



# Teachers' beliefs about assessment and accountability

Edit Tóth<sup>1</sup> · Benő Csapó<sup>1</sup>

Received: 11 March 2021 / Accepted: 10 July 2022 / Published online: 17 August 2022  
© The Author(s) 2022

## Abstract

One of the main aims of national assessment programmes is to improve the efficacy of education systems; realizing this aim often takes the form of implementing a variety of accountability measures. Using assessment results for accountability purposes is highly controversial, while one of its undesirable impacts is that it generates negative attitudes towards educational assessments among teachers. The aim of this study is to examine lower and upper secondary teachers' ( $N=1552$ ) opinions and beliefs about testing and, more specifically, about the national assessment programme in Hungary. A questionnaire was used to explore teachers' beliefs about the effects of the assessment system on how they teach, perceived pressure from stakeholders, teachers' acceptance of assessment programmes and the relationship between these beliefs. Results show that assessment programmes compel teachers to revise their teaching practices — some change to make meaningful gains in student learning, while others turn to practices that are not conducive to a genuine improvement in students' knowledge, focussing instead on assessment scores. Pressure from inside the school (colleagues and school leaders) and teachers' attitude towards assessments bring about changes in instruction, such as the reallocation of coaching and improvement in teaching. Sources of pressure outside school (local government and the media) have an indirect effect on changes in teaching because their pressure influences in-school motivators. Pressure from parents and students is felt directly by teachers, but only in limited areas. The results demonstrate that a national assessment programme has a more significant impact on teaching in lower than in upper secondary schools.

**Keywords** Teachers' beliefs · Test-based accountability · Source of pressure · Teaching practices

---

✉ Edit Tóth  
tothedit@edpsy.u-szeged.hu

<sup>1</sup> Institute of Education, University of Szeged, 30-34. Petőfi S. sgt, Szeged 6722, Hungary

## 1 Introduction

A large number of studies have highlighted the effects of national assessment systems on students' achievement (e.g. Dee & Jacob, 2011) and on teaching and learning processes. These studies indicate their positive and negative effects on teaching and learning practices (Au, 2007; Darling-Hammond & Rustique-Forrester, 2005; Jones, 2007; Popham, 2001; Stecher et al., 2000; Williams, 2009) and draw attention to the fact that these practices are greatly influenced by the consequences of assessment results on schools, teachers and students (Carnoy & Loeb, 2002; Certo, 2006; Hanushek & Raymond, 2002, 2005). Since these assessments are ubiquitous and produce a number of consequences, it can be assumed that teachers have a wide range of beliefs about assessment and assessment programmes and that these beliefs, in connection with others, have a significant impact on their teaching practices (Barnes et al., 2015).

In addition, research has focussed on the way teachers perceive pressure from stakeholders, such as parents, government and school leaders (Firestone et al., 2004; Moore & Waltman, 2007), and it has highlighted that they are greatly influenced by the nature of the accountability mechanism (Pedulla et al., 2003). However, it is not fully understood how teachers perceive the effects of interested parties (or stakeholders) on their teaching practices.

In this article, we explore how teachers change their teaching practice due to the national assessment programme and its accountability aspects, the perceived pressure from stakeholders and their effect on instructional practices. We explore these phenomena in a special accountability context — Hungary (Balázs & Ostorics, 2020) — where school-level reports are public, low-performing schools are sanctioned, and there is free school choice at the ISCED 2–3 level, which means the results of the national assessment could influence families' school choice-related decisions.

The present paper is structured as follows: first, we provide an overview of the objectives and associated stakes of a national assessment programme as part of accountability systems. Then, we review the literature on the empirically proven impact and perceived impact of accountability systems on teachers' instructional practices and the extent of pressure teachers perceive due to different — high- and low-stakes — accountability programmes. We then describe the main results of a questionnaire that explores Hungarian primary and secondary teachers' beliefs about changes in how they teach and perceived pressure from different sources. Finally, we highlight the effect of teachers' perceptions of this pressure on their instructional practices.

## 2 Theoretical background

By measuring educational outcomes, test-based accountability programmes primarily aim to improve students' achievement through a number of mechanisms — information, incentives and assistance — that bring about changes in classroom

practices (Hamilton et al., 2005). Furthermore, in a number of countries, these programmes have become a tool for measuring educational efficacy and for holding schools and teachers accountable for their students' performance. The stakes associated with the output can vary widely, so these assessment programmes increase pressure on teachers and principals to act to change teaching practices. Teachers' perceptions of different aspects of large-scale assessments influence their responses to reforms/programmes that aim to change instruction.

The assessment literature examines many aspects of teachers' beliefs about assessment programmes, especially beliefs about the purposes of assessment (i.e. Barnes et al., 2017; Brown, 2004, 2006; Brown & Harris, 2009; Önalın & Karagül, 2018), the value of large-scale assessments (Abrams et al., 2003), the use of assessment data (Copp, 2016) and the relationship between beliefs about classroom assessment and assessment programmes for accountability purposes (Brown & Harris, 2009). The literature also compares beliefs about classroom and large-scale tests from different perspectives (Leighton et al., 2010).

## 2.1 The impact of accountability programmes on instructional practices

The practice of high- and low-stakes testing goes back several decades in Anglophone countries. A great body of experience and research has shown that schools and other stakeholders respond to state-level accountability programmes in diverse ways. However, consistent patterns take form regarding the content and methods of teaching regardless of the accountability scheme used (e.g. Herman, 2004, 2008; Koretz et al., 2001; Stecher, 2002).

Based on a cross-check of the results of different test-based accountability programmes, numerous studies have pointed out that their introduction has had a positive impact on students' test scores (e.g. Carnoy & Loeb, 2002; Jacob, 2005; Linn & Dunbar, 1990; Nichols et al., 2012). Koretz and Hamilton (2003) claim that improvement in students' test scores on high-stakes tests is closely related to teachers' responses to the testing programmes manifested in the learning and instruction process. 'To the extent that teachers respond with approaches that bolster students' mastery of the domains about which users draw inferences, increases in scores will warrant inferences about improved student achievement' (Koretz & Hamilton, 2003, p. 3).

According to Firestone et al. (2002), advocates of test-based accountability systems can be divided into two groups. Representatives of the first group assume that the consequences of test results motivate teachers to work harder and more effectively to facilitate students' improvement. The other group attributes great importance to the assessment programmes themselves. In their view, assessments drive teachers' and principals' attention to teaching content, standards and versatile assessment methods and forms, thus shaping classroom processes and outputs. No matter which group they belong to, all proponents of accountability policies believe in the integration of assessment and instructional practice. In other words, they claim that 'what gets tested gets taught' (e.g. Resnick & Resnick,

1992). While this phrasing carries the promise of guaranteeing access to knowledge for students, in practice, this idea may have an adverse effect. For example, in an analysis of 138 standard assessments conducted since the introduction of the No Child Left Behind Act in the USA, Polikoff et al. (2011) found that only half of the standard content is included in the corresponding tests. Moreover, only half of the test content was covered by the standards. That is, when tests cover less than the standards, the implemented curriculum also becomes more restricted. Brown (2018) highlights that aligning assessment with the curriculum makes it possible for assessment to be ‘a guide to classroom instruction and learning’ (Brown, 2018, p. 15).

‘However, research suggests that the effects of testing have mostly failed to translate to fundamental changes in teachers’ pedagogy’ (Hamilton et al., 2013, p. 457). According to McNeil (2000), accountability policies and standardized testing ‘reduce the quality and quantity of what is learned in schools’ due to the pressure they bring to bear (p. 230). Accountability critics have consistently warned that accountability schemes may have an adverse impact on teaching content. For example, as early as the 1990s, Koretz et al. (1996) conducted a study in Kentucky which suggested that measures may lead to a narrowing of the curriculum and to the underrepresentation of non-assessed subjects, contents and domains. Asking teachers and principals to examine the effects of the standard-based reform introduced in Washington State at the end of the 1990s, Stecher et al. (2000) also showed that educators, by their own account, paid more attention to the tested content and test formats than to the represented standards of the test.

