementation of assessments with accountability purposes is closely intertwined with pressure conceived by educational stakeholders. This relationship is verbalized in Koretz’s (2008) conceptualization of high-stakes testing referring to tests which make stakeholders feel pressurized to increase performance. There is abundant evidence claiming that high-stakes testing serving accountability purposes pressurizes teachers (e.g. Koretz, 2008) and has unintended effects on the teaching practice (e.g. Nichols & Berliner, 2005). Although literature generally relates teachers’ perceived pressure to high-stakes testing, comparative research on the impact of centralized nation-wide testing – regardless of its aims and purposes – on teachers is scant.

Theoretical background

Teacher pressure has come to the forefront of research as a result of ever growing national efforts to hold stakeholders of education accountable for students’ performance. Tests along with the institutional-level publicity of assessment results and the related incentives provide the fundamentals of educational accountability. High-stakes testing relies on the assumption that the quality of education can be improved to a great extent “by introducing a system of rewards and sanctions that are triggered by students’ standardized test performance” (Nicols, Glass & Berliner, 2006:2). The well-known
principal-agent dilemma in economics (see Levačič, 2009) may also provide an adequate framing to a better understanding of the underlying mechanism of educational accountability. 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 assign 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).

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. Rewards and sanctions as sources of extrinsic motivation can have a powerful influence on teachers’ behavior. However, this impetus and its effect may be twofold. On the one hand, advocates of the system of holding stakeholders of education accountable for their work assume that “educators and their students will work harder and more effectively to enhance student learning when faced with large incentives and threatening punishments” (Finnigan & Gross, 2007). Numerous studies pointed out that – mainly in the USA – the introduction of test based accountability programs had an impact on students’ improvement (e.g. Jacob, 2005; Carnoy & Loeb, 2002; Hanushek & Raymond, 2005). In addition, data seem to underscore that the explicit incentivization had a more significant effect as opposed to implicit incentivization (Hanushek & Raymond, 2005; Dee & Jacob, 2011).

On the other hand, high-stakes testing may serve as an impediment to the development of intrinsic motivation as its introduction and implementation is generally accompanied by a high amount of pressure. Mounting evidence shows that pressure and high-stakes testing are closely intertwined and teachers are the most prone to pressure (Moore & Waltman, 2007). In line with this, Koretz
(2008) use the term high-stakes test to refer to tests which make teachers feel pressurized to increase students’ performance, even if the pressure mechanism is not realized in concrete rewards or sanctions. Teachers respond to accountability pressure in different ways. It may induce “fear of unwanted scrutiny, loss of flexibility in the classroom, pressure to teach to the test, and concern about jobs” in teachers (Fuller & Ladd, 2012), which in turn may threaten the validity of test scores, lead to corrupt educational practice (Nichols & Berliner, 2005), seriously compromise instructional practice (Pedulla, Abrams, Madaus, Russell, Ramos & Miao, 2003) and narrows down teachers’ attention to how well students score on tests leaving other needs unmet (Noddings, 2001, 2002). Cannell (1988) reported – as early as in the second half of the 80’s – that in many 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 NAEP. Studies trying to identify the factors that may account for this difference (see Jacob, 2007; Jacob & Levitt, 2003; Figlio & Gletzer, 2002) revealed that teaching to the test, reallocation or redistribution of resources and the use of forbidden tools explain this tendency (see Koretz, McCaffrey & Hamilton, 2001). Thus, the incentive schemes do not contribute to reaching the pre-defined aims of the principal in each and every case.

The scrutiny of the impact of two components – the Diagnostic Skills and Ability Assessment and the National Assessment of Basic Competencies – of the Hungarian educational accountability system on teachers’ responses to system-level testing forms the scope of our paper. Therefore, in what follows, the theoretical background will entail the brief comparison of these initiatives (Table 1).
Table 1. Characteristics of the Diagnostic Skills and Ability Assessment and the National Assessment of Basic Competencies

<table>
<thead>
<tr>
<th></th>
<th>DSAA</th>
<th>NABC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of assessment</td>
<td>criterion-referenced</td>
<td>norm-referenced</td>
</tr>
<tr>
<td>instrument</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objectives</td>
<td>diagnostic</td>
<td>diagnostic</td>
</tr>
<tr>
<td>Purpose</td>
<td>no stakes</td>
<td>high-stakes</td>
</tr>
<tr>
<td>Assessed domains</td>
<td>basic mathematical skills, reading skills</td>
<td>reading literacy, mathematical literacy</td>
</tr>
<tr>
<td></td>
<td>reasoning, writing skills</td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>4th graders</td>
<td>6th, 8th, 10th graders</td>
</tr>
<tr>
<td>Publicity of results</td>
<td>200 schools/data collection</td>
<td>nation-wide, school-level</td>
</tr>
<tr>
<td>Tests</td>
<td>20 test versions</td>
<td>anchor tests</td>
</tr>
<tr>
<td>Availability of framework</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>at school-level by means</td>
<td></td>
</tr>
<tr>
<td>Analysis of test results</td>
<td>of a special data analysis software</td>
<td>centralized</td>
</tr>
</tbody>
</table>

