

## Subjektívne alebo objektívne o národných testovaniach?

V posledných týždňoch rezonovala vo verejnosti téma externého testovania – či už išlo o testy externej časti maturitnej skúšky (EČ MS) alebo o Testovanie 9 (T9).

Náš článok vysvetľuje, ako sú koncipované testy národných meraní, a ako treba interpretovať ich výsledky. Venuje sa tiež podnetom a názorom, ktoré sa objavujú na verejnosti v súvislosti s národným testovaním.

Podme sa postupne pozrieť veci na koreň v snahe objektivizovať subjektívne.

### „Test bol zlý, lebo ani naši najlepší žiaci v ňom nedosiahli plný počet bodov!“

Národné testovania potvrdzujú to, čo je v školskej praxi známe – žiak, ktorý je v jednej triede či škole hodnotený známku výborný, inde môže byť ohodnotený horšie – a platí to i naopak, nech už sú príčiny a dôvody akékoľvek. Národné testy, ktoré pripravuje NÚCEM, sú tzv. **rozišovacie testy** (Norm-Referenced = NR-testy), ktorých cieľom nie je overenie miery osvojenia testovaných poznatkov a zručností žiakmi, ale porovnanie miery osvojenia poznatkov medzi žiakmi navzájom – teda vytvorenie poradia testovaných žiakov podľa miery úspešnosti v danom teste priradením percentilu každému žiakovi (percentil určuje percento žiakov, ktorí dosiahli horší alebo rovnaký výsledok ako daný žiak). **Celková priemerná úspešnosť žiakov v takto zostavenom teste by sa mala pohybovať od 40 % do 60 %**, čo je podľa teórie testovania hodnota všeobecne akceptovaná ako optimálna.

Test by mal obsahovať široké rozpätie úloh z hľadiska náročnosti, len tak je možné dostatočne rozlíšiť žiakov jednotlivých výkonnostných skupín. Na záver treba zdôrazniť, že ani reálne najlepší žiak školy nemusí v národnom teste dosiahnuť úspešnosť 100 %, a v žiadnom prípade to netreba od neho očakávať – ani zo strany školy či rodičov, ani zo strany žiaka samotného.

Tabuľka 1 ponúka prepojenie úspešnosti a percentilu v teste z anglického jazyka úrovne B1 z maturitnej skúšky v roku 2014, z ktorého je jasné, že iba jediný žiak dosiahol v teste maximálny počet bodov, pričom však aj napr. výsledky s priemernou úspešnosťou 78,3 % sa dajú považovať za vynikajúce vzhľadom na to, že percentil týchto žiakov je 91.

Tab. 1 – Prepojenie úspešnosti a percentilu v teste EČ MS z AJB1 (MS 2014)

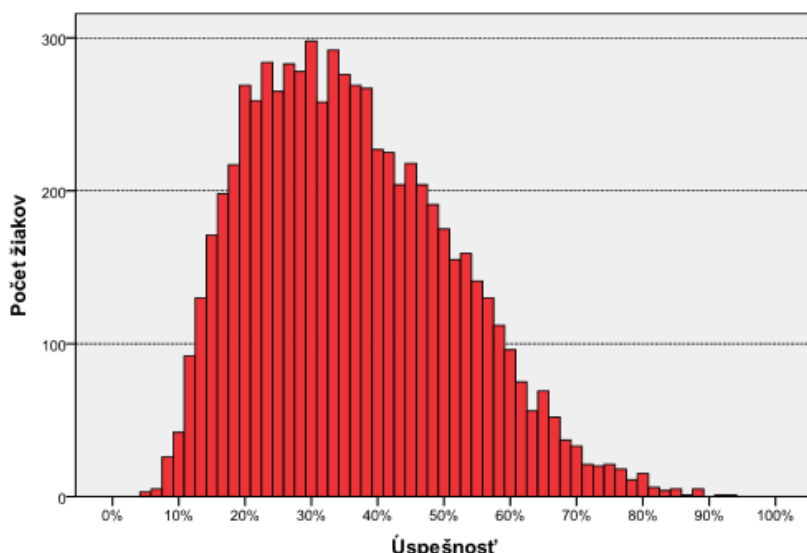
Úspešnosť	Percentil	Počet žiakov
3,3	,0	1
5,0	,0	2
6,7	,0	9
8,3	,1	11
10,0	,1	20
11,7	,2	43
13,3	,4	67
15,0	,7	97
16,7	1,1	191
18,3	1,9	205
20,0	2,8	263
21,7	3,9	293
23,3	5,2	336
25,0	6,6	411
26,7	8,4	438
28,3	10,3	514
30,0	12,5	533
31,7	14,8	594
32,5	17,4	1
33,3	17,4	593
35,0	19,9	589
36,7	22,5	675
38,3	25,4	649
40,0	28,2	662
41,7	31,0	765
42,5	34,3	1
43,3	34,3	760
45,0	37,6	770
46,7	40,9	753
47,5	44,2	1
48,3	44,2	784