Nevertheless, Au’s (2007) metasynthesis of 49 qualitative studies showed that the impact of high-stakes tests on the curriculum is ambivalent. According to the study, the primary impact of high-stakes testing is manifested through narrowing the curricular content down to the tested domains. Teaching the tested content receives high priority, and teachers shift towards a teacher-centred pedagogy. However, he also claims that ‘certain types of high-stakes tests have led to curricular context expansion, the integration of knowledge, and more student-centred, cooperative pedagogies’ (Au, 2007, p. 258). Nevertheless, the author emphasizes that ‘the nature of high-stakes test-induced curricular control is highly dependent on the structures of the tests themselves’ (Au, 2007, p. 259).

Furthermore, Williams (2009) found that linking teachers’ work to high-stakes tests that yield data which make accountability possible has resulted in discredited teaching methods and non-contextualized engagement with course content. Studies conducted by Jones (2007) and Pedulla et al. (2003) found that in high-stakes testing programmes many teachers spend a great deal of instructional time practising test-taking strategies, especially in low-performing schools. In addition, Jones’ (2007) study showed that teachers might have focussed on lower-level basic skills in their practice when they did not know or accept the aims of the testing programme.

Koretz et al. (2001) suggest that in terms of their effects on the validity of gains, teachers’ responses to testing can be grouped into seven categories of test preparation: (1) teaching more, (2) working harder, (3) working more effectively, (4) reallocation, (5) alignment, (6) coaching and (7) cheating. Teaching more implies spending more time on instructional activities (e.g. as opposed to carrying out

administrative tasks in class), and working harder translates into covering more material during class, for example. Working more effectively means using more effective methods and paying more attention to the quality of the method adopted. Reallocation refers to the regrouping of educational resources, such as teacher's teaching time and content, to increase students' achievement. Alignment is a special form of reallocation; curricula and tests are aligned, meaning that certain content is assigned a more prominent role in the teaching and learning process than other content. Coaching means that educational resources are used to focus on practising for certain aspects of the test. This method is also referred to as teaching to the test or teaching the test. Generally, the latter is often interpreted as a form of cheating. The first three of these responses are likely to produce unambiguously meaningful gains in scores. Reallocation, alignment and coaching may increase test scores 'without similarly increasing the achievement that the scores are intended to represent' (Koretz & Hamilton, 2003, p. 3). Cheating, however, definitely does not lead to a genuine and inherent change in students' knowledge or skills in the domain being measured.

Recent analyses further explore the impact of test-based accountability. Smith and Holloway (2020) used the data from the Teaching and Learning International Survey (TALIS) to examine the relationship between the general testing culture, the role of test scores in appraisals and satisfaction among teachers. They have found that overemphasizing the significance of test scores may reduce the positive effects of appraisals on teacher satisfaction. Similarly, when systematically reviewing 36 studies, Tuytens et al. (2020) remark 'we would warrant caution in concluding that teacher evaluation should rely on test scores too heavily' (p. 77).

## 2.2 Perceived pressure induced by assessment stakes

A number of studies have examined the impact of accountability programmes on teachers. The research questions have mainly addressed issues of the extent to which teachers feel pressured due to accountability programmes and how this perceived pressure influences their motivation, exhaustion (e.g. Certo, 2006; Cuevas et al., 2018; Finnigan & Gross, 2007; Herppich & Wittwer, 2018), level of stress (e.g. Saeki et al., 2018) and classroom instruction (e.g. Condliffe & Plank, 2013; Firestone et al., 2004; Pedulla et al., 2003). Most of the studies indicate that high stakes contribute to negative effects in instructional practice and the high amount of pressure can serve as an impediment to adopting effective teaching strategies. It may induce teachers' anxiety about unwanted intrusion, loss of flexibility in classroom practices, a feeling of coercion to teach to the test and fear for their jobs (Fuller & Ladd, 2012). On the premise that teachers feel pressured by the introduction of state-level assessment systems in the context of school, Pedulla et al. (2003) showed that the number of teachers experiencing pressure depends on the stakeholders and the nature of the testing environment.

Firestone et al. (2004) examined the effects of a test-based accountability system among mathematics and science teachers in New Jersey. They suggest that teachers perceive significant pressure from the formal hierarchy, principals and central office

staff. ‘Sometimes it comes directly through exhortation to do better; sometimes it is indirect. Pressure related to publicity and the formal hierarchy is not consistently supported from other directions, however. Some parents do care about their children’s test scores, but many have a broader view of their children’s welfare’ (Firestone et al., 2004, p. 87).

Pedulla et al. (2003) asked teachers working in different accountability systems about the perceived pressure they associated with mandatory state tests and how the perceived stress influenced their teaching practices and profession. In high-stakes environments, a significantly higher percentage of teachers indicated that they feel more pressured by the superintendent and school principal than teachers operating in a low-stakes testing environment. The same percentage of teachers experienced significant pressure from parents both in low- and high-stakes environments. This result suggests that the perception of parental pressure is independent of legal stakes. This outcome supports findings by Moore and Waltman (2007) that teachers greatly fear pressure from the government and school administration — and that parents and colleagues are rarely seen as sources of pressure. In all the testing environments examined, teachers’ instructional practice changed. Sometimes numerous teaching methods were adopted (e.g. cooperative teaching and individual work), especially in a high-stakes environment. However, pressure is more likely to lead to score inflation because teachers may feel compelled to use methods they had not considered acceptable before testing was introduced.

Pedulla et al. (2003) conclude that it is a challenge to find the satisfactory combination of incentives — stakes — that forces teachers to work more effectively but does not result in unexpected negative instructional practices.

In their 2-year investigation of 23 classrooms, Condliffe and Plank (2013) found that classroom quality is lower when classrooms are under great pressure to enhance students’ achievement. According to the researchers, teachers worry so much about the results of their students’ basic literacy and numeracy skills on high-stakes tests that higher-order problem-solving or maths problems are pushed into the background in classroom instruction. Test preparation activities are prioritized in the classroom at the expense of activities that foster teacher–student feedback interactions or concept development.

### 2.3 The context of the study

The proportion of pupils spending at least a year in kindergarten (pre-school) is relatively high in Hungary (with 95–98% of all children attending kindergarten at the age of 5 between 2001 and 2018). Students officially begin primary school (ISCED 1, Grades 1 to 4) at the age of six; however, parents prefer to let their children start school later. Therefore, the actual age that children enter school is higher. (However, stricter new legal measures will control this.) After 4 years of primary school (with class teachers), they continue their education in the lower secondary phase (ISCED 2, Grades 5 to 8, with subject teachers). Upper secondary education (ISCED 3, usually covering Grades 9 to 12/13) takes place in grammar schools, vocational secondary schools and vocational schools (Juhász et al., 2010). Assessment programmes

provide feedback for stakeholders at three levels. (1) Hungary participates in large-scale international assessment programmes (PIRLS, TIMSS and PISA) which provide system-level feedback. (2) The declared purpose of the annual student assessment system, Hungary's National Assessment of Basic Competencies (NABC), is to support schools in improving their efficacy. (3) An online diagnostic assessment system (the eDia) has been developed for the first six grades in reading, science and mathematics. The system has been fully functional since 2015, and schools may join the programme voluntarily. By 2019, around one-third of primary schools regularly utilize the online system (Csapó & Molnár, 2019). As in many European countries, there is a traditional school-leaving examination (Matura) at the end of upper secondary education for certain students in grammar school and vocational secondary school. The study presented in this paper is most closely related to the second level (2). As the NABC is mandatory, it is administered yearly, and it is only this assessment that is tied to accountability measures.

