

Co říkají zdrojová maturitní data o hodnocení písemných prací

Oldřich Botlík

Zastánci společné části maturitní zkoušky z českého jazyka a literatury tvrdí, že obě tyto dílčí zkoušky ověřují, co je opravdu **důležité**. Druhým zásadním argumentem hájícím státní maturitu je **srovnatelnost** výsledků zkoušek jako důsledek toho, čemu říkají objektivizace zkoušky. Doložím pomocí maturitních dat, že není možné, aby oba argumenty platily u obou dílčích zkoušek současně. Nejspíš však neplatí ani u jedné.

Mým cílem je ilustrovat, čím mohou být maturitní data užitečná v diskusích o tom, co státní maturita žákům a učitelům dává, a co jim naopak bere. Posouvají totiž abstraktní diskusi do roviny konkrétních příkladů a kvantifikovaných, nezpochybnitelných argumentů.

Výsledky hodnocení písemné práce vykazují anomálii

Srovnatelnost výsledků hodnocení písemné práce znamená, že když dva maturanti stejně ovládají český jazyk, pak by měli mít stejné hodnocení své písemné práce. **A to samozřejmě bez ohledu na to, které z deseti zadání si každý z nich vybral.** Takové tvrzení je ovšem formulováno neurčitě, nepřipouští výjimky a netoleruje mírné rozdíly – proto nelze očekávat, že by mohlo platit absolutně.

Neplatí ovšem ani tehdy, když se spokojíme s poměrně velkou tolerancí. Vlastně se tomu nelze divit: **vágnost bodové škály** kritérií hodnocení¹ je v některých případech, ovšem nijak ojedinělých, až směšná. Například v kritériu 3A – *větná syntax, textová koheze* je vliv nedostatků na čtenářský komfort adresáta odstupňován takto:

0 bodů: *Nedostatky mají zásadní vliv na čtenářský komfort adresáta.*

1 bod: *Nedostatky mají vliv na čtenářský komfort adresáta.*

2 body: *Nedostatky mají občas vliv na čtenářský komfort adresáta.*

3 body: *Nedostatky mají místy vliv na čtenářský komfort adresáta.*

4 body: *Nedostatky nemají vliv na čtenářský komfort adresáta.*

5 bodů: *Čtenářský komfort adresáta není narušován.*

Zkusil jsem také – jako naprostý laik – zapátrat po tom, ve kterém kritériu má být penalizován žák, který se v referátu (například o českém národním obrození) dopustí věcné chyby (například zařadí mezi národní buditele Karla Čapka). Byl jsem poučen, že „tento typ chyby je možné zohlednit v kritériu 3B“², které se ovšem týká nadvětné syntaxe a koherence textu. **Nevyjasněnost příslušnosti možných chyb k určitému kritériu** je tedy dalším potenciálním zdrojem rozdílů. Velmi nápadná anomálie ve výsledcích se ovšem paradoxně týká kritéria 2A – *pravopis, tvarosloví a slovo tvorba*. Paradoxně kvůli tomu, že kromě hrátek se slovním vyjádřením míry vlivu chyb na čtenářský komfort adresáta, jejichž obdobou v kritériu 3A jsem se už zabýval, obsahuje toto kritérium **jako jediné** aspoň nějakou kvantifikaci prohrěšků, která určuje, kolik bodů na pětibodové škále má maturant dostat. Odstupňování podle počtu chyb ovšem nezávisí na délce textu, který maturant vytvořil.

1 viz [Přílohy, které jsou nedílnou součástí kritérií hodnocení maturitní zkoušky ve školním roce 2016/2017](#)

2 Baumgartnerová G., Kapustová A.: Metodický materiál pro hodnotitele písemných prací z českého jazyka a literatury, Centrum pro zjišťování výsledků vzdělávání, 2016 – str. 17

V čem zmíněná anomálie spočívá? V tuto chvíli jsou veřejně k dispozici různě podrobné výsledky (detaily uvedu níže) hodnocení písemných prací v jarních kolech státní maturity 2015, 2016 a 2017. V každém z těchto tří ročníků byla zadána také **dvě vypravování, která pokaždé skončila jako dvě nejhůře hodnocená zadání** v kritériu 2A – *pravopis, tvarosloví a slovo tvorba*³. Navíc, a to je v roce 2017 výjimečné, se výrazně lišily také výsledky obou vypravování. V kritériu 2A dosáhlo zadání 2 (vypravování s prvky sci-fi navazující na výchozí text) průměrného výsledku 2,29 bodu, zatímco zadání 8 (vypravování navazující na výchozí text) získalo 2,85 bodu. Podrobné výsledky v kritériích 2A a 2B za všechny tři roky obsahují Tabulky 1–3 na konci textu.

