

Instruktažní materiál k testům kognitivních schopností a pre-gramotnosti aplikovatelných v sociologických výzkumech dospělé a dětské populace na 1. stupni ZŠ (on-line tutoriál)

Jiří Šafr, Magdaléna Gorčíková

Sociologický ústav AV ČR, v.v.i.

Abstrakt

Instruktažní materiál k psychometrickým testům (dostupný on-line) představuje vybrané nástroje pro měření kognice uplatnitelné v sociologických výzkumech typu survey v dospělé a dětské populaci. Nejprve krátce uvádíme teorii a metodologii měření kognice, vč. možnosti skórování výsledků v testu, posléze představujeme námi optimalizované nástroje měření kognitivních schopností u dospělých (adaptace testu WST-10A z amerického prostředí) a dětí (v oblasti slovní zásoby, vybavování si z paměti a rovněž verzi dWST-12A pro děti). Použité postupy optimalizace a validizace jsou uvedeny v příloze. Pro využití ve výzkumech dětí v první třídě (cca 6–7 let) uvádíme neoptimálnější test k měření pre-gramotnosti, kterým je měřítko (pasivní) slovní zásoby VOC, včetně jeho zkrácené verze. Podrobněji se věnujeme problematice užití nástrojů měření kognice a pre-gramotnosti v praxi a upozorňujeme, že tyto testy jsou určeny pouze k použití ve výběrových šetřeních (survey) k měření rozdílů mezi skupinami, nikoliv pro individuální diagnostiku žáka v poradenském zařízení.

Tento text vznikl v rámci vývoje a adaptace měřicích nástrojů (KA2.2) v projektu CSDA Research projektu CSDA Research – Výzkumný program Českého sociálněvědního datového archivu: Česká republika v Mezinárodním programu sociálních šetření ISSP, výzkum kvality dat a zdrojů dat (MŠMT ČR, evid. č. CZ.02.1.01/0.0/0.0/16_013/0001796).



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Obsah

Abstrakt.....	1
Obsah	2
1. Úvodní poznámky k metodologii měření	3
1a. Obecná východiska měření pomocí testů	3
2. Postup konstrukce testového skóre v datech – praktické rady a ukázky	4
3. Test kognitivních schopností Wordsum (WST-10A) – verze pro dospělé	9
4. Vybrané testy k měření pre-gramotnosti a kognitivních schopností v populaci dětí na prvním stupni základní školy	14
4a. Testy k měření pre-gramotnosti – kognitivních a jazykových schopností u dětí na počátku školní docházky (6–7 let).....	14
4b. Test kognitivních schopností Wordsum (dWST-12A) – verze pro děti (10-11 let).....	18
5. Použitá literatura	21
6. Přílohy	22
6a. Použitá data	22
6b. Zkrácené verze testu pre-gramotnosti VOC pro žáky 1. třídy	23

1. Úvodní poznámky k metodologii měření

Cílem tohoto tutoriálu je představení krátkých dotazníkových nástrojů měřících: 1) adaptovaný a validizovaný nástroj měření kognitivních dovedností dospělých (WST-10A) a 2) kognitivní schopnosti (dWST-12A) a ranou gramotnost u žáků nastupujících do základní školy (věk cca 6 let).¹

V sociálních vědách je stále větší pozornost věnována roli kognitivních schopností a úrovně gramotnostních dovedností v chování, rozhodování a dosahování životního úspěchu. K empirickému testování teorií příčin a důsledků bylo navrženo nepřeberné množství testů různých aspektů kognitivních a jazykových dovedností. Většina těchto testů vyvinutých v psychometrii a majících kořeny v klasické teorii měření (KTM) představuje poměrně rozsáhlé testové baterie, které jsou časově náročné a organizačně obtížné a při administraci často navíc kladoucí nároky na detailní proškolení administrátora testu. Pro sociologická šetření jsou pak tyto nástroje zcela nevhodné (délka, náročnost testování apod.). Takové nástroje dosud v českém výzkumném prostředí naprosto chyběly, přitom jsou velmi žádoucí, neboť ukazatele gramotnosti a kognitivních dovedností dětí a dospělých jsou důležitým prediktorem životních trajektorií, využití však nalézají i ve výzkumu dalších témat. Adaptovaný a validizovaný nástroj měření kognitivních dovedností dospělých v budoucnosti může sloužit dalším výzkumníkům, kteří budou ve výběrových šetřeních zjišťovat kognitivní schopnosti a jejich vztah např. k dosaženému vzdělání, reportovaným příjmům, pracovní pozici, zdraví a kvality života/ pocit štěstí (well-being). Dodejme, že tyto nástroje jsou určeny k měření distribuce hodnot, resp. relativního pořadí z hlediska kognitivních schopností v obecné populaci, tedy nikoliv pro diagnostiku kognitivních a jazykových schopností jednotlivců. Východiskem je rychlá a jednoduchá administrace testu.

1a. Obecná východiska měření pomocí testů

Vyhodnocení výsledků z psychometrických testů vyžaduje takové nástroje měření, které jsou platné (validní), spolehlivé (reliabilní) a citlivě reagující na změny (responsivní). Přesnost měření výsledku však určuje platnost a spolehlivost použitého nástroje.