Úspešnosť	Percentil	Počet žiakov
50,0	47,6	737
51,7	50,7	786
53,3	54,1	716
55,0	57,2	744
56,7	60,4	704
58,3	63,5	651
60,0	66,3	648
61,7	69,1	642
63,3	71,8	649
65,0	74,6	599
66,7	77,2	559
68,3	79,6	523
70,0	81,9	494
71,7	84,0	418
73,3	85,8	455
75,0	87,8	416
76,7	89,5	347
78,3	91,0	329
80,0	92,5	307
81,7	93,8	276
83,3	95,0	227
85,0	96,0	212
86,7	96,9	201
88,3	97,7	149
90,0	98,4	121
91,7	98,9	104
93,3	99,3	77
95,0	99,7	42
96,7	99,9	20
98,3	99,9	11
100,0	100,0	1

### „Žiakova úspešnosť 48 % v teste je slabá!“

Podľa toho, čo sme vyššie uviedli je zrejmé, že takýto výsledok v niektorom z testov národných testovaní **je v súlade s očakávaniami** a priemerom úspešnosti v NR-teste. V tomto prípade si treba všimnúť **hodnotu percentilu**. Ak by napríklad tento žiak mal v danom teste percentil 80,0, znamená to, že 80 % ostatných žiakov dosiahlo v tom istom teste horší alebo rovnaký výsledok. Z toho vyplýva, že daný test bol zrejme pre testovanú populáciu náročný, a žiak, s hoci zdanlivo nižšou priemernou úspešnosťou 48 %, patril k žiakom s celkovo najlepšimi nameranými výsledkami.

Pri pohľade na obrázok 1 vidíme, že rozloženie žiackej úspešnosti v teste z nemeckého jazyka úrovne B1 na maturitnej skúške v roku 2014 svedčí práve o takto interpretovanom vzťahu percentilu a celkovej náročnosti testu. Celý histogram je posunutý doľava od stredu, čo znamená, že **test bol pre testovaných pomerne náročný**.



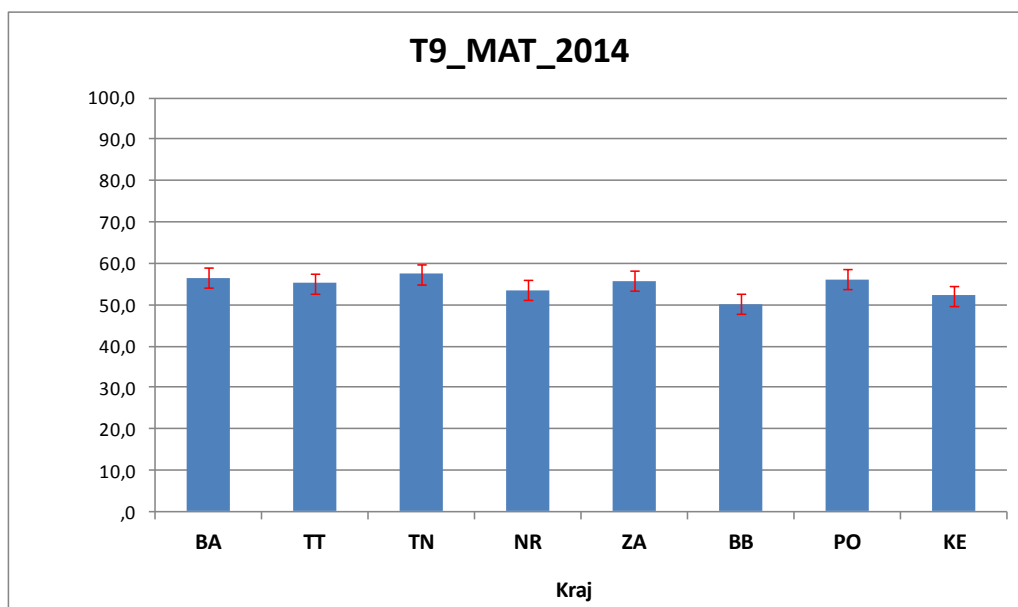
Obr. 1 – Histogram úspešnosti v teste NJB1 (MS 2014)