Anomálie ve výsledcích vypravování v roce 2017 a výsledky v testu

Za rok 2015 jsou k dispozici pouze agregovaná data, tedy průměrné výsledky hodnocení slohů podle zadání a kritéria hodnocení. Letos se ale poprvé podařilo dosáhnout zveřejnění úplně anonymizovaných prvotních dat za test ČJL i za písemnou práci z ČJL a to za roky 2016 a 2017. Údaje za test jsou v nich spojeny s údaji za sloh do jediného souboru. Jinak řečeno, jde o **data, která obsahují u každého žáka úplnou informaci** o tom, jak odpovídal v každé testové úloze, i o tom, které zadání písemné práce si zvolil a jakého výsledku dosáhl v každém z šesti kritérií hodnocení.

Nemyslím si, že by výsledek žáka v maturitním testu ČJL představoval příliš cennou informaci o jeho schopnosti používat český jazyk. Na druhé straně, je to zdaleka nejpodrobnější kvantifikovaná informace související s touto schopností, kterou máme k dispozici o každém žákovi, jenž psal maturitní sloh. Můžeme tedy – alespoň na chvíli – použít tuto informaci jako jistou míru žákovských schopností používat český jazyk a podívat se, zda rozdílné výsledky obou zadání nejsou například důsledkem toho, že zadání 2 (s horším výsledkem) psali v roce 2017 žáci s horšími schopnostmi používat český jazyk, než mají žáci, kteří psali zadání 8⁴.

Použil jsem dvě hlediska. Jedno pracuje s výsledkem za celý test, druhé jen s výsledkem za tzv. korektorské úlohy 3, 5, 8, 19, 28 a 32, které mají ke kritériu 2A nejbližší. Počítal jsem jen s těmi žáky, kteří měli průměrný výsledek (v celém testu, resp. v korektorských úlohách). V obou skupinách žáků s průměrným výsledkem zůstala v kritériu 2A obě vypravování výrazně nejhůře hodnocena a zůstal zachován také odstup zadání 2 od zadání 8. Jinak řečeno, i žáci se stejným výsledkem v celém testu, případně v korektorských úlohách, kteří psali vypravování, dosáhli v kritériu 2A průměrných výsledků, jež byly obdobně rozdílné a špatné jako výsledky za všechny žáky, kteří si tato zadání vybrali.

Co jsme tím zjistili? **Pokud vyjdeme z předpokladu, že test ČJL dává srovnatelné informace o tom, jak si maturanti osvojili znalosti a dovednosti důležité pro používání češtiny, stává se neudržitelným tvrzení o srovnatelnosti výsledků hodnocení písemných prací.**

Kde může být problém na straně maturitních slohů?

Zatím jsem se nezmínil o tom zásadním: tato dílčí zkouška podle mého názoru jen předstírá, že ověřuje, zda si žáci osvojili postupy potřebné k produkci kvalitních textů. Ve skutečnosti ověřuje, zda žáci dokážou napodobit jevovou stránku produkce kvalitních textů bez větších formálních vad. Nikdo se nenaučí napsat kvalitní reportáž nebo rozhovor tím, že si průběh události nebo setkání se zpovídanou osobou vymyslí a zaznamená své představy – takový postup se nejen neslučuje s novi-

3 Pro pořádek uvádím, že v roce 2017 dosáhl v kritériu 2A referát (zadání 6) stejného průměrného výsledku 2,85 bodu jako vypravování podle zadání 8. Referát psalo ovšem jen 33 žáků, zatímco zadání 8 psalo 11 839 žáků. V této fázi analýzy výsledků postrádá smysl zabývat se zadáními, která psalo tak málo žáků.