Klasická testová teorie psychometrického testování

Klasická testová teorie vývoje měřítek vychází z toho, že skutečné skóre dané osoby je neznámé. K odhadu skutečného skóre se každé osobě předkládá sada dílčích položek a na základě jednotlivých odpovědí na položky se vypočítá pozorované skóre. Díky chybě měření se ale toto pozorované skóre nerovná skutečnému skóre. Pro ideální nástroj by tato chyba měla být stejná co nejmenší. Původ této chyby je různý. Jedním jejím zdrojem je situace, kdy skóre odvozené z odpovědí na položky nereprezentuje přesně zamýšlenou vlastnost. Určení této složky chyby je otázkou testování validity. Druhým zdrojem chyby měření je náhodný rozptyl v odpovědích na položku, které se vyhodnocují testováním reliability.

Jedním z typů validity je kritériální validita, tu dosahuje nástroj tehdy, když skóre odvozené z nástroje odrážejí latentní (nepozorovatelný) konstrukt. V praxi je nejčastěji tato forma validity posuzována hodnocením velikosti korelací mezi skóre získanými z různých nástrojů. Důkazy o konstruktové validitě také lze dovodovat ze schopnosti nástroje rozlišit mezi dvěma nebo více skupinami osob, o nichž je obecně známo, že se v měřené vlastnosti liší.

¹ Nástroje jsou na vyžádání k dispozici v CESDA SOÚ AV ČR.

2. Postup konstrukce testového skóre v datech – praktické rady a ukázky

Při konstrukci hodnot získaných z měření v testu lze postupovat v zásadě dvěma postupy, buď na základě klasické teorie měření (KTM) nebo na základě postupů odvozených z teorie měření na položku (IRT). Vzhledem k tomu, že druhý zmíněný směr vyžaduje pokročilé znalosti a specializovaný software, budeme se v následujících příkladech věnovat především postupům vycházejícím z klasické teorie měření.

Postupů založených na KTM je vícero, v zásadě je lze rozdělit na ty, které jednotlivé položky považují za rovnocenné, tj. mající stejnou váhu, a na ty které, podobně jako postupy IRT, nějakým způsobem zohledňují obtížnost jednotlivých položek testu.

Především v případě postupu založeného na KTM pak většinou ještě následuje další krok, kterým je standardizace hrubých skóre či jakési „přeškálování“. To lze provést buď v závislosti na teoretických hodnotách odvozených z obecné populace získaných na základě hodnot z jiného reprezentativního výzkumu, ve kterém byly vytvořeny tzv. normativní hodnoty (většinou pro určité homogenní segmenty populace – strata, nejčastěji jde o věkové kohorty), nebo na pozici ve vztahu pouze ke zkoumané populaci (jako například rozdělení na percentily nebo percentilové pořadí – percentile ranks), které umožňují lepší interpretaci dosažených výsledků u jednotlivých případů (respondentů/subjektů).

Pokud k přístupu ke konstrukci skóre skrze klasickou teorii měření (KTM), pak je možno uvažovat od dvou přístupů k měření: hrubé skóre jako nevážené skóre z testu (v principu jde o součet správných odpovědí, o tom podrobněji níže) anebo určit pro každého respondenta/probanda skóre jako vážené, a to v principu podle obtížnosti jednotlivých položek. Zde je více možností jak na to, jednou je kupříkladu vážení podle podílu správných odpovědí u té které položky, jinou pak například vážení dané faktorovou zátěží u jednotlivých položek získaných v pomoci faktorové analýzy nebo PCA.

Konstrukci testového skóre si ukážeme na příkladu testu kognitivních schopností dWST-12A pro děti, v němž respondent/proband vybírá správnou odpověď ze čtyř možných alternativ, přičemž pouze vždy jen jedna je správná. V technické rovině spočívá příprava dat v zakódování správné odpovědi na hodnotu 1 a nesprávné na 0 (pokud bychom ale chtěli zjišťovat i jiné vlastnosti položek testu než jen obtížnost, např. odhadnutelnost odpovědi, pak musíme nejprve zakódovat jakoukoliv odpověď – proměnná pak nabývá hodnot 1–4 a využít postupy IRT).

Nejjednodušším postupem je nevážené skóre z testu, které lze získat v zásadě dvojnásobným způsobem. Buď jako součet správných odpovědí (tj. prostá suma, případně jako načítání, kdy je za každou správnou odpověď přiděleno/přičteno číslo 1, což je ale postup vyžadující důslednou kontrolu tzv. missingů), nebo, a to je namnoze lepší postup, jako průměr z řady všech položek testu, neboť automaticky řeší problém, pokud u jednotlivých případů chybí hodnoty u některých položek (missing) například z důvodu vynechání/ přeskočení otázky žákem. Většinou ovšem takové případy klasifikujeme jako špatné odpovědi (hodnota 0). Pokud využijeme pouze sumaci bez další kontroly pro ty, kteří test aktuálně nedělali nebo pro ně z nějakého důvodu nemáme data, pak se může stát, že postupu pomocí načítání může takový případ získat skóre nula, což ale není skutečná hodnota výsledku v testu (tu skutečně získá jen ten, kdo se testu prokazatelně zúčastnil a nezodpověděl ani jednu položku správně).

