

Monika Brozmanová

Matej Bel University, Slovak Republic

Beata Kosová

Matej Bel University, Slovak Republic

Consequences of National Testing of Primary School Pupils Through the Eyes of Teachers

Abstract

The study deals with the consequences of national testing at the primary school level on the processes of education, teachers' work in the classroom and the teaching profession. It is based on the international discourse about the consequences of large-scale testing and investigates Slovak teachers' opinions and measures they have implemented under the influence of testing in their own instruction and school. The research, carried out by the questionnaire method on a sample of 786 (2017) and 940 (2020) respondents, confirms that teachers perceive national testing as too far from instruction and inconsistent with educational standards. They do not believe it can fulfil the set goals, but, on the other hand, because of it, they narrow down instruction to prepare for tests and take measures that do not change the quality of pupil learning.

Keywords: *large-scale testing, national testing, results of education*

Introduction

With the development of reforms aimed at the decentralisation of education, the autonomy of schools, and reforms based on educational standards, there is increased pressure on monitoring the effectiveness of schools and school systems through national testing around the world. The growing culture of testing also carries over into schools by replicating test formats of large-scale testing in classroom work. It affects children's, parents' and teachers' prevalent views of life, and notions of the public about education (Volante et al., 2020), causing tension among

large-scale testing, instruction and classroom assessment (Wilson, 2018; Heritage, 2018; Shepard et al., 2018b).

Despite countries' different *histories of national testing* or its longstanding criticism, general fascination with standardised testing has persisted over the last decades (Simpson, 2017). The first national testing served for making decisions on pupils' school careers (Island 1946; Portuguese, North Ireland 1947). In the 1960s –70s, it was implemented in 6 countries and at the turn of the 21st century in 21 other European countries to monitor the quality of schools and education systems (National Testing of Pupils in Europe..., 2009). The Slovak Republic began nationwide testing in 2009 (ISCED 2) and 2015 (ISCED I). Finland is an exception, preferring formative assessment by teachers, testing only a sample of 5–10% of the age group to monitor the fulfilment of national curriculum objectives and equality of education (Volante et al., 2020).

Western democracies have been alternatively strengthening testing and introducing new types of national tests (Scotland, England, Denmark, France), even overvaluing external assessments instead of assessments by teachers (England), or abolishing tests (Portuguese for 35 years), revising or replacing them with other types of tests with different goals (Ireland, Iceland). However, since 2000, most countries have relied on large-scale testing to assess the effectiveness of their education systems (National Testing of Pupils in Europe..., 2009, p. 20). Assessment systems (Germany, Canada) are increasingly modelled to align with PISA (Volante et al., 2020, p. 27). The same is true of post-socialist countries that have not passed the phase of constructive criticism of the formats and purposes of large-scale testing and have begun large-scale testing without a deeper analysis. Slovakia is an example of a country where the education policy considers national testing without a professional debate to be an objective tool for determining the results of pupil education and the work of schools.

In general, states declare *three types of different goals for national testing*: 1) to decide on a further school career at the end of the education level; 2) to monitor the quality of schools and education system – both are summative tests; and 3) to support learning by clarifying pupils' specific needs and identifying appropriate instruction – assessment for formative purposes. Experts consider using one test for several purposes entirely inappropriate because each requires a different test design (National Testing of Pupils in Europe..., 2009). Slovakia, conversely, declares more purposes: to obtain objective information on pupils' performances on entering the lower secondary level, verify the level of their knowledge and skills, abilities to apply knowledge in practical tasks and abilities to think logically, provide feedback on pupils' knowledge to schools, the decision-making sphere and teachers for instruction improvement (Brozmanová & Kocisová, 2021, Testovanie 5), which, as explained below, is not feasible.