4 Totéž by bylo možné provést také pro data za rok 2016. Zatím jsem se k tomu nedostal.

nářskou etikou, ale především zcela pomíjí složité duševní činnosti autora, které vznik obou publicistických útvarů musejí provázet, má-li výsledek za něco stát. Ze stejných důvodů by ovšem nemělo být například psaní reportáže či rozhovoru tímto způsobem ani ověřováno. Dalším zjevným nedostatkem celého pojetí této dílčí zkoušky je nízká časová dotace a s ní spojený obvyklý rozsah vytvářených textů.

Některé problémy kritérií hodnocení jsem zmínil, případně naznačil už na začátku. Pokud jde o souhrnné výsledky roku 2017, obsahuje je Tabulka 4. Uvedených pět zadání reprezentuje 58 295 žáků z celkem 61 415 – zbývajících pět zadání můžeme na této úrovni zkoumání ignorovat. Je zcela zřejmé, že letos bylo, z hlediska naděje na dobré hodnocení, zdaleka nejvýhodnější konvertovat oznámení na zprávu (zadání 10) – také jeden ze zcela formálních a nereálných úkolů.

Pokud jde o příčiny anomálie u obou vypravování v kritériu 2A, konzultoval jsem problém s několika středoškolskými učiteli. Shodli se na tom, že napsat kvalitní krátké vypravování je obtížné. **Tento útvar tedy nejspíš bývá v maturitních písemných pracích delší než útvary ostatní, takže žáci dopouštějící se občas chyb posuzovaných v kritériu 2A jich nejspíš udělají víc.** Kritérium 2A je ovšem absolutní – délku vzniklého textu nebere v úvahu. Zadání 2 ale dopadlo výrazně hůře než zadání 8 právě jen v kritériu 2A – v ostatních kritériích jsou rozdíly nepatrné. **Byly tedy texty podle 2 výrazně delší než texty podle 8, anebo je problém ještě v něčem dalším?**

Několik mých konzultantů uvedlo, že pro hodnotitele (i pro ně osobně) bývá **problém penalizovat jednu chybu jen v jediném kritériu.** Zmiňovali tendenci pohlížet celkově přísněji na práci, která je v něčem dílčím zklamala (například na vypravování, které neobsahovalo dostatečně přesvědčivé prvky sci-fi), a naopak celkově vstřícněji na práci, která se jim „líbí“. Vágnost, gumovost a nevyjasněnost kritérií hodnocení to samozřejmě usnadňuje.

A konečně, další možné vysvětlení „propadu vypravování v kritériu 2A“ souvisí s rysem zadání, který pracovně nazývám **frekvence pravopisných pastí.** Třeba ve zprávě (zadání 10) je ta frekvence poměrně nízká – výchozí oznámení vlastně dává například návod, jak psát slova, v nichž by slabší žáci nejspíš mohli udělat chybu (třeba časové údaje, „kánoích“, „grilování“, „regionální“ apod.). Vypravování naopak obsahuje pasti v podobě přímých řečí s uvozovací větou uprostřed či na konci. Pastí je také možnost používat nespisovnou češtinu – žáci nemusejí uhlídat, aby se vyskytovala jen v promluvách jednajících postav, a někteří hodnotitelé ji zase mohou (nesprávně) penalizovat i tam.

Uvědomuji si, že „počítání s různými průměry“ je příliš hrubým analytickým nástrojem – přesto upozornilo na zásadní koncepční chybu v konstrukci kritéria 2A. Data, která jsou k dispozici, lze zkoumat i jemnějšími nástroji – když ovšem analytik netuší, co přesně má hledat, je to časově dost náročné. Proto **dávám zájemcům k dispozici interaktivní tabulku s výsledky všech zadání ve všech kritériích hodnocení,** a to za tři poslední jarní maturitní kola. Ti, kdo více rozumějí souvislostem mezi kritérii hodnocení a mají bohaté zkušenosti se slohovými pracemi středoškoláků, si tam nejspíš povšimnou rovněž dalších „podezřelých“ míst. Rád se jejich detailnější analýze budu věnovat.