Než přistoupíme k samotné konstrukci proměnné měřící výsledek v testu u každého případu a její případné standardizaci, je třeba ověřit kvalitu měření – tzv. psychometrické parametry výsledného skóre z testu. Postup konstrukce proměnné měřící výsledky v testu a ověření jejich vlastností z hlediska kvality měření uvádí následující tabulka 1.

Tabulka 1. Postup konstrukce testového skóre a ověření kvality měření, KTM.

1. kontrola a případné vyčištění dat (odstranění nelogických hodnot)
2. zakódování položek testu na hodnoty 0/1
3. vytvoření filtru pro ty, kteří test z nějakého důvodu nedělali (touse)
4. ověření konzistence testu, např. pomocí koeficientu položkové reliability
5. ověření homogenity testu, např. pomocí explorační faktorové analýzy nebo PCA
6. případná kalibrace jednotlivých položek (vynechání) za účelem maximalizace reliability
7. samotná konstrukce latentní proměnné měřící výkon v testu, např. pomocí sumace do hrubého skóru nebo zprůměrování položek kódovaných 0/1. Následovat může ještě také jeho standardizace za zkoumanou celou populaci (výzkumný vzorek) např. na percentilové rozpětí.
8. ověření a kontrola hodnot (střední hodnoty, rozptyl, rozložení hodnot). Následovat by měly ještě další kroky ověřující psychometrické atributy latentní proměnné z testu.
9. ověření homogenity nebo heterogenity měření v populaci a v relevantních subpopulacích, např. rozložení hodnot podle pohlaví. Ptáme se: je tvar rozložení hodnot stejný u dívek a u chlapců? Jsou střední hodnoty stejné, anebo pokud je mezi nimi rozdíl, je v souladu s teoretickým očekáváním? Nedosahují maximální či naopak minimální hodnoty v jednotlivých subpopulacích kritických hodnot stropu a podlahy (tzv. ceiling a flooring effect).
10. ověření validity a to konstruktové (vztah k paralelně měřeným konceptům) nebo prediktivní (predikce výsledků v hodnotách jiné závislé proměnné v souladu s teorií).
11. pokud nová proměnná vykazuje parametry kvality měření (reliabilita, diskriminační schopnost, homogenita průběhu měření v subpopulacích), pak můžeme přikročit k standardizaci hodnot pro jednotlivé případy tak, abychom mohli snáze interpretovat výsledky za různé podskupiny, například podle věkových kohort (např. na percentile-rank).

Prvních osm bodů konstrukce hrubého skóre z testu si ukážeme na příkladu testu dWST-12A – Wordsum měřeném v 4. vlně děti Panel ZŠ. Jeho specifikem je, že se jedná o test, kde je na výběr z více odpovědí. Uvedené kroky přípravy testu demonstrujeme s využitím statistického programu Stata (veze 16) a uvádíme kódy v příkazovém řádku (jednotlivé kroky tvoří sekvenci zaznamenanou v tzv. dofile). Pro zachování anonymity testu jsou v příkladu použita jiná slova než ve skutečném testu.

1. Nejprve výpis části dat (10 žáků, kteří se zúčastnili čtvrté vlny), zobrazeno je jen osm z dvanácti položek.

```
. list ID d4S01 d4S02 d4S03 d4S04 d4S05 d4S06 d4S07 d4S08 ///
      if d4ucastT4==1 in 1/35, sepby(d4ucastT4) noobs
```

ID	d4S01	d4S02	d4S03	d4S04	d4S05	d4S06	d4S07	d4S08
7	pole	drbat se	vypít	koupit si	mručet	bezpečný	batoh	síla
9	pole	drbat se	vypít	zlomit	mručet	slušný	batoh	zloba
14	pole	dloubat se	vypít	koupit si	mručet	bezpečný	batoh	slušnost
15	pole	drbat se	vypít	koupit si	nutit	bezpečný	batoh	slušnost
17	pole	dloubat se	sníst	koupit si	matlat se	bezpečný	úsměv	zloba
24	pole	drbat se	sníst	koupit si	mručet	bezpečný	úder	slušnost
25	pole	dloubat se	vypít	zlomit	mručet	bezpečný	batoh	síla
28	pole	dloubat se	vypít	užít	mručet	slušný	batoh	nenávisť
30	pole	drbat se	vyprat	koupit si	mručet	bezpečný	úsměv	slušnost
33	pole	drbat se	vypít	užít	mručet	bezpečný	výhra	nenávisť