Both the names and the characteristics of the programs show the distinct purposes of the assessments. The diagnostic nature of DSAA excludes the opportunity for using its results for accountability purposes. Nevertheless, NABC is a diagnostic assessment testing students mandatorily at the end of the academic year. Besides a national report, school level and school provider level reports are also accessible to the general public and parents can reach the individual results of their children. Both the publicity of the results and the sanctions associated with the weak results of NABC also signify the accountability of the Hungarian education system. On the other hand, the DSAA is aiming at informing educators on students’ developmental level at the end of ISCED2 and providing guidelines for operating special development programs at individual and/or group level. It was compulsory for all students at the time of our survey. However, it did not serve any accountability purposes. A representative sample (200 schools) chosen from the entire 4th grade population were centrally processed. In addition, national standards were published to which schools could compare their results. However, as of 2013, the legal obligation of schools to carry out the DSAA, was terminated. Its instruments have been made publicly available online and schools process its results by themselves. The data collection of the NABC and the DSAA were carried out on the same day.

In the present study we are particularly interested in mapping out how centralized compulsory testing procedures shape teachers’ teaching practices. As the above description clearly shows the DSAA and NABC serves two distinct purposes. According to our hypothesis, the former, bearing a diagnostic, low-stakes nature carried out in lower elementary school years, should not evoke any pressure in
teachers, therefore, may not influence teachers’ behavior whereas NABC, as a high-stakes test administered in upper grades, may trigger patterns similar to those underpinned by international research. The following research questions were addressed to examine how teachers in different stages of formal education respond to centralized testing and test-based school accountability pressure.

1. Do system-level assessments evoke pressure in teachers in Hungary?
2. With which stakeholders can teacher pressure be associated with?
3. Is there a difference between teachers’ perceived pressure in lower and upper elementary schooling?
4. Does teachers’ perception of pressure have any impact on the instructional practice in different stages of formal education?

Methods

Sample, data collection

The study was carried out in a multi-method approach in two steps. We conducted the first empirical study in spring 2010. The sampling unit of our survey was the school (N=256), the sample was representative for region and settlement type. From every school two classroom teachers, one mathematics teacher and one Hungarian grammar and literature teacher filled in an online questionnaire anonymously. The second survey was conducted in spring 2012 in ten elementary schools in a culture bearing unit. In accordance with the target populations of the DSAA and NABC, one classroom teacher, one mathematics, one Hungarian grammar and literature teacher, and one school principal took part from each school in the second round of the survey.

Instruments

The survey consisted of two instruments: a questionnaire and a semi-structured interview. The questionnaire was based on numerous international surveys – the subscales were collected and linked in the International Project for the Study of Educational Accountability Project. At the adaptation we took into consideration the characteristics of the Hungarian education system. Three parts of the questionnaire was used in the present study. In accordance with thematic coverage, the first part measured teachers’ perceived pressure exerted by different types of assessments (2 items), the
second part investigated the sources of perceived pressure to increase test scores (8 items), and the third part aimed at gaining information about the changes in instruction due to DSAA/NABC (15 items). Teachers’ opinions were mainly assessed on a four level Likert scale (1=disagree; 4=agree a lot). Teachers were addressed the same questions about both the DSAA and the NABC.

The semi-structured interview with its subjects targeted the further interpretation and the resolution of the contradictions in the answers of the questionnaire. Content analysis was used to examine the interviews.

**Analysis and results**

The answers suggest that lower elementary school teachers (in other words: classroom teachers) perceive significant pressure due to the DSAA (M=2.90, SD=0.87) and upper elementary school teachers feel pressurized due to the NABC (M=2.90, SD=0.87) to a great extent. Although legally the NABC has a significantly higher stake than the DSAA, the answers of the lower elementary school teachers do not differ significantly from that of upper elementary school teachers (p>0.05). Results show that 23.1% of lower elementary school teachers and 24.1% of upper elementary school teachers feel pressurized to a great extent. On the other hand, 10.3% of respondents who teach in the first four years of elementary schooling and 12.1% of teachers active in further years of compulsory elementary schooling do not perceive accountability pressure at all.