The last period is typical of a *fierce professional debate on the rehabilitation of formative assessment in classrooms* and efforts to create a strong, well-supported assessment culture for learning amid the dominating influence of national testing programmes (Volante et al., 2020). It was triggered in 2017 by M. Wilson and his Presidential Address at the US National Council for Measurement in Education meeting, stressing the priority of formative classroom assessment and calling to seek its coherence with large-scale testing by creating formative testing tools facilitating pupils' progress in learning. He pointed to negative results of inappropriate use of summative measurements by states and teachers. National testing can signal teachers what standards they should be teaching and give them concrete examples of what the standards mean, but they do not allow them to make diagnostic conclusions about individual pupils or the results of their education (Wilson, 2018, pp. 6–7). The criticism of excessive preference for testing follows from research findings that *the tests do not measure pupils' actual knowledge* but general cognitive skills (Popham, 1999), making 80% of reliable variances of scores in mathematics, reading and writing (Sicolý, 2002, p. 22). The tests do not focus on many aspects of cognition important for success, such as the organisation of knowledge, the use of thinking strategies, and self-monitoring skills (Volante et al., 2020). In many items, success depends on out-of-school experience (Popham, 1999), and then socioeconomic status is mistaken for pupils' results (Noble et al., 2012, p. 17). Pupils fail because they cannot read or comprehend (Sigborsson, 2017). Research shows incoherence between actual knowledge and test success. Through the use and analysis of verbal protocols¹, Bond (2020) proved that tests do not arrive at valid conclusions about pupils. Failing pupils were mostly unable to create a structure of problem-solving (generate a set of sub-goals) but had no problems correctly answering sub-questions leading to the solution, which was related to their learning methods (Bond, 2022, pp. 88–89). Noble et al. (2012, p. 791) subjected six items of the national test to the verbal protocol and confirmed that the tests reflect sociocultural differences and not a lack of knowledge because of the mismatch between the tests and the language and life experience of pupils or because of misunderstanding the context that did not make sense to them. False negative answers (content knowledge but incorrect answer) were statistically significantly more frequent in non-dominant communities.

Since the open letter, *OECD and PISA Tests are Damaging Education Worldwide (2014)*, signed by 83 academics from around the world, experts have been devoted to analysing the *consequences of large-scale testing on education*. Education policy

¹ Verbal protocol (think aloud protocol) – supplementing the test by a subsequent set of sub-questions leading to the correct solution

is criticised for assessing education quality by quantifying measurable aspects of education to the detriment of immeasurable objectives (Sigþórsson, 2017). Narrowing curriculum, prioritising content and competence that can be easily tested and leaving out standards difficult to test, teaching test items, narrowing classroom assessment and practising test-taking skills (Bond, 2020; Heritage, 2018; Noble et al., 2012; Sigþórsson, 2017; Wilson, 2018) are considered the most severe long-term consequences. Tests direct attention to a small part of abilities and prefer a minimalist conception of literacy (Simpson, 2017), thus narrowing the ways that teachers think about what a certain standard means (Wilson, 2018). It jeopardises pupil learning and engagement. Concentrating only on content knowledge is against inspiring, creative, interpretative and critical thinking (Simpson, 2017), prevents realistic self-awareness, reduces children's self-esteem, and impairs identity and motivation (Sigþórsson, 2017; Heritage, 2018). The *consequences of large-scale testing on the status of teachers and the teaching profession* are also at the centre of the professional debate. The emphasis on the significance of external assessment places teachers in political-power relations, public rhetoric challenges teachers' professional autonomy and knowledge, teachers are intensely monitored, neoliberal distrust of them is encouraged, and there is over-simplification of what the teaching profession consists of (Sigþórsson, 2017; Simpson, 2017). Testing depicts teachers as ineffective. Thus, they defend their position by challenging assessment instruments (Popham, 1999), refusing national tests as too distant from instruction with no effects on change in daily practice (Marion, 2018). It especially concerns primary education teachers because their main goal is the immeasurable holistic development of pupils, and viewing their work through the lens of testing only discredits their professional efforts.