2. Přirazení hodnoty 1 při správné odpovědi, 0 nesprávné podle klíče testu.

```
. recode d4S01 (5 = 1) (nonmiss =0), gen(d4wst01)
. recode d4S02 (4 = 1) (nonmiss =0), gen(d4wst02)
. recode d4S03 (3 = 1) (nonmiss =0), gen(d4wst03)
. recode d4S04 (2 = 1) (nonmiss =0), gen(d4wst04)
. recode d4S05 (1 = 1) (nonmiss =0), gen(d4wst05)
. recode d4S06 (1 = 1) (nonmiss =0), gen(d4wst06)
. recode d4S07 (2 = 1) (nonmiss =0), gen(d4wst07)
. recode d4S08 (3 = 1) (nonmiss =0), gen(d4wst08)
. recode d4S09 (4 = 1) (nonmiss =0), gen(d4wst09)
. recode d4S10 (5 = 1) (nonmiss =0), gen(d4wst10)
. recode d4S11 (2 = 1) (nonmiss =0), gen(d4wst11)
. recode d4S12 (1 = 1) (nonmiss =0), gen(d4wst12)
```

```
. lab def d4wst 0 „spatne“ 1 „spravne“ .d NEDELAL TEST, modify
. lab val d4S01-d4S12 d4wst
. format d4wst01-d4wst12 %8.2f // formát zobrazení
. desc d4wst01-d4wst12
. tabstat d4wst01-d4wst12 , stats(n mean min max) column(s) f
```

Výsledkem je tabulka, v níž v prvním kroku zkontrolujeme počet validních případů a rozsah minimálních a maximálních hodnot.

	N	mean	min	max
d4wst01	219.00	0.91	0.00	1.00
d4wst02	219.00	0.30	0.00	1.00
d4wst03	219.00	0.81	0.00	1.00
d4wst04	219.00	0.73	0.00	1.00
d4wst05	219.00	0.47	0.00	1.00
d4wst06	219.00	0.75	0.00	1.00
d4wst07	219.00	0.30	0.00	1.00
d4wst08	219.00	0.28	0.00	1.00
d4wst09	219.00	0.34	0.00	1.00
d4wst10	219.00	0.79	0.00	1.00
d4wst11	219.00	0.55	0.00	1.00
d4wst12	219.00	0.41	0.00	1.00

3. V případě, že z nějakého důvodu chceme vytvořit celkové skóre z testu pouze pro ty případy, kdy respondent odpověděl na všechny položky testu, resp. (a to především – test z nějakého důvodu vůbec nedělal) je nejprve potřeba vytvořit filtr pro ty, kteří některé z položek nedělali. Zde je nicméně třeba rozvaha jak s nezodpovězenými položkami zacházet, nejčastěji je totiž v případě nezodpovězení označujeme jako nesprávnou odpověď a kódujeme je tak jako 0 (k opravdu jednoduchému vytvoření příslušného filtru, zde pojmenovaného `toused4wst`, je vhodné v programu Stata mít nainstalovaný uživatelský stata modul `marktouse`; `ssc install nmissing`).

```
marktouse toused4wst d4wst01-d4wst12, label("Filtr: Nonmissing d4wst01-d4wst12")
fre toused4wst
```

```
toused4wst -- Filtr: Nonmissing d4wst01-d4wst12
```

	Freq.	Percent	Valid	Cum.
Valid 0	227	50.90	50.90	50.90
1	219	49.10	49.10	100.00
Total	446	100.00	100.00	

4. Před samotnou konstrukcí testového skóre přikročíme k ověření vnitřní konzistence testu prostřednictvím výpočtu položkové reliability pomocí koeficientu KR-20 pro dichotomické položky (zobecnění koef. Alfa) a provedeme případné vyloučení nekonzistentních položek.

```
. alpha d4wst01-d4wst12, item label std
```

Test scale = mean(standardized items)

Item	Obs	Sign	item-test corr.	item-rest corr.	interitem corr.	alpha	Label
d4wst01	219	+	0.3093	0.0882	0.0635	0.4272	
d4wst02	219	+	0.3515	0.1337	0.0601	0.4130	
d4wst03	219	+	0.3750	0.1597	0.0582	0.4047	
d4wst04	219	+	0.3862	0.1721	0.0573	0.4007	
d4wst05	219	+	0.4545	0.2494	0.0518	0.3753	
d4wst06	219	+	0.2573	0.0333	0.0677	0.4441	
d4wst07	219	+	0.3596	0.1426	0.0594	0.4101	
d4wst08	219	+	0.3790	0.1641	0.0579	0.4033	
d4wst09	219	+	0.3924	0.1790	0.0568	0.3984	
d4wst10	219	+	0.3430	0.1246	0.0608	0.4159	
d4wst11	219	+	0.3689	0.1528	0.0587	0.4069	
d4wst12	219	+	0.4656	0.2622	0.0509	0.3710	
Test scale					0.0586	0.4275	mean(standardized items)

Koeficient reliability Alfa resp. zde pro dichotomické proměnné KR-20 dosahuje za testové skóre se všemi 12 položkami 0.43, což je sice hodnota značně nízká (doporučená hodnota je nad 0,7), ale zde logická s ohledem na nízký počet položek testu. Kontrolujeme také potenciální změnu hodnoty koeficientu po vyloučení dané položky. V případě, že dosahuje vyšší hodnoty než za celou škálu (Test scale) můžeme zvážit její vyřazení (kupříkladu zde položka 6).