V průměrech se také leccos ztratí: když například jeden hodnotitel soustavně hodnotí přísně, zatímco druhý soustavně hodnotí mírně, průměr za oba přestává vybočovat z normálu, ačkoli každý hodnotitel jednotlivě z normálu vybočuje. O přesnosti hodnocení jako takové nevíme vůbec nic a já se obávám, že o ní téměř nic neví ani Cermat. Bylo by možné ji odhadnout (a porovnat tak například přesnost centrálního hodnocení s přesností hodnocení na školách) na základě „slepého“ druhého hodnocení dostatečně reprezentativního vzorku prací (z hlediska zadání, škol, na kterých vznikla, i z hlediska kvality) dostatečně reprezentativním vzorkem hodnotitelů. Z důvodů, které uvedu v závěru, si ovšem nemyslím, že k tomu někdy dojde. Navíc je zřejmé, že na „uctívání posvátné krávy srovnatelnosti“ doplácíme ve věcech, které jsou mnohem důležitější.

Kde může být problém na straně testů?

Účelem závěrečné zkoušky a maturitní zkoušky je ověřit, jak žáci dosáhli cílů vzdělávání stanovených rámcovým a školním vzdělávacím programem v příslušném oboru vzdělání, zejména ověřit úroveň klíčových vědomostí, a dovedností žáka, které jsou důležité pro jeho další vzdělávání nebo výkon povolání nebo odborných činností⁵.

(§ 73 školského zákona)

Respektování citovaného ustanovení školského zákona nepatří podle mého názoru k nejsilnějším stránkám maturitního testu z ČJL. Uvedu několik příkladů.

Již zmíněné **úlohy ověřující tzv. korektorské dovednosti** jsou tradiční součástí testu. V roce 2016 šlo o úlohy 1, 8, 11, 16, 17, 24 a 32 s celkovou dotací 10 bodů⁶, v roce 2017 to byly úlohy 3, 5, 8, 19, 28 a 32 s celkovou dotací 9 bodů. Žáci v nich měli odhalovat různé jazykové chyby v *cizích* textech. RVP SOŠ ani RVP G ovšem podobné specializované korektorské dovednosti nepožadují – podle nich se žák nemá dopouštět jazykových chyb ve *vlastních* textech. Jde samozřejmě o rozdílné dovednosti. Například v roce 2016 měli žáci v úloze 8 najít na patnácti (!) řádcích textu o Ezopovi čtyři slova zapsaná s pravopisnou chybou (a zapsat je správně). Ačkoli nešlo o žádné chytáky (v historce se pravý; zásluhou kusích informací, nezapomělo se, zkončil v trezoru), bez jediné chyby se úkolu zhostilo pouze 11,4 % žáků. I kdyby šlo výhradně o gymnazisty (informaci o typu školy data neobsahují), byla by to jen třetina ze všech maturantů na gymnáziích. Problém je přitom snadno řešitelný technicky: stačí například, aby žáci posuzovali různé varianty psaní (například „v historce se pravý/praví“ a odpovídali ANO, pokud je správná verze uvedena na prvním místě – zde by tedy správnou odpovědí bylo NE). Nabídnutých dvojic variant přitom může být samozřejmě více.

S dalším příkladem, **přísudkem jmenným se sponou**, možná u některých čtenářů nepochodím, ale odpustit si ho nedokážu. Ve stejném testu měli žáci v podúloze 22.2 najít přísudek jmenný se sponou v těchto větách z Burgessova Mechanického pomeranče. *Vypadal dost vypuganě, když nás čtyři maníky zlukoval, jak ho tak tiše obestupujeme, ale řekl: „Ano? O co jde?“ Promluvil strašně nahlas učitelským vojsem, jako by nám chtěl ukázat, že není vypuganej.* Také zde byla úspěšnost žáků podobně „tragická“ – opět platí, že kdyby uspěli pouze gymnazisté, stejně by šlo jen o každého třetího maturanta z gymnázia. Výsledek ovšem tentokrát chápu také jako informaci, co si žáci myslí o užitečnosti tohoto odborného pojmu pro sebe. Stejně jsem výsledek interpretoval také v jedné internetové diskusi a byl jsem dost neurvale poučen, jak užitečný je tento pojem například pro správné psaní shody přísudku s podmětem. Aby nedošlo k nepochopení: já sám jsem dokázal spojení „není vypuganej“ identifikovat bez nejmenších potíží. Vzpomněl jsem si ale na maturitní test z jara 2013, v němž Cermat ponechal v úloze zaměřené na porozumění textu, tedy nikoli na pravopis, větu *Cesare a Lucrezia byly děti papeže Alexandra VI.* Byl jsem tehdy jediný, kdo na chybu upozornil – přehlédl ji i J. Kostečka ve své recenzi testu.⁷ Nejspíš to tedy s užitečností (pro správné psaní – nikoli pro následné odůvodňování) nebude tak žhavé. Problém podobně koncipovaných úloh spočívá v tom, že kladou nepřekonatelnou *odbornou* překážku před žáky, kteří by možná větu napsali správně (třeba díky pomůcce „byli dětmi“).