Hungary has been developing its own national assessment system (Balázsi & Ostorics, 2020), the NABC, since 2001. The main objective of assessing students with the NABC in Grades 6, 8 and 10 is to provide feedback for schools and other stakeholders. The NABC measures reading and mathematical literacy. Development of its framework has been influenced by that of PISA (Cresswell, 2016); it intends to measure similar competencies (applicable knowledge). The majority of teachers may contribute to the development of these competencies as there are opportunities for continuous improvement in most school subjects.

Since 2008, data for all students has been centrally processed (following PISA scaling and data analysis methods). In addition, due to a secret measurement ID (generated via a sophisticated cryptographic process), it has been possible to connect consecutive assessments at student level. The database is thus suitable for longitudinal analyses. Results are published in national, school-level and school provider-level reports, which are accessible to the general public at the Educational Authority homepage. Furthermore, parents can access the individual results of their children. In addition to detailed feedback, a school-level database is also available to schools for further analyses. Schools have to prepare an action plan if the results from the NABC show that half of the students in their institution perform below the minimum criterion set forth by law. The action plan must be approved by the particular local authority, which is responsible for financing local elementary schools, overseeing whether these schools are operated in accordance with the legislation, and frequently assessing teachers' work with instruments, such as the NABC. Both the publication of the results and the sanctions associated with weak results on the NABC make accountability possible in the Hungarian education system. This exposure counts as a high-stakes motivational form because in Hungary, due to free school choice, published assessment data could have a remarkable effect on schools as they may influence the number of students enrolled in the institution and its income as well. At the time of the research presented here, the vast majority of the schools' budget came from centralized support calculated on the basis of the number of enrolled students, which was supplemented by the school provider/local government's own financial resources. Because of this financial support, the number of enrolled students is a key issue for schools.

### 3 Aims of the present study and research questions

In this study, we explore the effect of the NABC on the teaching and learning process, teaching methods and teaching content. For this purpose, we administered a questionnaire that focuses on teachers' beliefs about three different aspects: (1) the acceptance and usefulness of large-scale school assessment programmes, (2) the effect of the NABC on teaching and (3) perceived pressure from stakeholders at the lower and upper secondary levels. Although the context of the study seems specific to Hungary, it has many areas in common with international developments; therefore, the research questions are conceptualized so that the analyses may result in generalizable findings.

In the present study, we seek to answer the following research questions:

1. What beliefs do teachers have about the usefulness and reliability of large-scale educational assessments?
2. What stakeholder pressure do teachers perceive to increase their pupils' achievement on the NABC?
3. Do teachers report changes in their instructional behaviour, and, if so, do they attribute these to the effects of the NABC?
4. How do pressure from stakeholders interested in the efficacy of the education system and acceptance of large-scale educational assessments together affect the perceived change in teachers' teaching practices?

## 4 Methods

### 4.1 Participants

Data was analysed from a total of 1552 participants. Seven hundred twenty-six Hungarian lower secondary teachers responded to our survey from 256 schools (ISCED 2, the percentage of females in the sample is 86.4%; this percentage is 87.5% in the teacher population), and 826 upper secondary teachers took part in our survey from 97 schools (ISCED 3) teaching in Grades 9 to 12/13 (the percentage of females in the sample is 76.8%; this percentage is 76.2% in the teacher population). The two subsamples were selected on the basis of the regional distribution of schools; at the upper secondary level, the school type was taken into consideration as well. 36.54% of teachers teach in grammar school programmes, 44.12% in secondary technical schools and 19.33% in vocational schools (the proportions in the population are 37.9%, 40.8% and 21.3%, respectively). We collected data from three teachers in Grades 5 to 8 (ISCED 2) from every elementary school, and the questionnaire was completed by nine teachers from every upper secondary school (ISCED 3). The selected teachers teach mathematics ( $N_{\text{Lower}} = 244$ ,  $N_{\text{Upper}} = 276$ ), Hungarian language and literature ( $N_{\text{Lower}} = 245$ ,  $N_{\text{Upper}} = 278$ ), and one or more subjects in the natural sciences ( $N_{\text{Lower}} = 237$ ,  $N_{\text{Upper}} = 272$ ).



## 4.2 Instruments

This study forms part of a larger project which has aimed to explore teachers' opinions and attitudes towards different levels and types of assessment; each block of the questionnaire used in the study represents a particular assessment or accountability procedure. Teachers' opinions were assessed on a four-point Likert scale (1 = disagree; 4 = agree). The questionnaire elicited information on the following themes.

- (1) *Changes in instructional practice* (17 items): Changes in teacher's teaching practices due to the national assessment system were assessed with a subscale based on Hamilton et al. (2005). Following the classification developed by Koretz et al. (2001), the first six statements represent efforts towards better and more instruction. The others express reallocation of time or resources across topics (7–11 items), activities (11–14 items) or students (15–17 items) within a subject — these practices are not necessarily harmful, but they have the potential to lead to a narrowing of instruction and thereby inflated test scores (Koretz, 2002). Internal consistency was acceptable (Cronbach's  $\alpha = 0.85$ ).
- (2) *Perceived pressure from different stakeholders* (8 items): The extent to which teachers feel pressured by different stakeholders to increase students' achievement on the NABC was measured with a subscale adapted from the Administrator Questionnaire Iowa I Survey (e.g. Moore & Waltman, 2007). The list of stakeholders was expanded with two additional elements: the municipality (the government was divided into local- and state-level agents) and students. Internal consistency was acceptable (Cronbach's  $\alpha = 0.86$ ).
- (3) *View of large-scale educational assessments* (8 items): Teachers' view of large-scale tests was assessed using a relevant adapted subscale from the Teacher Survey of the German VERA 2007 assessment. This subscale assesses teachers' acceptance of large-scale national and international assessment programmes in general and their beliefs on their usefulness. Scale reliability was high (Cronbach's  $\alpha = 0.88$ ).

## 4.3 Procedures

Data were collected online individually and anonymously. Participation in the survey was voluntary. An email was sent to every contact person in selected schools, in which they were asked to choose five teachers according to given criteria and to pass on to them the URL address for the questionnaire and a previously generated account. It took 35–40 min to complete the questionnaire.

## 4.4 Analyses

In this paper, we describe teachers' beliefs about the effects of the national assessment programme on their teaching practices, perceived pressure from stakeholders and teachers' acceptance of assessment programmes and analyse differences

between lower and upper secondary teachers to establish mean score differences. The questionnaire included negatively formulated items; in these cases, scores have been recoded (so high scores indicate a positive view). The independent-samples *T*-test was used to establish statistical differences between groups.

Confirmatory factor analysis (CFA) was used to examine the existence of dimensions describing (1) changes in teaching practices, (2) sources of pressure and (3) acceptance of large-scale assessments. Structural equation modelling was used to analyse the relationships between the pressure sources, the perceived change in teacher's instructional practices and acceptance of large-scale educational assessments. To identify whether teachers' perceptions of changes in instructional practice and sources of pressure were statistically equivalent within different levels of education, we examined measurement invariance (Cheung & Rensvold, 2002): the configural equivalence (the basic model structure is invariant across groups), the metric factorial invariance (factor loadings are constrained to be equal across groups) and strong factorial invariance (all intercepts of item loadings are constrained to be equal across groups) (Hu & Bentler, 1999; Wu et al., 2007).