7. konstrukce latentní proměnné měřící výkon v testu pomocí *hrubého testového skóre* s využitím zprůměrování. Jde o dokonalejší způsob než prostá sumace hodnot jednotlivých položek, v případě kódování hodnot 0/1 tak totiž dostáváme podíly na úspěšnosti z celého rozsahu testu (po vynásobení stem v procentech).

```
. egen d4WSTmea = rowmean(d4wst01-d4wst12) if toused4wst==1
. lab var d4WSTmea "WST-t4 dítě, raw-sc.(12×mean)"
. format d4WSTmea %6.2f
. fre d4WSTmea
```

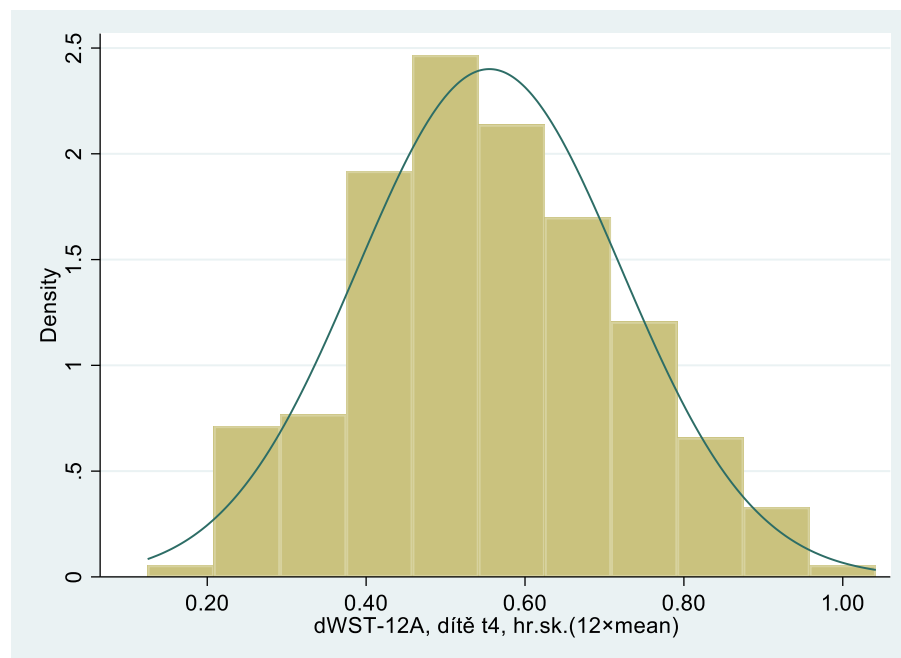
		Freq.	Percent	Valid	Cum.
Valid	0.17	1	0.22	0.46	0.46
	0.25	13	2.91	5.94	6.39
	0.33	14	3.14	6.39	12.79
	0.42	35	7.85	15.98	28.77
	0.50	45	10.09	20.55	49.32
	0.58	39	8.74	17.81	67.12
	0.67	31	6.95	14.16	81.28
	0.75	22	4.93	10.05	91.32
	0.83	12	2.69	5.48	96.80
	0.92	6	1.35	2.74	99.54
	1.00	1	0.22	0.46	100.00
Total		219	49.10	100.00	
Missing	.	227	50.90		
Total		446	100.00		

8. Po vytvoření proměnné hrubého skóre testu provedeme ověření rozložení jeho hodnot. Skewness-kurtosis test ověřující hypotézu o normalitě rozložení hodnot vždy doplníme vizuálním zobrazením např. pomocí histogramu. Zde zamítáme H_0 předpokládající, že distribuce hodnot je signifikantně odlišná od normálního rozdělení ($p > 0.05$) a to v případě šikmosti (skewness) tak i špičatosti (kurtosis).

```
. sktest d4WSTmea
. histogram d4WSTmea, discrete normal xtitle("WST-t4 dítě (12×mean)")
```

Skewness/Kurtosis tests for Normality					
Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	joint Prob>chi2
d4WSTmea	219	0.3367	0.2111	2.51	0.2849

Rozložení hodnot hrubého skóre testu dWST-12A pro děti t4 (1. pololetí páté třídy), podíl úspěšnosti



Tímto máme vytvořeno hrubé skóre z testu dWST-12A pro děti, s názvem proměnné d4WSTmea. Tím ale zpracování dat z testu většinou nekončí, je vhodné ho také dále ještě standardizovat, kupříkladu na percentilové pořadí (percentile range) případně na procentní rozpětí (percentage range).

Podrobněji se tímto testem zabýváme dále v části – Test kognitivních schopností dWST -12A verze pro děti na konci prvního stupně ZŠ (věk 10-11 let), kde ho představujeme podrobněji, včetně ověření validity (krok 10). V následujících částech tohoto tutoriálu představíme jednotlivé měřicí nástroje, nejprve test WST-10A k měření kognice v reprezentativních sociologických výzkumech v dospělé populaci a poté vybrané testy jazykových a kognitivních schopností či pre-gramotnosti určené dětem mladšího školního věku (počátek 1. stupně ZŠ).