Snad je zřejmé, že podobných příkladů *odborných* překážek zbytečně hozených žákům pod nohy se dá v maturitních testech najít mnohem více. Patří k nim také již proslulá úloha o epizeuxi. Podíl maturantů, kteří budou podobný termín zásadně potřebovat ve svém dalším vzdělávání, je naprosto

5 Doslovné znění příslušného ustanovení, včetně interpunkce.

6 Na problém upozornili autoři materiálu *Státní maturita porušuje zákon a poškozuje žáky i výuku ve školách*, který je ke stažení na adrese [Maturitní data – odtajněno](#), a to na konci sekce **Ke stažení** (na kartě Analytický materiál k testům z jara 2016).

7 <http://www.ascestinaru.cz/jiri-kostecka-recenze-maturitniho-testu-jaro-2013/>

zanedbatelný – jde nanejvýš o několik promile z celkového počtu žáků. Na druhé straně by bylo jistě dobře, kdyby se všichni maturanti dokázali zamýšlet nad účinkem, kterého autor básně hodlal dosáhnout opakováním výrazů ve verši nebo v různých verších. A kdyby ten účinek při čtení dokázali vnímat. K tomu ale žáci pojmy jako epizeuxis, anafora apod. primárně nepotřebují.

Použil jsem – na krátkou chvíli – výsledky žáků v testu jako měřítko jejich schopnosti používat český jazyk v písemné práci. Nyní jsou zřejmé přinejmenším dva důvody, proč to nemůže fungovat. Za prvé, významnou část maximálního bodového zisku v testu mohli žáci získat jedině tehdy, když správně vyřešili úlohy zaměřené na něco, co s vytvářením kvalitních textů příliš nesouvisí. Za druhé, takto koncipovaný test je silně nespravedlivý vůči negymnazistům, mezi nimiž se budoucí lingvisté nevyskytují vůbec. Tito žáci, i vinou nesprávné výuky na základní škole, nejen nezvládli teorii natolik, aby ji mohli při tvorbě vlastních textů správně užívat, ale často ani nechápou, čím jim může být užitečná. Koncept srovnatelnosti výsledků testu je proto udržitelný jedině tehdy, když přistoupíme na to, že by snad mohl platit při srovnávání mezi gymnazisty. A potom, odděleně a na nesmyslně nastavené úrovni, při srovnávání mezi negymnazisty. **Daň, kterou za takovou „objektivizovanou srovnatelnost“ platí výuka češtiny na všech typech škol, je přitom neúnosná.**

Hodnocení slohů: na školách, nebo v centru?

Nechci na tu otázku odpovídat, protože si myslím, že se především mají psát jiné „slohy“ a mají být také jinak hodnoceny. Mezititulek je mým oslím můstkem k novele školského zákona, která vrátila hodnocení písemných prací do centra. Chci na tom ilustrovat ještě jedno užití maturitních dat.

Podle důvodové zprávy ke zmíněné novele *došlo v letech 2013 a 2014 k nárůstu počtu písemných prací maturujících žáků, které byly učiteli z kmenových škol hodnoceny takovým nejnižším možným bodovým hodnocením, aby žák ještě u této dílčí zkoušky uspěl a nemusel zkoušku opakovat*. Důvodová zpráva uvádí, že na kontrolním vzorku z roku 2014 vzrostla po přehodnocení neúspěšnost z hodnoty 1,8 % na 10,1 %. To byl také **jediný** oficiální argument pro návrat k centrálním hodnotitelům. Důvodová zpráva ovšem nepodává žádnou informaci o výběru hodnotitelů, kteří prováděli přehodnocení, ani o instrukcích, které hodnotitelé dostali. Nepopisuje dokonce ani velikost vzorku a jeho další vlastnosti. Předkladatel zákona přesto argument použil a naši zákonodárci ho přijali.