Our questionnaire also sought answers on teachers’ views about the extent and nature of pressure different stakeholders exert in order to improve students’ NABC test scores. Factor analysis was carried out to identify sources of pressure (KMO=.818, Bartlett’s test=2152.16, explained variance=71.68). Three factors were extracted from the dataset: (1) direct sources of pressure at school-level (teachers themselves, colleagues, and administration), (2) indirect sources of pressure outside school (public, government, and municipality) and (3) the beneficiaries, the consumers of education and the direct utilizers of results (parents, students). Analyses were carried out in order to examine the impact of these stakeholders on teachers’ perceived pressure.
Teachers highlighted the pressure exerted by school administrators (Table 2). To translate this phenomenon into numbers: 40% felt extremely pressurized to improve students’ achievement. The school administrators represent an external source of pressure, however, the extent an internal source, that is teachers’ self-imposed expectations, generates pressure does not differ significantly (p>0.05). The second most pressurizing stakeholders are the municipalities. The regular control and assessment of teachers’ work is the legal obligation of the municipality. The data yielded by national system-level assessments are used in teachers’ evaluation; therefore, these results have an impact on their educational decisions. Typically, the municipality communicates with the school administration, teachers perceive the pressurizing force of the municipality through the school administration. Teachers also consider their colleagues and their expectations as sources of pressure. Possibly, having good school-level results is important for the whole teacher community and teachers impose pressure on each other as well. Answers show that teachers also perceive pressure from the public. 17.9 of classroom teachers and 27.5% of upper elementary school teachers reported on experiencing public pressure. The reasons for this may be that the results of the DSAA and NABC are major determinants in the shaping of public opinion and parents’ elementary school choice. At the time of our survey (2012) the municipalities of the schools were the municipalities and a centrally allocated sum based on the number of students formed an integral part of the school budget and this proportion was complemented by
the own resources of the municipality. The schools are obliged to accept the application of every single student residing in the schooling district. However, they can also enroll students belonging to other districts to fill the vacancies. The number of enrolled students plays a determinative role due to the above mentioned system of the allocation of financial resources. Nevertheless, this is also associated with the public opinion of the particular school. Even students residing in the schooling district may abandon schools considered to be weak which can lead to a further decrease in the number of students attending the school. In extreme circumstances this may lead to a dramatic fall in the headcount.

Our teacher interviews confirmed that teachers are overwhelmed by the role the results of the NBAC may play in the shaping of public opinion on schools. The following quote is used to best illustrate this phenomenon:

„Publicizing good school results can easily lead to considering the school as a hyper-super institution. Obviously this has consequences. A teacher teaching very clever students is labelled as a very good teacher, whereas who sweats blood with disadvantaged students – obviously with lot weaker results is labelled as a bad teacher or indulgent (upper elementary school teacher, Hungarian literature and grammar teacher)“.

Teachers feel the least pressurized by parents and students. This implies that parents and students are not particularly interested in the results of DSAA and NABC or that these stakeholders exhibit some shortcomings in advocacy. 12.2% of classroom teachers and 13.6% of upper elementary school teachers experience pressure by parents and almost the same proportion of teachers (12.6% and 13.1% subsequently) feel pressurized by students.

If we take a look at the differences in classroom teachers’ and upper elementary school teachers’ answers, we can see that significant difference is only exhibited in case of the pressurizing source attributed to the government. Lower elementary school teachers feel pressurized by the government to a greater extent than their upper elementary counterparts. These tendencies are surprising considering the fact that the assessments differ in many aspects (see Table 1). Results of the interview show that the extent teachers feel pressurized is influenced by the discourse at the schools.

Factor analysis was also used to indentify changes in teachers’ instructional practice as a result of system-level testing. Results (KMO=.783; Bartlett’s test=3350.81 p<.00; explained variance=62.8%) showed that our dataset is adequate to carry out the factor analysis. Five factors were extracted from the dataset regarding teachers’ response to these initiatives: (1) instructional changes, (2) homework, (3) alignment, (4) teaching to the test, (5)
reallocating teachers’ attention among outliers. In what follows further explanation and illustrative examples are provided for the better understanding of the factors.

1. More efficient teaching – Changes in the teaching contents and methods that contribute to students’ development; e.g. “I introduce more efficient teaching method.”, “I pay more attention to the training requirements.”

2. Homework – Allocating more time; e.g. “I give more homework assignments.”, “I give more difficult homework assignments.”

3. Reallocation of teachers’ attention among outliers – Teachers pay more attention to good and poor performing students during the teaching and learning process; e.g. “I put greater focus on poor performing students.”