3. Test kognitivních schopností Wordsum (WST-10A) – verze pro dospělé

V této části představujeme první výsledky české adaptace krátkého testu verbální inteligence Wordsum (my jej zde označujeme jako verzi WST-10A) používaného od 70. let v americkém výzkumu GSS. Tento test nachází široké uplatnění v reprezentativním sociologickém šetření, ale i v oblasti ekonomické a politické vědy (indikátor schopností v predikci volebního chování, vysvětlení variability v příjmech, aj.). V českém prostředí jsme tento nástroj připravili za účelem měření kognitivních schopností rodičů v kontextu vzdělávání (primárně pro panelový výzkum žáků základních škol), uplatnění najde ale i při řešení jiných sociologických úloh a témat zkoumaných v rámci reprezentativních dotazníkových šetření v dospělé populaci. Za účelem adaptace WST-10A na české jazykové prostředí jsme uskutečnili předvýzkum na populaci vysokoškolských studentů. Tímto způsobem jsme provedli první odladění v překladu konkrétních položek. Poté proběhla další kalibrace na reprezentativním vzorku dospělé populace (ve spolupráci s Centrem pro výzkum veřejného mínění SOÚ AV ČR, v.v.i.). V první fázi adaptace testů měřících schopnosti byla náročná zejména přesná kalibrace obtížnosti jednotlivých položek, překlad významu u některých položek byl proto postupně doupravován a doladován. Na základě výsledků z této přípravné fáze byla navržena česká verze WST-10A, která byla dále podrobena kalibrování v reprezentativním vzorku dospělé populace, kde byly souběžně zjišťovány různé proměnné reprezentující výsledky nebo determinanty související s kognicí, včetně sociodemografických charakteristik respondentů.

Nejprve jsme tedy provedli rozsáhlou pre-testovací fázi výzkumu v populaci studentů pražských vysokých škol (FSV, FF, FHS, PedF UK, PA ČR), kdy byla administrována rozsáhlejší databanka položek, celkem 20. Jednalo se o všechna slova ze seznamu A i B, který navrhl Thorndike (1942). Ukázalo se, že valná většina položek tak, jak byla přeložena, byla pro respondenty studenty velmi lehkých, proto jsme pro pilotní studii v dospělé populaci vybrali obtížnější položky a provedli první vlnu sběru dat, kam bylo zařazeno již jen 13 položek. Následně jsme provedli analýzu celkového skóre i jednotlivých položek, a protože celkové skóre stále vykazovalo z hlediska rozložení hodnot vychýlení doprava, tak jsme v sadě 13 slov upravili odpovědi u tří z nich tak, aby odpovědi byly obtížnější. Jednalo se o ta slova, u nichž jsme navíc měli výsledky z výzkumu GSS v populaci USA a věděli jsme proto jakou mají mít přibližnou váhu v testu. Poté proběhla další vlna sběru dat ve spolupráci s CVVM tentokrát v dospělé populaci ČR (kvótní výběr). Následovala analýza, jejímž cílem bylo vybrat 10 finálních položek, které by nejlépe reprezentovaly vlastnosti měřené testem WST. Vedle kritéria obtížnosti, kdy jsme preferovali obtížnější položky, jsme také z části přihlíželi k validitě reliabilitě/časové stabilitě jednotlivých položek. Vznikla tak první verze testu WST-10A adaptovaná pro dospělou populaci České republiky (předpokládáme, že v její kalibraci budeme třeba ještě pokračovat, viz doporučení).

Zadání a použití testu

Úkolem respondenta je najít synonymum k 10 slovům. Správnou odpověď vybírá z pěti alternativních slov, kdy vždy je jen jedna správná odpověď. Celkem test netrvá déle než 6-8 minut. V módu sběru dat pomocí face-to-face je dotázanému nejprve vysvětlen princip a ukázán příklad, jak test funguje. Následně je mu postupně předloženo 10 slov, každé z nich je zobrazeno na kartě. V případě sběru dat s využitím CAPI je mu ukázána obrazovka na zařízení. Respondenti jsou pak požádáni, aby zvolili volbu odpovědi, která se podle nich se blíží

významu klíčového slova. Pro odpověď není zaveden časový limit ani nabízena možnost reakce, že dotázaný neví.

Ukázka zadání a příkladu testu

"Chtěli bychom vědět něco o tom, jak lidé hádají slova, která neznají. Možná některá z nich znáte, jiná ne. Na každém řádku je uvedeno první slovo velkými písmeny a dalších pět slov. Prosím, označte vždy to slovo, které se podle Vás nejvíce blíží smyslu slova napsaného velkými písmeny. V příkladu v rámečku máme velkými písmeny slovo ŠELMA, takže z nabídky odpovědí označíte slovo „zvíře“. Slova v seznamu jsou ve skutečnosti pro všechny velmi obtížná, pokud si tedy nejste jistí odpovědí, zkuste si tipnout." POKYN: UKAŽTE RESPONDENTOVI PŘÍKLAD.

Příklad:

ŠELMA:


strach výrok velký zvíře vyloučený

POKYN: UKAŽTE RESPONDENTOVI MOŽNOSTI ODPOVĚDÍ.

Skóre testu

Podíl úspěšnosti v jednotlivých položkách uvádí tabulka 2, kde z důvodu zachování unikátnosti, resp. anonymity znění testu uvádíme jednotlivé položky pouze s označením pomocí slovních druhů. Na první pohled je patrné, že v použité verzi testu převažují lehčí položky, tj. takové, u nichž je vysoká úspěšnost. Více jak 85% úspěšností se vyznačuje sedm z deseti položek, pouze jedna má procento úspěšnosti menší než 50%.