## 5 Results

### 5.1 The structure of teachers' perceptions of changes in instruction, sources of pressure and acceptance of large-scale educational assessments

Following the literature, five dimensions of changes in instruction were defined due to assessment programmes: (1) better instruction, (2) homework, (3) reallocation of content resources, (4) reallocation of teacher's attention and (5) test-taking strategies. Sources of pressure were organized into three groups: (1) school-level (internal) pressure sources (school management and colleagues) — these professional stakeholders have a direct effect on teachers' teaching; (2) external sources of pressure (the local authority, government and general public) — they have an indirect effect on teachers, mainly through school management; and (3) beneficiaries (parents and students). Furthermore, we examined the dimension of acceptance of large-scale educational assessments. We conducted three CFAs to test the structure of the three scales under examination.

In the first stage of measurement model development, one poorly fitting item was identified (error variance > 0.80, factor loadings < 0.50 and relatively high modification indexes) in the large-scale educational assessment dimension. The item was eliminated based on theoretical and logical considerations. Each of the final measurement models tested in stage 1 had an acceptable to excellent goodness of fit (Table 1).

The lower and upper secondary teachers' perceptions of changes in instruction were arranged in five dimensions. We examined whether the model of changes in instruction is statistically equivalent in the subgroups. We can state that the model is configurally invariant (RMSEA = 0.013). The regression weights from the factors to the items were equivalent ( $\Delta$ CFI = 0.001), as were the factor intercepts

**Table 1** Goodness-of-fit indices of measurement models

Model	School level	$\chi^2$	$p$	$df$	$\chi^2/df$	CFI	TLI	RMSEA	SRMR
Acceptance of large-scale educational assessments factor	Lower	36.37	<.001	9	4.0	.97	.94	.05	.03
	Upper	27.15	<.001	9	3.0	.99	.98	.04	.02
Changes in instruction factors	Lower	54.50	<.001	16	3.4	.97	.93	.05	.04
	Upper	76.67	<.001	16	4.8	.96	.93	.05	.03
Pressure factors	Lower	156.94	<.001	67	2.3	.95	.93	.05	.04
	Upper	173.82	<.001	67	2.6	.96	.95	.05	.04

( $\Delta CFI=0.190$ ). The model differences represent real differences in lower secondary and upper secondary teachers' perceptions.

## 5.2 Group differences for the changes in instruction dimension

In Table 2, we summarize the descriptive statistics and the results of group differences for the examined effects of the national assessment system on the teaching process; the statements are arranged by dimension. The results show that the NABC has an impact on teaching practices and a major effect on teaching methods. In addition, teachers claim that the assessment programmes have brought the curriculum and the assessed competencies into the forefront. The assessments also highlight the necessity of supporting and developing poor performers. Data do not support the possibility that teachers modify the topics and content of their teaching because of the assessment programmes. Teachers pay more attention to practising general test-taking strategies and the test, and item formats used on large-scale educational assessments also appear in everyday teaching practices. For five variables, there is no significant difference in the means for the two subsamples, and three of these statements/variables fall within the reallocation of content resources dimension. For the other variables, the means are higher for lower secondary teachers than for upper secondary teachers.

## 5.3 Group differences for sources of pressure components

Teachers' views about the degree of pressure they feel from different stakeholders to improve students' NABC test scores were analysed, highlighting the differences between lower and upper secondary teachers' answers (see Table 3). According to the responses, the stakeholders that exercise the most pressure are school administrators. Teachers highlighted pressure from the local authority. To translate this phenomenon into numbers, 43.4% of lower secondary teachers and 37.5% of upper secondary teachers felt extremely pressured by their school administration to improve students' achievement, and 33.8% and 27.2% of them, respectively, felt this same pressure from the local authority. Teachers in each subgroup feel the most moderate pressure from parents and students. Lower secondary teachers perceive higher levels

**Table 2** Means, standard deviations and group differences for changes in instruction variables

Items	Lower secondary <i>M (SD)</i>	Upper secondary <i>M (SD)</i>	<i>t</i>	<i>p</i>
<b>Better instruction</b>				
I search for more effective teaching methods	3.23 (.77)	3.00 (0.83)	5.81	.001
I focus more strongly on curriculum requirements	3.04 (.82)	2.93 (0.86)	2.53	.001
I focus more strongly on key competencies	2.72 (.88)	2.57 (0.89)	2.52	.001
I focus more strongly on the competencies that are assessed as part of the educational concept in large-scale assessment programmes	3.17 (0.74)	2.85 (0.83)	8.16	.001
<b>Homework</b>				
I assign more homework	1.47 (0.72)	1.56 (0.74)	-2.27	.002
I assign more difficult homework	1.41 (0.60)	1.42 (0.65)	n.s.	
<b>Reallocation of content resources</b>				
I reduce instructional content	1.99 (0.80)	1.95 (0.90)	n.s.	
I take fewer liberties in how I design the content of my lessons	1.82 (0.81)	1.88 (0.84)	n.s.	
I set aside or spend less time on content which will not be tested	1.58 (0.71)	1.66 (0.77)	1.96	.005
I have narrowed down the curricular content of my instruction	1.85 (0.87)	1.84 (0.87)	n.s.	
My instruction focuses more strongly on competencies rather than content	1.93 (0.83)	1.89 (0.81)	4.07	.001
<b>Reallocation of teacher's attention among students</b>				
I focus more strongly on low achievers	2.86 (0.86)	2.67 (0.88)	4.21	.001
I focus more strongly on high achievers	2.71 (0.91)	2.44 (0.91)	5.83	.001
I offer more assistance outside classroom instruction	2.70 (0.96)	2.42 (0.94)	5.74	.001
<b>Test-taking strategies</b>				
I focus more strongly on multiple-choice tests	2.38 (0.91)	2.12 (0.86)	5.60	.001
I focus more strongly on tests with open-ended questions	2.31 (0.88)	2.31 (0.95)	n.s.	
I spend more time teaching general test-taking strategies	2.47 (0.89)	2.25 (0.93)	4.77	.001

*n.s.*, non-significant

**Table 3** Means, standard deviations and group differences for sources of pressure variables

Items	Lower secondary <i>M (SD)</i>	Upper secondary <i>M (SD)</i>	<i>t</i>	<i>p</i>
School-level pressure sources				
Yourself	3.31 (0.77)	2.95 (0.87)	8.75	.001
Colleagues	2.78 (0.85)	2.52 (0.86)	5.96	.001
School management	3.36 (0.76)	3.13 (0.86)	4.95	.001
External sources of pressure				
Local authority	3.05 (0.93)	2.74 (0.98)	6.22	.001
Government	2.59 (1.08)	2.24 (0.99)	5.89	.001
General public	2.66 (0.98)	2.27 (0.99)	7.76	.001
Beneficiaries				
Parents	2.39 (0.98)	1.99 (0.94)	8.29	.001
Students	2.29 (0.99)	2.01 (0.98)	5.06	.001

of pressure from all the stakeholders than their colleagues at the upper secondary level.

#### 5.4 Group differences for teachers' acceptance on large-scale assessment variables

Having examined acceptance of large-scale educational assessments (Table 4), we can state that teachers acknowledge the importance of large-scale educational assessments and their impact on their work. Upper secondary teachers accept large-scale educational assessments better than their colleagues in lower secondary schools. However, we can identify a group of teachers who consider assessments as a source of trouble and feel that they produce more drawbacks than solutions.