Opačný prípad by bol, keby sme mali žiaka s priemernou úspešnosťou v teste 80 %, ale jeho percentil by bol napr. 55. Toto by znamenalo, že v danom teste bolo až 45 % žiakov ešte lepších než sledovaný žiak. Vzhľadom na koncepciu NR-testov by sa dal takýto test pokladať za veľmi ľahký pre daný testovaný súbor. Ideálne je, keď vrchol histogramu je v oblasti medzi 40 – 60 % úspešnosti.

### „Naša škola má najhoršie výsledky v regióne v porovnaní s inými školami!“

V záverečných správach k jednotlivým testovaniam, ktoré NÚCEM vydáva každý rok, sa okrem iného spomína aj tzv. štatistická významnosť. V našom článku sa jej nebudeme podrobnejšie venovať, podstatné však je **chápať rozdiely medzi výsledkami škôl na pozadí štatisticky podložených výstupov a interpretácií**. Prezentovať svojvoľne na základe rozdielov medzi úspešnosťami určité poradie škôl – či už v rámci regiónu, alebo v rámci celej krajiny – je neobjektívne a nesprávne.

Pre ilustráciu na nasledujúcom obrázku znázorňujeme úspešnosť v teste z matematiky (Testovanie 9 2014) podľa krajov, na ktorom vidíme medzi krajinami minimálne rozdiely priemerných úspešností, ktoré nie sú zo štatistického hľadiska významné. Tvrdiť, že žiaci Banskobystrického kraja boli v teste „horší“ než žiaci Trenčianskeho kraja, by bolo **nepresné**, neprofesionálne a v konečnom dôsledku neférové a subjektívne.



Obr. 2 – Porovnanie priemernej úspešnosti podľa krajov v teste T9 z MAT (Testovanie 9-2014)

Veľmi dobre rozumieme ohlasom vyučujúcich jednotlivých predmetov, ktorí môžu mať pocit „nátlaku“ napr. zo strany vedenia školy, ktoré sa – úplne pochopiteľne – snaží prezentovať čo najlepšie výsledky testovania. Práve z tohto dôvodu predmetoví koordinátori NÚCEM-u na každom školení i v záverečných správach za jednotlivé testované predmety zdôrazňujú charakter testovania, ako aj jeho zmysel, čo však, zdá sa, ešte nie je úplne vžitú do pedagogickej praxe a zmýšľania – či už vedenia školy, alebo aj samotných vyučujúcich.

### „V teste neboli podchytené všetky oblasti predmetu.“

Ako sme spomenuli vyššie, testy národných meraní (podobne ako napr. testy PISA či PIRLS) patria k tzv. NR-testom a ich **cieľom nie je pokrytie všetkého preberaného učiva predmetu**. To je skôr zmyslom tzv. kritériálnych testov (Criterion-Referenced = CR-testy). Zastúpenie jednotlivých oblastí predmetu v externom teste je dané špecifikáciou testu, ktorá je vždy v predstihu zverejnená, a ktorá, prirodzene, rešpektuje jednak záväzné pedagogické dokumenty, a jednak samotnú koncepciu NR-testu. Ak by sme mali pripravovať národné CR-testy, ktoré by overovali celý rozsah učiva na výstupe z jednotlivých stupňov vzdelania, takéto testy by museli obsahovať oveľa väčší počet úloh a ich časová dotácia by bola oveľa väčšia, ako je to pri NR-testoch. Kým teraz napr. test z matematiky pre Testovanie 9 trvá 60 minút, pri CR testoch by to bolo 150 minút, prípadne aj viac.