Tabulka 2. Test WST-10A dospělá populace ČR, verze 1 (CVVM 2017 pilot II.), úspěšnost u jednotlivých položek

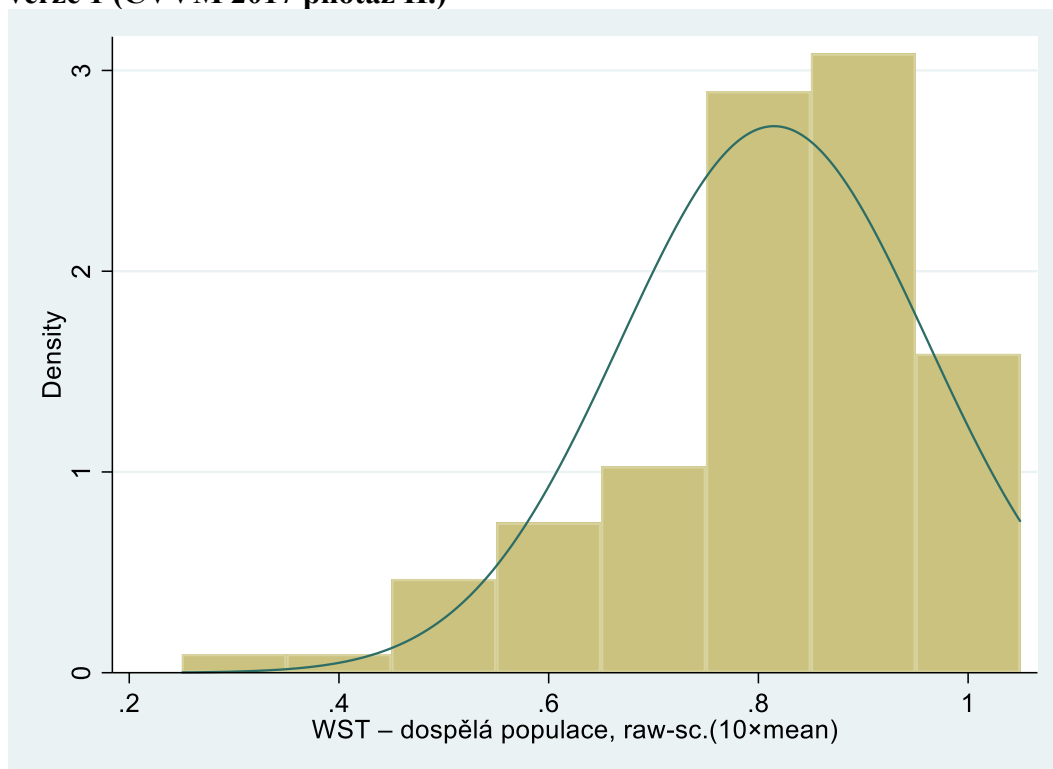
	p	
wst2_1	podstatné jméno	0.96
wst2_3	sloveso	0.96
wst2_13	přídavné jméno	0.90
wst2_9	podstatné jméno	0.89
wst2_5	sloveso	0.89
wst2_11	přídavné jméno	0.86
wst2_12	podstatné jméno	0.86
wst2_4	sloveso	0.85
wst2_6	přídavné jméno	0.67
wst2_2	sloveso	0.31

Zdroj: CVVM 2017 pilot II., N = 107

Ze správných odpovědí (viz předchozí část s podrobným návodem, jak to udělat na obdobném testu adaptovaném pro děti) vytvoříme pomocí zprůměrování hodnot (0/1) odpovědí (nebo pomocí součtu správných odpovědí) hrubé skóre z testu WST-10A. Převažující počet položek lehkých se promítá i do průměrné hodnoty skóre z testu, která je v dospělé populaci (ve vzorku v pilotáži CVVM II.) 0,81 (SD 0,15), což odpovídá osmdesátiprocentní úspěšnosti správných odpovědí pro verzi testu s finálně vybranými deseti položkami. Nejnížší úspěšnost byla až na 30%, tj. min. 3 položky zodpovědělo pouze jedno procento respondentů (a 5, tedy polovinu

testu dokonce jen 7%), zatímco nejvyšší 100% úspěšnosti dosáhlo 16 % respondentů. Nerovnoměrné rozložení v úspěšnosti správných odpovědí je také patrné z histogramu (graf 1), který dokumentuje jednoznačné vychýlení směrem doprava. Tuto skutečnost potvrzuje i test normality rozložení hodnot, který nezamítá H_0 předpokládající, že distribuce hodnot je signifikantně odlišná od normálního rozdělení ($p > 0.05$) jak u šikmosti (skewness) tak i špičatosti (kurtosis). Dodejme, že u originální verze testu použité v americké populaci (USA 2014) je průměrná úspěšnost za pro celou populaci 6,00 (SD 1,97 viz graf 2), pro bělochy ale 6,31 (SD 1,87) tj. je rovněž nepatrně vychýlena, přesto rozložení správných odpovědí zde dosahuje normální rozložení hodnot (výsledky z našich vlastních výpočtů).