**Table 4** Means, standard deviations and group differences for teachers' acceptance on large-scale educational assessments variables

Items	Lower secondary <i>M (SD)</i>	Upper secondary <i>M (SD)</i>	<i>t</i>	<i>p</i>
Large-scale educational assessments...				
...should be conducted on a regular basis	3.02 (0.98)	3.10 (0.92)	n.s	
...are important for work in schools	2.85 (0.97)	2.89 (0.92)	n.s	
...only cause trouble in schools	2.19 (0.99)	1.93 (0.96)	5.16	.001
...create more problems than solutions	2.21 (0.99)	2.02 (0.94)	4.01	.001
...contribute to an increased effort in schools	2.79 (0.91)	2.71 (0.88)	n.s	n.s
...provide an objective basis to evaluate schools/school systems	2.16 (1.01)	2.45 (0.97)	-5.78	.001
...are not useful for my job as a teacher	2.11 (0.98)	2.17 (0.97)	n.s	

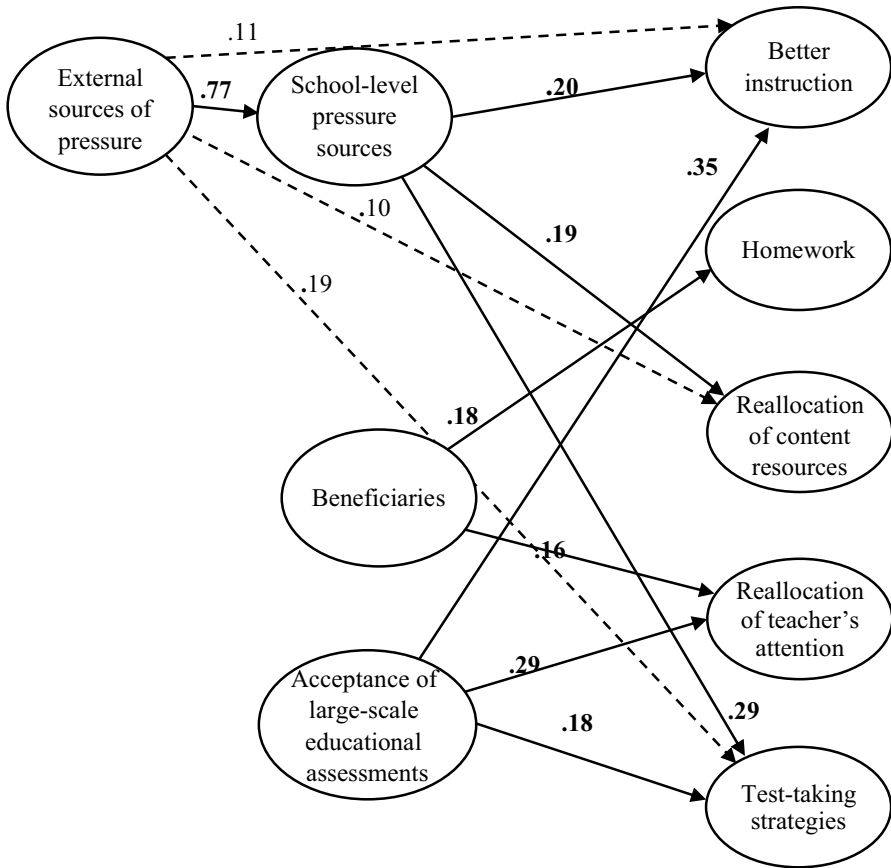
n.s. = non-significant

## 5.5 Relations between sources of pressure, acceptance of large-scale educational assessments and teaching practices

We analysed how perceived pressure from stakeholders and teacher's attitude towards large-scale surveys predicted changes in teaching practices. Multiple regression analysis was used for each of the descriptive factors of instructional practices as dependent variables along with the three pressure factors and the factor of acceptance of large-scale educational assessments as predictors. In both models, the regression analysis showed that better instruction was predicted by the school-level pressure sources ( $\beta_{\text{Lower\_secondary}}=0.33$ ,  $\beta_{\text{Upper\_secondary}}=0.50$ ) and acceptance of large-scale educational assessments factors ( $\beta_{\text{L}}=0.23$ ,  $\beta_{\text{U}}=0.29$ ), while homework was only predicted by the beneficiaries factor ( $\beta_{\text{L}}=0.14$ ,  $\beta_{\text{U}}=0.32$ ). Reallocation of content resources was predicted by the school-level pressure sources factor ( $\beta_{\text{L}}=0.25$ ,  $\beta_{\text{U}}=0.34$ ). Reallocation of teacher's attention was predicted by the school-level pressure sources factor ( $\beta_{\text{L}}=0.18$ ,  $\beta_{\text{U}}=0.31$ ) and by the beneficiaries factor ( $\beta_{\text{L}}=0.17$ ,  $\beta_{\text{U}}=0.14$ ). The school-level pressure sources ( $\beta_{\text{L}}=0.28$ ,  $\beta_{\text{U}}=0.54$ ) and acceptance of large-scale educational assessments factors ( $\beta_{\text{L}}=0.16$ ,  $\beta_{\text{U}}=0.16$ ) represented predictors for test-taking strategies. The external sources of pressure factor had no direct effect on any of the five change factors.

Based on the regression analyses, subsamples were analysed for relations (direct and indirect connections) between perceived pressure from stakeholders, acceptance of large-scale educational assessment and changes in teaching practices using a structural equation model. We managed to build well-fitting models to describe the change in teaching for the lower (Fig. 1) and upper (Fig. 2) secondary teachers' sample. The models had reasonably acceptable fit indices (lower secondary level:  $\chi^2(300, N=720)=534.92$ ,  $\chi^2/df=1.78$ ,  $p<0.001$ , TLI=0.93, CFI=0.92, RMSEA=0.04, SRMR=0.04; upper secondary level:  $\chi^2(300, N=826)=655.74$ ,  $\chi^2/df=2.19$ ,  $p<0.001$ , TLI=0.94, CFI=0.93, RMSEA=0.04, SRMR=0.04). Broken lines indicate the indirect effect of external sources of pressure. Individual items are the manifest variables.

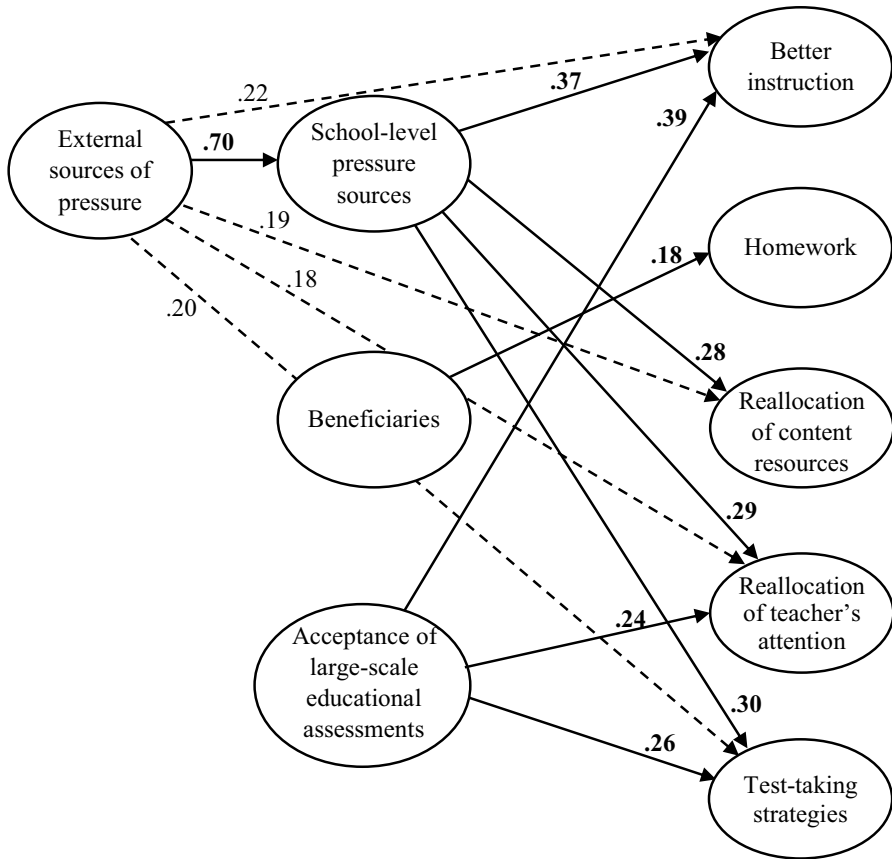
At both school levels under examination, the external sources of pressure factor have a strong effect on school-level pressure sources but have no direct effect on the changes in instruction factors. External sources of pressure have an indirect effect on several factors that indicate changes in instruction through the school-level pressure sources factor. The influence of the stakeholders is different on the changes in instruction factors at the two school levels. The school-level pressure sources factor was positively associated with test-taking strategies, better instruction and reallocation of content resources at both school levels but not with the reallocation of teacher's attention factor at the upper secondary level. The effect of internal pressure sources is more significant at the upper secondary level than at the lower secondary level. The beneficiaries factor (parents and students) influences the amount and difficulty of homework at both school levels, and also has an effect on reallocation of teacher's attention at the lower secondary level. Acceptance of large-scale educational assessments was associated with the better instruction, reallocation of teacher's attention and test-taking strategies factors at both school levels.