A similar professional debate is not available in post-socialist countries²; there are almost no publications about this issue in Slovakia. Pilot testing projects have raised a wave of teachers' criticism at the primary education level since 1989, promoting the individual approach to pupils and inclusive education the most. Nevertheless, Testing 5 (hereinafter referred to as T5) in the subjects Mother Tongue and Mathematics has been implemented as compulsory. Instruction at the primary education level is carried out according to the state educational programme (2008), based on standards aiming to develop pupils' competencies. In 2015, the programme was updated, reinforcing the emphasis on reading comprehension, conceptual knowledge and higher dimensions of cognitive processes. Despite this, between 2016 and 2019, there was

² For instance, so far there has been only one study published in *The New Educational Review*, about pupils' attitudes toward international testing in Mathematics (Konarzewski, 2019).

only a slight change in pupils' achievements from 63.1% to 64.8% in Mother Tongue and 62.3% to 63.4% in Mathematics.

Since national testing has not been researched in Slovakia so far, our research is the first probe to find out what teachers think about national testing, how they respond to its results with their instruction and whether there are changes in their opinions and the consequences of testing after the updates.

Research Methods

The research was carried out using an electronic questionnaire of our own design in 2017 and 2020. The questionnaire contained 15 questions, of which only those were selected for the study that concerned that aim. The research sample consisted of two balanced groups of primary and lower secondary education teachers teaching Mother Tongue and Mathematics in the 5th grade (Table 1).

Table 1. Composition of the research sample

Year	Number	Women	Experience 11-20 years	Experience 20 + years	Primary level	School in town
2017	786	93.5%	31%	53.8%	49.7%	46.7%
2020	940	94.3%	26%	56.7%	50.2%	51%

The statistical methods used to compare the two groups in the nominal variable were the Chi Square test of independence and Cramer's V coefficient to measure the effect size. To compare the two groups in the ordinal variable, the non-parametric Mann-Whitney U test was used, and the coefficient Cohen d was used to determine the effect size.

Research Results

Teachers' opinions and statements were very compact. No statistically significant differences were found between their lengths of experience, workplace, or level of education at which they were teaching. Due to the overwhelming predominance of women in the research, the research could not be evaluated by teachers' gender. Therefore, especially descriptive statistics were used.

Teachers' opinions on national testing showed marked polarisation, where they divided into almost balanced groups in both measurements, according to which T5

fulfilled or failed to fulfil set goals. Almost half of the respondents did not believe the set goals could be achieved by testing (Table 2).

Table 2. Does T5 fulfil the set goals?

	2017		2020	
	N	%	N	%
It fulfils the set goals	23	2.9	45	4.8
It does partially, gives feedback about a certain performance	389	49.5	464	49.4
Rather does not, gives results only about the test writing skill	271	34.5	288	30.6
No, it does not fulfil the set goals at all	98	12.5	135	14.4
No answer	5	0.6	8	0.9
Total	786	100	940	100

$$\chi^2 = 7.113, df = 4, p = 0.13$$

There were statistically significant changes in teachers' opinions on the consistency with educational standards in instruction; in 2020, 68% of respondents very much or mostly believed that there was a mismatch between standards and national testing, against 48.3% in 2017. Thus, most teachers expressed opinions that T5 was too distant from instruction and assessed its relevance accordingly (Table 3).

Table 3. Do the tasks in T5 correspond to educational standards in instruction?

	2017		2020	
	N	%	N	%
Yes, always	13	1.7	70	7.4
Yes, mostly	441	56.1	200	21.3
No, mostly not	293	37.3	577	61.4
No, not at all	24	3.1	62	6.6
No answer	15	1.9	31	3.3
Total	786	100	940	100

$$\chi^2 = 232.932, df = 4, p = 0.000, \text{Cramer's } V = 0.184, \text{medium practical significance}$$

In both years, 90% of respondents thought that T5 could not improve the quality of education in the future. One-third of them predicted a decrease in quality due to increased testing. Thus, teachers did not consider national testing an appropriate instrument for increasing the quality of education. It could also be seen in their open assessments of national testing. For both years, 641 statements were obtained about its positives, however, with the fact prevailing that pupils acquired experience in testing. In 870 statements on the negatives, however, the stress of

parents, teachers and pupils prevailed (50%), and non-objectivity of testing, which indicated how negatively they were experiencing it.