### „Testy ani zadania PFIČ MS nereflektovali aktuálne udalosti a dianie na Slovensku/vo svete.“

Príprava testu externej časti maturitnej skúšky alebo testu Testovania 9 je **dlhodobým procesom**, pri ktorom sa neprihliada napr. na to, aké kultúrne udalosti budú mať možno vplyv na reálne testovanie v danom čase. Naopak, pri výbere tém PFIČ MS z vyučovacích a cudzích jazykov sa **zámerne nevyberajú** témy reflektujúce aktuálnu situáciu, ktoré by bolo možné ľahko predpovedať a študenti pripravujúci sa na maturitnú skúšku by mohli byť oproti predchádzajúcim či budúcim ročníkom vo výhode.

### „Vyžrebované maturitné témy neobsahovali konkrétny žáner.“

Vo vyučovacích i cudzích jazykoch je presne v cieľových požiadavkách stanovené, ktoré žáner musia maturanti zvládať. V jednotlivých obálkach, z ktorých sa žrebuje pri výbere tém písomnej formy internej časti maturitnej skúšky napr. zo slovenského jazyka a literatúry, sú

zastúpené všetky povinné žánre, určenie konkrétneho je už vecou náhody, **závisí od žrebovania.**

### „Tvorcovia testov sú odtrhnutí od školskej reality.“

Autormi testov zo všetkých testovaných predmetov pre Testovanie 5, Testovanie 9 a maturitnú skúšku sú **učitelia** základných, resp. stredných škôl z praxe, ktorí majú skúsenosti s tvorbou testov, metodikou či s autorstvom učebníc. Okrem autorov testových položiek sa na procese prípravy testovania podieľajú aj ďalší odborníci – **recenzenti** testov z radov pedagógov škôl rozličného zamerania. V neposlednom rade sa na tvorbe testov do istej miery podieľajú aj samotní **žiaci a študenti**, ktorí absolvujú pilotné testovanie, pričom sa ich pripomienkam a podnetom venuje patričná pozornosť.

### „Test bol ťažší/ľahší než minulý rok.“

Porovnávať náročnosť testov medziročne nie je v našich podmienkach možné. **Keďže NÚCEM na rozdiel od iných krajín zverejňuje testy pre potreby žiakov a učiteľov, nemôžu sa v nasledujúcich ročníkoch testovania objaviť v testoch rovnaké, tzv. kotviace položky, na základe analýzy ktorých by sa dalo porovnávať medzi dvoma rozdielnymi testovanými populáciami a priniesť relevantné výsledky.** Takto zostávajú všetky úvahy o zvýšenej/zníženej náročnosti testov iba v rovine dojmov.

Pravdou zostáva, že napr. do tohtoročného testu EČ MS zo slovenského jazyka a literatúry bolo aj na základe požiadaviek vyučujúcich SJL zaradených menej úloh na čítanie s porozumením v porovnaní s predchádzajúcimi rokmi, čo, pochopiteľne, spôsobilo, že v teste pribudlo viac úloh z jazyka a literatúry. A keďže čítanie s porozumením bolo tou zložkou, ktorá v minulosti zvyšovala priemernú percentuálnu úspešnosť testu EČ MS zo SJL (a zároveň do istej miery znižovala jeho rozlišovaciu schopnosť), tento rok priemerná úspešnosť testu očakávane poklesla. V teste EČ MS z anglického jazyka úroveň B2 je zase žiacka úspešnosť vyššia, keďže vzhľadom na zrušenie uznávania náhrad maturitnej skúšky z cudzieho jazyka test riešili aj maturanti, ktorí majú certifikáty z jazykových škôl, a preto sa môže na prvý pohľad zdať, že test bol oproti minulým rokom ľahší.

### „Niektoré úlohy v testoch boli veľmi ťažké až nezvládnuteľné.“

Zloženie testových úloh predpisuje už spomínaná špecifikácia testu. **NR-test musí obsahovať aj veľmi náročné položky, pri ktorých je predpoklad, že kvalitne rozlíšia práve žiakov najlepších výkonnostných skupín.** Všetky úlohy prechádzajú pilotným testovaním na relevantnej vzorke žiakov, a ak sa po analýze výsledkov pilotného testovania ukáže, že je úloha extrémne náročná (max. 10 % žiakov odpovedalo správne), resp. extrémne ľahká (aspoň 90 % žiakov odpovedalo správne), je zo súboru úloh vylúčená. Obťažnosť položiek (percentuálny podiel žiakov, ktorí správne riešili položku) vyhodnocujeme nasledovne:

Tab. 2 – Obťažnosť testových položiek

Hodnota obťažnosti v %	Obťažnosť položky
<0; 20)	veľmi obťažná
(20; 40)	obťažná
(40; 60)	stredne obťažná
(60; 80)	ľahká
(80; 100)	veľmi ľahká

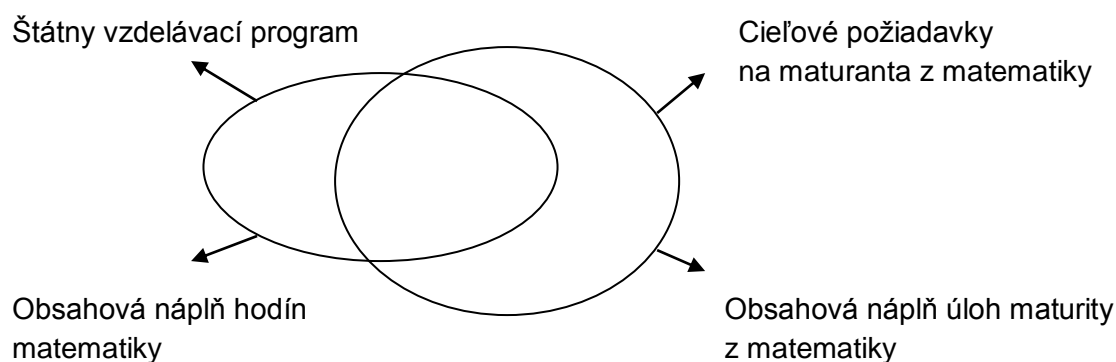
Každý z testov národných testovaní musí obsahovať veľmi obťažné i veľmi ľahké položky, pričom **väčšina položiek by sa mala pohybovať v rozpätí hodnôt stredne obťažnej položky.**

**„Žiaci prídu o body, ak pri hodnotení úloh s krátkou odpoveďou nemôže komisia doplniť kľúč.“**

Pri hodnotení odpovedí na úlohy s krátkou odpoveďou v testoch EČ MS z vyučovacích a cudzích jazykov je potrebné dodržiavať pokyny na hodnotenie a prácu s kľúčom správnych odpovedí. Ak je maturitná komisia pod vedením predsedu predmetovej maturitnej komisie presvedčená, že by do úvahy prichádzala aj iná odpoveď, než je uvedené, predseda PMK môže túto otázku konzultovať s odborníkmi NÚCEM-u, avšak **akékoľvek doplnenie kľúča a následná reoprava** (opakovaná oprava) **žiackych odpovedí sú zabezpečené centrálné**, aby sa predišlo situácii, v ktorej by niektoré školy mali odpoveď uznanú ako správnu a žiaci iných škôl by ju mali ohodnotenú ako nesprávnu. Žiaci určite o body neprídu, NÚCEM má mechanizmy a možnosti na to, aby **skontroloval všetky odpoveďové hárky**, vyselekoval hárky s odpoveďami vyžadujúcimi reopravu a samotnú reopravu aj uskutočnil. Kompletne kľúče správnych odpovedí sú aj z tohto dôvodu vždy zverejňované neskôr.

**„V teste EČ MS z matematiky boli úlohy, s ktorými sa žiaci na hodinách nestretli.“**

Obsah vyučovania predmetu matematika určuje štátny a školský vzdelávací program. Svojou náplňou je **zameraný na všeobecného žiaka**, ktorý sa po skončení strednej školy nebude ďalej venovať štúdiu matematiky na vysokej škole. Rozsah vedomostí a zručností požadovaných k maturitnej skúške z matematiky je stanovený Cieľovými požiadavkami na vedomosti a zručnosti maturantov z matematiky. Spomedzi predmetov, ktoré majú externú časť maturitnej skúšky, je špecifikum predmetu matematika v tom, že **rozsah cieľových požiadaviek je obsahovo rozdielny a oveľa väčší ako rozsah štátneho vzdelávacieho programu z matematiky** (obrázok 3). Preto sa žiaci na maturitnú skúšku z matematiky špeciálne pripravujú na voliteľných predmetoch v poslednom ročníku štúdia, kde si dopĺňajú vedomosti a zručnosti získané na základných hodinách matematiky o tie okruhy poznatkov, ktoré nie sú obsiahnuté v štátnom vzdelávacom programe matematiky pre jednotlivé druhy stredných škôl. **Absolvovanie základných hodín matematiky teda nepostačuje na úspešné zvládnutie maturitnej skúšky z matematiky** a žiak by mal pred maturitou z matematiky absolvovať aspoň 8 voliteľných hodín zameraných na matematiku, ktoré môže škola žiakovi poskytnúť v rámci disponibilných hodín.