**Fig. 1** Model for lower secondary teachers' beliefs about changes in instruction and perceived pressure from various stakeholders

## 6 Discussion

This article has presented the case of the NABC in an attempt to understand and model teachers' beliefs about acceptance of testing and accountability and to understand the underlying mechanisms between teachers' perceived pressure induced by different stakeholders and the impact of pressure on the teaching process at the lower and upper secondary levels. Using confirmatory factor analyses, evidence was found that five dimensions of instructional practices (better instruction, more homework, reallocation of resources and topics, reallocation of teacher's attention among students and test-taking strategies) could be used to describe how teachers changed their practices — as suggested by Koretz et al. (2001). Stakeholders were divided into three groups (school-level pressure sources, external sources of pressure and beneficiaries), and there is a separate factor that covers the general attitude towards national and international assessment programmes. Results imply that different



**Fig. 2** Model for upper secondary teachers' beliefs about changes in instruction and perceived pressure from various stakeholders

stakeholders and acceptance of large-scale educational assessments have an effect on the different dimensions of instructional practices.

Hungarian teachers generally tend to accept large-scale educational surveys to some extent and find them useful. Nevertheless, we can identify a group of teachers (about a third of them) who consider assessments as a source of trouble in school and feel that they produce more drawbacks than benefits. Similarly, according to a considerable proportion of teachers, large-scale school assessment programmes cannot significantly contribute to an objective judgement of the performance of schools and school systems. These data have shown that teachers do not have an anti-assessment attitude — they rather show a willingness to integrate assessment into their professional duties of improved teaching and learning. As regards education-level differences, our findings indicate that upper secondary teachers are more likely to accept large-scale educational assessment than their lower secondary counterparts.



The findings show that the NABC has a major effect on teaching methods. In addition, teachers mostly perceive the NABC as bringing the curriculum and training requirements to the forefront, with their work becoming more focussed as a result. However, in their view, assigning more homework and/or more difficult homework is not an acceptable consequence of the introduction of the NABC. These results are consistent with the conclusion of Hamilton et al. (2005) that teachers do not burden students with extra homework (extra tasks outside of school time) for better results. It would be worth investigating whether any modification occurred in the content and aims of homework assignments. Learning time is characteristically increased by written homework in the Hungarian education system, especially at upper secondary level, with the aim of consolidating or practising the material taught. Presumably, this is also one of the reasons why teachers do not want to burden students with further tasks.

The teachers agree with many of the statements which could point to positive or negative reactions depending on the context: the assessments highlight the necessity of supporting and developing poor performers and talented students. As a consequence, teachers pay more attention to these students in the form of extra-curricular tutoring. The assessment may draw attention to the importance of students with extreme performance and makes it possible to identify them. Teachers typically refuse to narrow down the curricular content due to the NABC, and a notable number of them also refuse to align teaching content with measured content. In addition, due to the national assessment programme, teachers indicate that they pay more attention to the development of competencies assessed as part of these very assessment programmes. At the same time, their responses do not suggest any potential neglect of content or competencies that are not measured. This might be related to the circumstance of the secondary school entrance exam (which represents high stakes for both students and schools) primarily targeting specialist/disciplinary knowledge, while competency assessments focus on skill-based knowledge. Teachers pay more attention to test-taking strategies and multiple-choice tests in classroom instruction. This is a finding related to the fact that the format of the NABC differs widely from that of conventional classroom tests. Based on the answers, 52% of the lower secondary teachers and 41% of their upper secondary peers pay more attention to practising in general. In addition, universal test-taking strategies and the test formats used in national and international assessments also appear to be more emphasized in everyday assessment practice.

Lower and upper secondary teachers differ in their beliefs about what they change in their instructional behaviour. Responses suggest that upper secondary teachers are less likely to change their teaching practices in the three dimensions under examination than lower secondary teachers. This can be explained by the difference in external reference points at these two levels. In lower secondary education, the NABC is the only such reference, while there are also other exams at the upper secondary level.

The results support the conclusions of international surveys (Moore & Waltman, 2007; Pedulla et al., 2003; Stecher et al., 2000) that teachers at both educational levels experience a great deal of pressure to boost test scores from the formal hierarchy, school administration and local government (or municipality). Typically,

the municipality has direct ties to the school administration, and teachers probably perceive pressure from the municipality through that level. Teachers' own expectations about their work also represent a significant driving force. However, in contrast with previous results (Moore & Waltman, 2007), colleagues' expectations are not insignificant pressure factors, as teachers put pressure on each other. This may be due to the fact that the stakes of assessment results are high in schools. Research shows that, according to teachers, students' results on the NABC depend mostly on the work of teachers majoring in mathematics and Hungarian language and literature, and it is their task to have their students practise and prepare for the assessment (Tóth, 2015). They believe this despite the fact that the NABC measures key competencies, the development of which is officially a task in every lesson. For lower secondary teachers, pressure from the public is only moderate. This may be related to the fact that NABC results are published in a school-level report in Hungary, these are accessible to the general public, and even schools have to publish them on their websites. Due to publicity, judgements about a school may also be affected by the NABC. Both the public and the local government have a direct impact on the school, and the teacher perceives their pressure to improve students' results on the NABC indirectly due to pressure from the school principal and colleagues. Parents and students do not seem to be a driving force. Nevertheless, data also suggests that the results from the NABC play a role in conscious school choice among some parents although not many adults consider it to be relevant (Balázsi & Horváth, 2011).

Teachers at the lower secondary level also feel more pressured by the different stakeholders to increase students' achievement on the NABC than their upper secondary counterparts. This difference might be explained by the finding that among all the student assessments at primary school, it is the NABC which places the most pressure on teachers, while at secondary schools, it is the school-leaving exam (see Tóth & Csapó, 2011). At secondary schools, even today, it is rather the school-leaving exam that is considered as the most important feedback on the efficacy of an institution's work. Also, due to its prestige, it is still treated as the indicator of the quality of education at secondary school much more than the NABC.

Our well-fitting structural model provides insights into how teachers' attitudes towards large-scale educational assessments, teaching practices and perceived pressure induced by different stakeholders relate to each other.