4. Alignment – Giving priority to and laying more emphasis on the content of the assessments instead of the requirements of the curriculum; e.g. “I skip or spend less time on the contents that are not assessed by the centralized assessments.”

5. Teaching to the test / Practicing test-taking strategies – The teachers pay more attention and dedicate more time to practicing test-taking strategies during the teaching and learning process; e.g. “I focus on the multiple choice tests more during my work.”

Teachers put greater focus on teaching methods and teaching standards ($M_{\text{Factor}1}$=3.03 and 3.07), they focus their attention on students who are close to proficient or whose performance is weak ($M_{\text{Factor}5}$=2.82 and 2.77), and teach their students to the test ($M_{\text{Factor}4}$=2.54 and 2.44). They do not align the taught and measured contents because of the assessment. ($M_{\text{Factor}3}$=2.54 and 2.44) and do not assign more and more difficult homework ($M_{\text{Factor}2}$=2.54 and 2.44). Lower elementary school teachers are more likely to teach to the test ($t=2.13$ $p<0.03$), whereas teachers in upper elementary grades are more inclined to align taught and measured contents ($t=3.2$ $p<0.00$). During the interviews teachers reported that due to the assessment their focus shifted to domains that had not received that much attention during the teaching-learning process before. In nine schools out of ten teaching to the test prevails by means of the available tests. Results of the interviews conducted in upper elementary school grades highlight that harmonizing the framework of the NABC and the requirements of the curriculum poses challenges for teachers, which leads to inadequate practices.

We examined the relationship between the sources of pressure and the changes in teaching practices due to the assessment (Table 3). In lower elementary school grades data show a weak correlation ($p<.01$) between the factors of the sources and the factors of
changes in teaching in four cases. In upper elementary school grades data reveal that there is a significant relationship ($p<.01$) between all factors but one. The instructional changes, alignment and teaching to the test factors have stronger correlation with sources of pressure at school level than with the other two sources. The relationship of the reallocation of attention factor with beneficiaries suggests that assessment results encouraged parents to ask teachers to pay more attention to their poor or high performing child.

<table>
<thead>
<tr>
<th>Factors</th>
<th>DSAA School level</th>
<th>DSAA Outside school</th>
<th>DSAA Beneficiary</th>
<th>NABC School level</th>
<th>NABC Outside school</th>
<th>NABC Beneficiary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructional changes</td>
<td>.167</td>
<td>.155*</td>
<td>.125*</td>
<td>.314</td>
<td>.174</td>
<td>.190</td>
</tr>
<tr>
<td>Homework</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>.179</td>
<td>.195</td>
<td>.171</td>
</tr>
<tr>
<td>Alignment</td>
<td>n.s.</td>
<td>.108*</td>
<td>.109*</td>
<td>.242</td>
<td>.170</td>
<td>n.s.</td>
</tr>
<tr>
<td>Reallocation of attention</td>
<td>n.s.</td>
<td>n.s.</td>
<td>.145</td>
<td>.221</td>
<td>.163</td>
<td>.252</td>
</tr>
<tr>
<td>Teaching to the test</td>
<td>.177</td>
<td>.208</td>
<td>n.s.</td>
<td>.213</td>
<td>.155</td>
<td>.182</td>
</tr>
</tbody>
</table>

*Note: * $p<.05$, n.s. = non-significant, in the other cases $p<.01$

**Summary and conclusion**

In line with international trends, Hungary has also been developing its own national assessment system with no and high-stakes components facilitating educational accountability alike. Teacher pressure related to school accountability has come to the forefront although this is a relatively new field in Hungary.

The findings imply that the majority of teachers feel pressurized by system-level assessments – regardless of their high or no stakes purpose – at all levels of formal education. The main source of pressure related to testing is the school administration but teachers are also subject to self-imposed pressure to a great extent. With the exception of self-imposed pressure and government, there is no significant difference in the perceived pressure in lower and upper elementary grades, reflecting that any tests imposed on schools and teachers by an external authority may bring about unintended effects. However, in the case of DSAA the lack of relationship between changes in instructional practice and pressure suggests that the extent to which sources exert pressure does not inevitably draw along shifts in teachers’ practices. Nevertheless, in upper elementary school grades there is a relationship between school level sources of
pressure with changes in instructional practice, so teachers who feel pressurized by the leadership of the institutions are inclined to change their methods and overall work.

The results of the present study complement previous findings on teachers’ pressure and responses to system-level assessments. On the other hand, findings may pave the way to shaping contemporary perspectives on and the conceptualization of the one-to-one relationship between high-stakes-testing and perceived pressure. In addition, data also suggests that it is imperative to better inform teachers and principals about both the objectives of the assessment programs and the opportunities for utilizing their results.

This research was 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