Obr. 3 – Prienik rozsahu cieľových požiadaviek na maturanta a ŠVP z MAT

**„Test obsahoval niektoré veľmi ľahké úlohy, ktoré žiakov pomýlili, keďže v nich hľadali chyták.“**

Ako sme spomenuli vyššie, tento typ testov musí obsahovať aj náročné, aj ľahké úlohy. Ľahké úlohy v rozlišovacom teste rozlišujú žiakov slabších výkonnostných skupín a celkovo spĺňajú najmä **motivujúcu funkciu**, bývajú spravidla umiestnené medzi prvé položky

testu/časti testu, nie je potrebné hľadať v nich „chytáky“ a, ako sa pri nich často stáva, „prekombinovať“ riešenie.

### **„Texty ukážok v testoch EČ MS z vyučovacích jazykov sú príliš dlhé.“**

Test z vyučovacích jazykov a slovenského jazyka a slovenskej literatúry má stanovené v špecifikácii, aké typy textov bude obsahovať. Pre ilustráciu uvádzame, že napr. pri testoch zo SJL sa ani v pilotnom testovaní (pri ktorom majú žiaci rozhodne menšiu motiváciu než pri „ostrom“ testovaní) neukázalo, že by posledné úlohy testu zaznamenali **zvýšenú neriešenosť** (vynechanosť alebo nedosiahnutosť), ktorá by svedčila o tom, že test bol neprimerane dlhý vzhľadom na čas poskytnutý na jeho riešenie. Žiaci, ktorí absolvovali pilotné testovania, nemali problém s časovým limitom 90 minút a prečítaním ukážok, ktoré v teste EČ MS zo SJL 2015 **spolu obsahovali 1 397 slov**. Ak by sa však zvýšená neriešenosť v testovaní preukázala, autori by dĺžku ukážok prehodnotili. Ako vhodné porovnanie nám v tomto prípade poslúži i **navrhovaný rozsah textov v teste z druhého vyučovacieho jazyka (cudzí jazyk na bilingválnych školách, úroveň C1), ktorý predstavuje 3 350 slov** (do tohto počtu nerátame slová textov na počúvanie s porozumením), pričom na riešenie testu budú mať žiaci iba o tretinu viac času ako vo vyučovacích jazykoch. Aj z tohto porovnania vyplýva, že rozsah textov vo vyučovacích jazykoch a SJSL je primeraný.

Predmetoví koordinátori Národného ústavu certifikovaných meraní vzdelávania sa každoročne zúčastňujú na školeniach predsedov školských a predmetových maturitných komisií, ako aj na konferenciách k výučbe a testovaniu v jednotlivých predmetoch na rôznych stupňoch škôl. Všade zdôrazňujú odpovede na najpálčivejšie otázky, celkovo zhrnuté do dvoch vyjadrení, ktoré si dovoľíme na záver nášho článku uviesť.

Zo strany NÚCEM-u nejde o úmysel poškodiť žiakov pri testovaní, ubrať im body, poukazovať na to, čo nevedia a pod. No zároveň nie je v našom záujme znižovať nastavenú latku úrovne vzdelávania – nie je možné prispôbovať testy „momentálnej dispozícii“ testovanej populácie – v takom prípade by sme sa podieľali na tom, čo je pravým opakom zmyslu objektívneho testovania v modernej európskej spoločnosti, a *zvyšovanie kvality vzdelávania* by zostalo iba dobre znejúcim slovným spojením.

Snahou NÚCEM-u je každý rok pripraviť z jednotlivých predmetov kvalitné a zároveň dostatočne náročné testy, porovnateľné s európskymi štandardmi, a v tomto zámere budeme aj naďalej pokračovať.

NÚCEM