The external sources of pressure factor have no direct effect on the changes in instruction factors; however, it has some indirect effects. We can find an indirect relation of external pressure sources to better instruction, reallocation of content resources and test-taking strategies at the upper and lower secondary school levels and to reallocation of teacher's attention through school-level pressure sources at the upper secondary level. The model suggests that the central government, the local municipality and the public can assert their interests through the school administration. Therefore, the central and local governments can put pressure (on teachers) at the school level to change the teaching process.

From school-level sources of pressure, there are direct paths to better instruction, reallocation of content resources and test-taking strategies at both the lower and upper secondary levels and to reallocation of teacher's attention at the upper

secondary level. Our results show that the more pressure a teacher feels from the school administration and their colleagues as regards their students achieving better, the more they pay attention to their methods and the core curriculum and the better they prepare their students for the assessments. The correlation is more remarkable for upper secondary teachers in the case of better instruction and reallocation of content resources than for lower secondary teachers. Our results indicate that school-level stakeholders have the most significant effect on test-taking strategies at the upper secondary level.

Acceptance of large-scale educational assessments tended to be a positive predictor of changes in teaching practices, which is described in terms of dimensions: better instruction, test-taking strategies and reallocation of teacher's attention at both the lower and upper secondary levels. That is, the more supportive teachers are of large-scale school assessment programmes, the more likely they are to revise their instructional methods in favour of better achievement and the more they focus on the core curriculum, thus leading to more effective teaching. Teachers who are more supportive of large-scale educational surveys also tend to focus more on low and high achievers, even outside the classroom, and teach directly to the test. The beta value is the highest for better instruction, so it is possible that teachers' positive attitude towards large-scale educational assessments has the strongest results in terms of quality of teaching. The extent to which teachers have changed the content resources because of the NABC is unrelated to their acceptance of it. It is possible that teachers believe that it is not modifying teaching content that leads to good results. Neither the school-level pressure sources nor the acceptance of large-scale assessments was associated with homework assignment practice. Based on the results of this study, it appears that extra homework is not a noteworthy means of improving students' efficacy at schools. However, the path from beneficiaries to homework turned out to be significant. According to our results, teachers presumably give extra tasks to their students or extend practice to the home environment due to influence from clients — parents and students. At lower secondary level, parents and students impact how teachers reallocate their attention among their students. The explanation for this effect may be captured in parent–child as student and parent–teacher communication. After all, out of the five instructional behaviours under examination, the amount and quality of homework and an extra focus on low or high achievers, which is aimed at nurturing talent or catching up, appear in parent–child conversations about school and in teacher–parent communication during the parents' evening at school.

A multitude of research has examined teachers' perceptions of national assessment programmes along a number of dimensions, especially in the last two decades. There is a large body of literature to highlight how national assessment programmes affect teaching practice, so their positive and negative effects on the teaching and learning process are known. The novelty of our research is that it describes the phenomenon in a special context, a school system, in which enrolment in any school in any district is open, admission to schools outside the district can only be denied if there is a lack of openings, and the results of the national educational assessment programme are public at school level. It is thus only the schools, not the students, for whom the stakes of the assessment

programme are direct. Our results are particularly noteworthy for pointing out the differences between the beliefs of teachers at different levels of education: upper secondary teachers are more accepting of the large-scale assessment, are less sensitive to stakeholder pressure, and do not change their teaching practices as significantly because of the national assessment as upper secondary teachers. Our analyses are also novel in that they highlight the relationships between teachers' perceived pressure from different stakeholders, their attitude towards large-scale assessments and changes in their teaching practices. Our results draw attention to the fact that teachers' instructional behaviour is influenced by how they perceive the pressure and expectations of stakeholders to improve students' results on the NABC.

## 7 Limitations of the study

As data were obtained with self-report questionnaires, the design only allows us to draw a reliable picture of teachers' beliefs and attitudes, while the actions they reported require confirmation from other sources. Furthermore, no information is available on the extent of teachers' conformity or willingness to adjust their attitudes and behaviour to a hypothetical social norm. We do not know the extent to which the information we have collected is based on teachers' temporary impressions. Additionally, we have to examine if daily teaching practices changed due to teachers measured beliefs or factors beyond these. Our results do not yield direct evidence on how the underlying mechanism affects assessments of the teaching process. Based on the findings of this study, the next step should be to examine teachers' behaviour.

Since participation in the survey was voluntary, our results are probably positively distorted as well. Certain answers from teachers might presumably depend on similar causes (e.g. their professionalism and conscientiousness). In our present survey, we do not have sufficient instrumental variables to solve the problem of endogeneity.

This study has focussed on the impact of a national assessment programme, which is a paper-based summative assessment; however, formative and classroom assessments may have a more significant impact on student learning, as well as the newly available online diagnostic assessment. Further research may expand the scope of study to other forms of assessment as well.

**Funding** Open access funding provided by University of Szeged.

**Data availability** The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory into Practice*, 42(1), 18–29.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258–267. <https://doi.org/10.3102/0013189X07306523>
- Balázs, I., & Horváth, Z. (2011). A közoktatás minősége és eredményessége [Quality and efficacy of public education]. In É. Balázs, M. Kocsis, & I. Vágó (Eds.), *Jelentés a magyar közoktatásról 2010 [Report on public education 2010]* (pp. 325–362). Oktatókutató és Fejlesztő Intézet.
- Balázs, I., & Ostorics, L. (2020). The Hungarian Educational Assessment System. In H. Harju-Luukkainen, N. McElvany, & J. Stang (Eds.), *Monitoring student achievement in the 21st century* (pp. 157–169). Springer. [https://doi.org/10.1007/978-3-030-38969-7\\_13](https://doi.org/10.1007/978-3-030-38969-7_13)
- Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. In H. Fives & M. G. Gill (Eds.), *International handbook of research on teachers' beliefs* (pp. 284–300). Routledge.
- Barnes, N., Fives, H., & Dacey, C. M. (2017). US teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education*, 65, 107–116. <https://doi.org/10.1016/j.tate.2017.02.017>
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301–318. <https://doi.org/10.1080/0969594042000304609>
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged version. *Psychological Reports*, 99(1), 166–170. <https://doi.org/10.2466/pr0.99.1.166-170>
- Brown, G. T. L. (2018). *Assessment of Student Achievement*. Routledge.
- Brown, G. T. L., & Harris, L. R. (2009). Unintended consequences of using tests to improve learning: How improvement-oriented resources heighten conceptions of assessment as school accountability. *Journal of MultiDisciplinary Evaluation*, 6(12), 68–91.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24(4), 305–331. <https://doi.org/10.3102/01623737024004305>
- Certo, J. L. (2006). Beginning teacher concerns in an accountability-based testing environment. *Journal of Research in Childhood Education*, 20(4), 331–349. <https://doi.org/10.1080/02568540609594571>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255. [https://doi.org/10.1207/S15328007SEM0902\\_5](https://doi.org/10.1207/S15328007SEM0902_5)
- Condliffe, B., & Plank, S. (2013). Pressures of the season: An examination of classroom quality and high-stakes accountability. *American Educational Research Journal*, 50(5), 1152–1182. <https://doi.org/10.3102/0002831213500691>
- Copp, D. T. (2016). The impact of teacher attitudes and beliefs about large-scale assessment on the use of provincial data for instructional change. *Education Policy Analysis Archives*, 24(109), 1–29. <https://doi.org/10.14507/epaa.24.2522>
- Cresswell, J. (2016, July 24). *System-level assessment and educational policy*. Australian Council for Educational Research. <http://research.acer.edu.au/assessgems/10>
- Csapó, B., & Molnár, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: The eDia System. *Frontiers in Psychology*, 10, 1522. <https://doi.org/10.3389/fpsyg.2019.01522>