The opinions were related to *how national testing affected the educational process and teachers*. Fortunately, in both years, 60% of teachers reported that T5 results had no effect on the evaluation of a particular teacher, and according to a third of them, the T5 results had only a partial effect as one of the criteria. However, further answers showed that the situation in schools in Slovakia was developing in a way criticised as a negative narrowing of the curriculum (Tables 4 and 5).

Table 4. Did T5 influence instruction and evaluation in your school?

	2017		2020	
	N	%	N	%
Yes, considerably; we began to examine mostly in the form of tests	76	9.7	128	13.6
Yes, partially, mainly in months before the date of testing	362	46.1	503	53.5
No, rather not; sometimes we include the T5 tasks in the instruction	206	26.2	193	20.5
No, instruction goes on as before the T5 implementation	134	17.0	107	11.4
No answer	8	1	9	1
Total	786	100	940	100

$\chi^2 = 26.214$, $df = 4$, $p = 0.000$, *Cramer's V* = 0.055, low practical significance

The fact that against 2017 the results statistically significantly worsened toward learning for tests (2017 total – 55.8%; 2020 total – 67.1%) means that Slovakia did not learn from international experience. Answers reflected a flat, less analytical view on the ways to achieve better results in national testing by further measures with increased preparation for testing (2017 – 42.6%; 2020 – 51.3%) without change in the very essence of instruction.

Table 5. Do you pay attention to testing by further specific activities?

	2017		2020	
	N	%	N	%
We started a test prep course	41	5.2	95	10.1
We increased the number of language and math lessons	145	18.4	172	18.3
We provide tutoring for pupils before preparation for T5	149	19	215	22.9
We bought (parents paid) publications with T5 tests	162	20.6	188	20
Other	268	34.1	252	26.8
No answer	21	2.7	18	1.9
Total	786	100	940	100

$\chi^2 = 24.82$, $df = 5$, $p = 0.000$, *Cramer's V* = 0.055, low practical significance

According to the statements, *teachers did not reflect on the possibilities of using the national testing results*. On the one hand, they were aware of the problem because 3/4 of the teachers (2017 – 75.3%; 2020 – 76.2%) answered that it was reading the task assignment with comprehension. Paradoxically, however, not even one respondent answering “Other”, who were more than 500 in both years, did not report activities precisely aimed at improving reading comprehension. Most open answers reported changes in the timetable before testing, solving older tests or private tutoring.

Discussion

According to the results, national testing in Slovakia is strongly separated from classroom education processes and is perceived by teachers as such or objected to.

According to experts, *linking instruction, pupil learning, and assessment* is crucial for testing to support learning processes and improve education (Wilson, 2018; Shepard et al., 2018a; Marion, 2018). However, this requires the curriculum, instruction and assessment to be prepared and developed together (Wilson, 2018), thus conceptually compatible (Shepard et al., 2018a), and the assessment system to be coherent with the ways pupils are expected to progress (Marion, 2018).

Slovakia has a big conceptual discrepancy between education and national testing in primary education. In the last decades, there have been evident efforts to realise primary education based on cognitive and socio-cognitive theories with a personalistic orientation on pupils’ holistic development, but in the sociocultural context. However, national testing is based on a behaviourist perspective, with a tendency to decompose content learning into tiny bits and to test it cross-sectionally against standards (Shepard et al., 2018b). It is precisely why replication of national tests in instruction is dangerous. Although there is a debate on personalised (electronic) tests, reacting directly to the pupil’s response, experts agree that the curriculum, pupils’ learning and assessment should be based on the sociocultural theory. It is because it explains the social nature of learning and the creation of meanings the best, links cognitive aspects with motivation and the creation of identity, and supports self-regulation and engagement (Heritage, 2018; Noble et al., 2012; Shepard et al., 2018b). The sociocultural theory is the framework, but a coherent collective curriculum design, instruction, and assessment require more subtle, discipline-specific learning models (Shepard et al., 2018a). For the testing to be used for formative purposes, the collective development should be based on *learning progressions*. Learning progressions are descriptions of successive and