Korektní experiment s přehodnocením by měl dodržet mimo jiné pravidla uvedená pod čarou⁸. Nemám data z roku 2014 k dispozici, ovšem analyzoval jsem data z let 2016 a 2017. Výsledky zachycuje Tabulka 5. Kdyby měl argument z důvodové zprávy platit i v roce 2016 pro celý soubor (tj. nejen pro blíže nespecifikovaný vzorek), museli by hodnotitelé na školách nadřžováním zachránit před propadnutím 8,75 % všech žáků. To znamená, že by „správně“ mělo propadnout 6 843 žáků namísto „pouhých“ 833 žáků. Absurdnost argumentace dokreslují výsledky práce centrálních hodnotitelů v roce 2017. Navzdory nijak neskrývané snaze odstranit „schod“ mezi výsledky 11 a 12 bodů se ho podařilo jen mírně snížit. Přesto zůstal zcela zjevný, což Jiří Zíka, ředitel Cermatu, komentoval v MF Dnes slovy: „Není žádný výrazný nárůst neúspěšnosti, kterého se někteří hrozili. Hodnotitelé jsou lidé, když mají šanci, studentovi přidají.“ Toleroval tedy centrálním hodnotitelům to, kvůli čemu důvodová zpráva hodnotitele na školách kritizovala. Počet neúspěšných žáků vzrostl třikrát.

⁸ Hodnotitelé provádějící přehodnocení nesmějí vědět, co je cílem experimentu. Nesmějí tudíž obdržet k přehodnocení žádné jiné instrukce než ty, které mají k dispozici při hodnocení běžném. Cíl experimentu nesmějí odhalit ani z převládající úrovně hodnocených písemných prací. Proto musejí dostat k přehodnocení jak práce slabé, tak práce průměrné i nadprůměrné. Práce musejí být vybrány náhodně a vzorek má být dostatečně velký, aby bylo možné zobecnit získaný výsledek na celý soubor. Zhruba šesti-násobný nárůst neúspěšnosti se totiž může projevit i na vzorku pouhých deseti prací. Náhodný musí být i samotný výběr hodnotitelů. Mezi hodnotiteli na školách byli i lidé, kteří se veřejně zasazovali o návrat centrálního hodnocení. Pokud by převládli v týmu, který prováděl přehodnocení, už jenom kvůli tomuto nedostatku by nešlo objektivitu celého experimentu úspěšně hájit.

Závěr

Podle mého názoru je dnes již zcela zřejmé, že státní maturita nebyla připravena s cílem spravedlivě měřit důležité výsledky vzdělávání. Jde o projekt, který měl především uspokojit osobní ambice velmi úzkého okruhu lidí, jež ani nechci nazvat odbornými. Druhým cílem, který projekt sledoval a nadále sleduje, je odradit významnou část žáků ukončujících základní vzdělávání od studia maturitních oborů. Třetím, který se dá vytušit, je – přinejmenším u části podporovatelů a centrálních hodnotitelů – vytvořit si zdroj dalších příjmů.

Argumentovat za těchto okolností jinak než „tvrdými“ daty mi připadá jako házení hrachu na zeď. Sám jsem se před více než dvěma lety rozhodl jít cestou tlaku na zveřejňování maturitních dat. Na začátku mi trvalo téměř tři čtvrtě roku, než jsem z Cermatu vydoloval díky zákonu č. 106/1999 Sb., o svobodném přístupu k informacím, souhrnné výsledky jedné jediné testové úlohy. V tuto chvíli už jsou na webu částečně anonymizovaná data prvotní data z jednotných přijímacích testů, která obsahují informaci o typu školy, na niž se uchazeč hlásí.

Žádal jsem, aby data obsahovala také informaci o pohlaví uchazeče, protože je důležitá. Ministerští úředníci (ve skutečnosti ovšem spíše Cermat) tuto informaci zatím vypustili, a to navzdory příkazu ministra školství. Musím tedy podat novou stížnost na porušení zákona. Ovšem odpor ke zveřejňování důležitých informací také naznačuje, že jsem našel slabé místo: data nejspíš umožní odhalit nedostatky, jejichž rozsah si mnozí ani nedokážou připustit a představit.