- Cuevas, R., Ntoumanis, N., Fernandez-Bustos, J. G., & Bartholomew, K. (2018). Does teacher evaluation based on student performance predict motivation, well-being, and ill-being? *Journal of School Psychology, 68*(6), 154–162. <https://doi.org/10.1016/j.jsp.2018.03.005>
- Darling-Hammond, L., & Rustique-Forrester, E. (2005). The consequences of student testing for teaching and teacher quality. *Yearbook of the National Society for the Study of Education, 104*(2), 289–319. <https://doi.org/10.1111/j.1744-7984.2005.00034.x>
- Dee, S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management, 30*(3), 418–446. <https://doi.org/10.1002/pam.20586>
- Finnigan, K. S., & Gross, B. (2007). Do accountability policy sanctions influence teacher motivation? Lessons from Chicago's low-performing schools. *American Educational Research Journal, 44*, 594–630. <https://doi.org/10.3102/0002831207306767>
- Firestone, W. A., Monfils, L., Camilli, G., Schorr, R. Y., Hicks, J. E., & Mayrowetz, D. (2002). The ambiguity of test preparation: A multimethod analysis in one state. *Teachers College Record, 104*(7), 1485–1523.
- Firestone, W. A., Monfils, L. F., Shorr, R. Y., Hicks, J. E., Martinez, M. C. (2004). Pressure and support. In: W. Firestone, R. Schorr & L. Monfils (Eds.), *The ambiguity of teaching to the test: Standards, assessment, and educational reform* (pp. 63–89). Lawrence Erlbaum Associates.
- Fuller, S. C., & Ladd, H. F. (2013). *School based accountability and the distribution of teacher quality among grades in elementary schools*. Working Paper No. 75. CALDER.
- Hamilton, L. S., Berends, M., & Stecher, B. (2005). *Teachers' responses to standards-based accountability. Working Paper*. RAND Corporation.
- Hamilton, L. S., Schwartz, H. L., Stecher, B. M., & Steele, J. L. (2013). Improving accountability through expanded measures of performance. *Journal of Educational Administration, 51*(4), 453–475. <https://doi.org/10.1108/09578231311325659>
- Hanushek, E. A., & Raymond, M. E. (2002). Sorting out accountability systems. In W. Evers & H. Walberg (Eds.), *School accountability* (pp. 75–104). Hoover Institution Press.
- Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance. *Journal of Policy Analysis and Management, 24*(2), 297–327. <https://doi.org/10.1002/pam.20091>
- Herman, J. L. (2004). The effects of testing on instruction. In S. Fuhrman & R. Elmore (Eds.), *Redesigning accountability* (pp. 141–166). Teachers College Press.
- Herman, J. L. (2008). Accountability and assessment: Is public interest in K–12 education being served? In K. E. Ryan & L. A. Shepard (Eds.), *The future of test-based educational accountability* (pp. 211–232). Routledge.
- Herppich, S., & Wittwer, J. (2018). Preservice teachers' beliefs about students' mathematical knowledge structure as a foundation for formative assessments. *Teaching and Teacher Education, 65*, 107–116. <https://doi.org/10.1016/j.tate.2018.06.011>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.
- 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), 297–327.
- Jones, B. D. (2007). The unintended outcomes of high-stakes testing. *Journal of Applied School Psychology, 23*(2), 65–86. [https://doi.org/10.1300/J370v23n02\\_05](https://doi.org/10.1300/J370v23n02_05)
- Juhász, J., Lannert, E., Tot, E. (2010). *Structures of education and training systems in Europe, Hungary*. 2009/2010 Edition. European Commission. [https://www.tudok.hu/file/Structures/structures\\_en\\_100325-1.pdf](https://www.tudok.hu/file/Structures/structures_en_100325-1.pdf)
- Koretz, D. (2002). Limitation in the use of achievement tests as measures of educators' productivity. *Journal of Human Resources, 37*(4), 752–777. <https://doi.org/10.2307/3069616>
- Koretz, D., Barron, S., Mitchell, K., & Stecher, B. (1996). *The perceived effects of the Kentucky Instructional Results Information System (KIRIS)*. RAND Corporation.
- Koretz, D., & Hamilton, L. S. (2003). *Teachers' responses to high-stakes testing and the validity of gains: A pilot study*. CSE Technical Report. N. 610. Center for the Study of Evaluation, University of California.
- Koretz, D., McCaffrey, D., & 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.

- Leighton, J. P., Gokiert, R. J., Cor, M. K., & Heffernan, C. (2010). Teacher beliefs about the cognitive diagnostic information of classroom versus large-scale tests: Implications for assessment literacy. *Assessment in Education: Principles, Policy & Practice*, 17(1), 7–21. <https://doi.org/10.1080/09695940903565362>
- Linn, R. L., & Dunbar, S. B. (1990). The Nation's report card goes home: Good news and bad about trends in achievement. *Phi Delta Kappan*, 72(2), 127–133.
- McNeil, L. (2000). *Contradictions of school reform educational costs of standardized testing*. Routledge.
- Moore, J., & 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.
- Nichols, S., Glass, G., & Berliner, D. (2012). High-stakes testing and student achievement: Updated analyses with NAEP data. *Education Policy Analysis Archives*, 20(20), 1–35.
- Önal, O., & Karagül, A. E. (2018). A study on Turkish EFL teachers' beliefs about assessment and its different uses in teaching English. *Journal of Language and Linguistic Studies*, 14(3), 190–201.
- Pedulla, J., Abrams L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. National Board on Educational Testing and Public Policy, Lynch School of Education, Boston College.
- Polikoff, M. S., Porter, A. P., & Smithson, J. (2011). How well aligned are state assessments of student achievement with state content standards? *American Educational Research Journal*, 48(4), 965–995. <https://doi.org/10.3102/0002831211410684>
- Popham, W. J. (2001). Teaching to the test. *Educational Leadership*, 58(6), 16–20.
- Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37–75). Kluwer Academic Publishers.
- Saeki, E., Segool, N., Pendergast, L., & von der Embse, N. (2018). The influence of test-based accountability policies on early elementary teachers: School climate, environmental stress, and teacher stress. *Psychology in the Schools*, 55(4), 391–403. <https://doi.org/10.1002/pits.22112>
- Smith, W. C., & Holloway, J. (2020). School testing culture and teacher satisfaction. *Educational Assessment, Evaluation and Accountability*, 32, 461–479. <https://doi.org/10.1007/s11092-020-09342-8>
- Stecher, B. M. (2002). Consequences of large-scale high-stakes testing on school and classroom practice. In L. S. Hamilton, B. M. Stecher, & S. P. Klein (Eds.), *Making sense of test-based accountability in education* (pp. 79–100). RAND Corporation.
- Stecher, B. M., Barron, S., Chun, T., & Ross, K. (2000). *The effects of the Washington state education reform on school and classroom* (CRESST Report 525). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Students Testing.
- Tóth, E. (2015). Az Országos kompetenciamérés hatása a tanítási munkára pedagógusinterjúk alapján [The impact of the National Assessment of Basic Competencies on teaching based on teacher interviews]. *Magyar Pedagógia*, 115(3), 115–138. <https://doi.org/10.17670/MPed.2015.2.115>
- Tóth, E. & Csapó, B. (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, 30 August - 3 September, 2011. 57.
- Tuytens, M., Devos, G., & Vanblaere, B. (2020). An integral perspective on teacher evaluation: A review of empirical studies. *Educational Assessment, Evaluation and Accountability*, 32, 153–183. <https://doi.org/10.1007/s11092-020-09321-z>
- Williams, L. (2009). Down and dirty with grammar. *Literacy Learning: the Middle Years*, 17(1), 11–20. <https://doi.org/10.3316/aeipt.173168>
- Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research, and Evaluation*, 12(3), 1–26. <https://doi.org/10.7275/mhqa-cd89>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.