gradually more complicated ways of thinking about a specific domain, schemes (maps) of progress, defining the order of learning experience, learning aims as interlinked steps along the route the pupil needs to take to get closer to that aim from his or her current position (Wilson, 2018, pp. 8–9). The curriculum structure, classroom instruction and assessment must be conceived rather longitudinally than cross-sectionally (Shepard et al., 208b). Unfortunately, in Slovakia, neither teachers nor education policy think this way. Standards and textbooks are conceived by the contents of scientific disciplines, not by the development of pupils' thinking in them.

An example is the issue of reading comprehension. On the one hand, T5 added separate tasks for reading comprehension to the extent of up to 40%, although it was part of each task. On the other hand, after increasing learning for tests, there was a decrease in success in those tasks (2017 – 72.8%; 2020 – 68.4%). Pupils perceive what really matters from what their teachers prioritise. Davis & Vehabovic (2017, p. 581) found out about reading instruction that if tested standards are prioritised, test-formatted passages used, or test items taught (to the detriment of more comprehensive comprehension methods related to critical reading, searching, social affiliation, own identity and interest in self-selected reading), pupils learn that it is not important to read multiple texts for comparison, or build knowledge by non-linear experience, perspectives, overcoming prejudices; that comprehension is about the categorisation of small chunks of texts and agreeing with an authoritative interpretation of the author's message or that answering questions right and not thinking is the most important.

Conclusion

Our research confirms that national testing is perceived by most teachers as separated from and conceptually inconsistent with their classroom instruction. Adopted measures lead to an increase in learning for tests instead of developing pupils' comprehension of the text. Teachers use the information from testing only to repeat test tasks without information about problems in pupils' thinking. Slovakia would need a constructive professional debate and teachers' quality improvement in formatively oriented tests designed according to learning progressions.

The limitations of the questionnaire method based on subjective statements do not allow for drawing more specific conclusions. However, they open many inspirations for future education research, such as verification of the national testing validity, investigation of false success and failure through verbal protocols,

deeper research on reading literacy instruction in relation to testing, experimental joint development of curriculum, instruction and assessment.

We are aware, in agreement with the cited authors, that assessment at the state level cannot be closely aligned with classroom assessment because they serve different purposes. However, they could be loosely interlinked and conceptually coherent (Wilson, 2018; Marion, 2018; and others), based on collaboration of practice, research and education policy, that would re-direct the efforts to gather evidence about the reputed effectiveness of the education system to the efforts to contribute to the central mission of education – teach children what they need to know. Unless this is the case, national testing should at least cease to be claimed to tell about the quality of education and be an objective instrument for measuring its results.

References

- Bond, L. (2020). Assessment in the Service of Student Learning: Three Cases in Point. *Educational Measurement: Issues and Practice*, 39(3), 86–89. DOI: 10.1111/emip.12373
- Brozmanová, M., & Kocisová, L. (2021). Hodnotenie kvality primárneho vzdelávania Testovaním 5 z pohľadu učiteľov [Evaluation of the Quality of Primary Education by Testing 5 from the Teachers' Perspective]. *Slavonic Pedagogical Studies Journal*, 10(1), 1–16. DOI: 10.18355/PG.2021.10.1.1
- Davis, D. S., & Vehabovic, N. (2017). The Dangers of Test Preparation: What Students Learn (and Don't Learn) About Reading Comprehension from Test-Centric Literacy Instruction. *The Reading Teacher*, 71(5), 579–588. DOI: 10.1002/trtr.1641
- Heritage, M. (2018) Making Assessment Work for Teachers. *Educational Measurement: Issues and Practice*, 37(1), 39–41. DOI: 10.1111/emip.12191
- Konarzewski, K. (2019). Primary School Pupils' Attitude toward Mathematics and their Achievement in Mathematics. *The New Educational Review*, 56(2), 17–28. DOI: 10.15804/tner.2019.56.2.01
- Marion, S. F. (2018). The Opportunities and Challenges of a Systems Approach to Assessment. *Educational Measurement: Issues and Practice*, 37(1), 45–48. DOI: 10.1111/emip.12193
- National Testing of Pupils in Europe: Objectives, Organisation and Use of Results*. (2009). EACEA P9 Eurydice.
- Noble, T., Suarez, C., Rosebery, A., O'Connor, M. C., Warren, B., & Hudicourt-Barnes, J. (2012). "I Never Thought of it as Freezing": How Students Answer Questions on Large-Scale Science Tests and What They Know about Science. *Journal of Research in Science Teaching*. 49(6), 778–803. DOI: 10.1002/tea.21026
- Popham, W. J. (1999). Where Large Scale Educational Assessment is Heading and Why it shouldn't. *Educational Measurement: Issues and Practice*, 18(3), 4–35.

- Shepard L. A., Penuel W. R., & Pellegrino, J. W. (2018a). Classroom Assessment Principles to Support Learning and Avoid the Harms of Testing. *Educational Measurement: Issues and Practice*, 37(1), 52–57. DOI: 10.1111/emip.12195
- Shepard L. A., Penuel W. R., & Pellegrino, J. W. (2018b). Using Learning and Motivation Theories to Coherently Link Formative Assessment, Grading Practices, and Large-Scale Assessment. *Educational Measurement: Issues and Practice*, 37(1), 21–34. DOI: 10.1111/emip.12189
- Sicoli, F. (2002). What Do School-level Scores from Large-Scale Assessments Really Measure? *Educational Measurement: Issues and Practice*, 2(4), 17–26.
- Sigþórsson, R. (2017). Testing or transforming practice: probing an Icelandic national initiative to improve literacy education. *Literacy*, 51(2), 65–73. DOI: 10.1111/lit.12107
- Simpson, A. (2017). Teachers negotiating professional agency through literature-based assessment. *Literacy*, 51(2), 111–119. DOI: 10.1111/lit.12114
- Testovanie 5. (n.d.). *NÚCEM Národný ústav certifikovaných meraní. [Testing 5. National Institute for Certified Educational Measurements]*. <https://www.nucem.sk/sk/merania/narodne-merania/testovanie-5/o-testovani-5>
- The Guardian*. (2014). *OECD and PISA Tests are Damaging Education Worldwide*. <https://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics>
- Volante, L., DeLuca, L., Adie, L., Baker, E., Harju-Loukkainen, H., Heritage, M., Schneider, Ch., Stobart, G., Tan, K., & Wyatt-Smith, C. (2020). Synergy and Tension between Large-Scale and Classroom Assessment: International Trends. *Educational Measurement: Issues and Practice*, 39(4), 21–29. DOI: 10.1111/emip.12382
- Wilson, M. (2018). Making Measurement Important for Education: The Crucial Role of Classroom Assessment. *Educational Measurement: Issues and Practice*, 37(1), 5–20. DOI: 10.1111/emip.12188

AUTHORS

MONIKA BROZMANOVÁ

PhD., Matej Bel University, Faculty of Education
Banská Bystrica, Slovakia
e-mail: monika.brozmanova@umb.sk
ORCID: 0000-0001-7526-8675

BEATA KOSOVÁ

PhD., Matej Bel University, Faculty of Education
Banská Bystrica, Slovakia
e-mail: beata.kosova@umb.sk
ORCID: 0000-0002-4221-6235