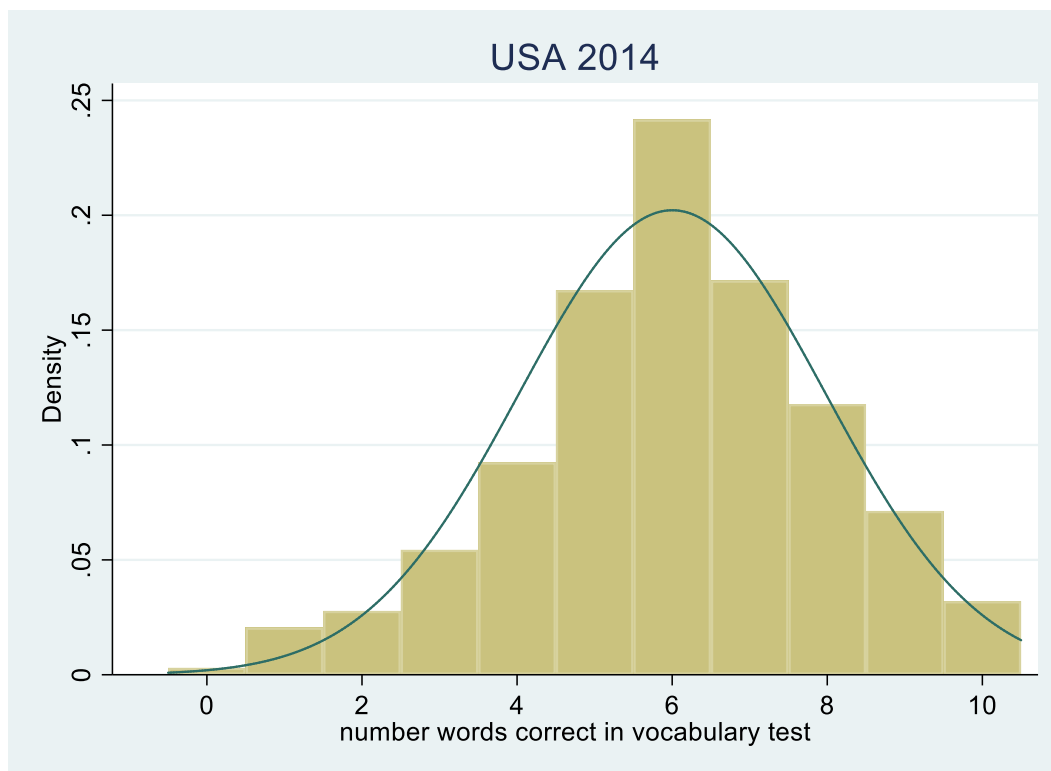
Graf 1. Rozložení hodnot hrubého skóre testu WST-10A pro dospělou populaci ČR, verze 1 (CVVM 2017 pilotáž II.)



Zdroj: CVVM 2017 pilot II., N = 107

Skewness/Kurtosis tests for Normality						
Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	adj	joint	Prob>chi2
				chi2(2)		
WSt2_~0xMean	107	0.0001	0.0648	14.87	0.0006	

Graf 2. Rozložení hodnot hrubého skóre originálního testu WST-10A pro dospělou populaci USA (GSS 2014)



Zdroj: GSS 2014 (vlastní výpočty), N = 1655

Reliabilita a validita

Dřívější výzkum v USA využívající testu Wordsum ukázal, že dosažené skóre je vysoce korelované (min. 0,75) s rozsáhlejšími a mnohem delšími testy obecné míry kognitivních dovedností (Miner 1957), což jen dokazuje, že krátký test slovní zásoby může sloužit jako efektivní proxy indikátor pro obecnou inteligenci a kognitivní dovednosti v rozsáhlejších reprezentativních výzkumech.

Pilotní studie v dospělé populaci sledovala také mnoho vlastností, které s kognitivními dovednostmi souvisí a mohou tak pomoci validizovat WST-10A. Souběžnou validitu opíráme především o výsledky spjaté se vzděláváním, k dispozici máme mimo jiné nejvyšší stupeň dosaženého vzdělání (9 stupňů), retrospektivní odhad celkového studijního průměru na konci školní docházky, příjem a dále výsledky ze tří paralelních testů poznávacích schopností (testu verbální inteligence z testu struktury inteligence IST – subtest výběr z obrazců a subtest analogie² a dále náš vlastní test kulturních znalostí měřící kulturní kapitál).

Deset vybraných položek v druhé vlně dosahuje relativně nízké položkové reliability, koeficient Kuder-Richardson 20 pro dichotomická data (KR-20) je 0,534. To ale není překvapivý výsledek, s ohledem na značně nízký počet položek jde o reliabilitu rozhodně uspokojivou (dodejme, že reliabilita WST v USA pro rok 2014 je vyšší $KR-20 = 0.625$).

Souběžnou validitu opíráme především o skutečnost, že v souladu s teoretickým očekáváním kognitivní úroveň pozitivně souvisí s dosaženými výsledky ve vzdělávání (dosažené vzdělání a studijní průměr na konci školní docházky), dále s osobním příjmem, socioekonomickým statusem. Také předpokládáme kladný vztah s velikostí bydliště, ale i nekognitivním

² Test struktury inteligence I-S-T 2000 R (Amthauer et al., 2015) měří více oblastí poznávacích schopností, mj. také verbální inteligenci (konkrétně subtesty výběr obrazců a analogie, která má nejvyšší validitu) a tu jsme využili pro validizaci testu WST-10A.

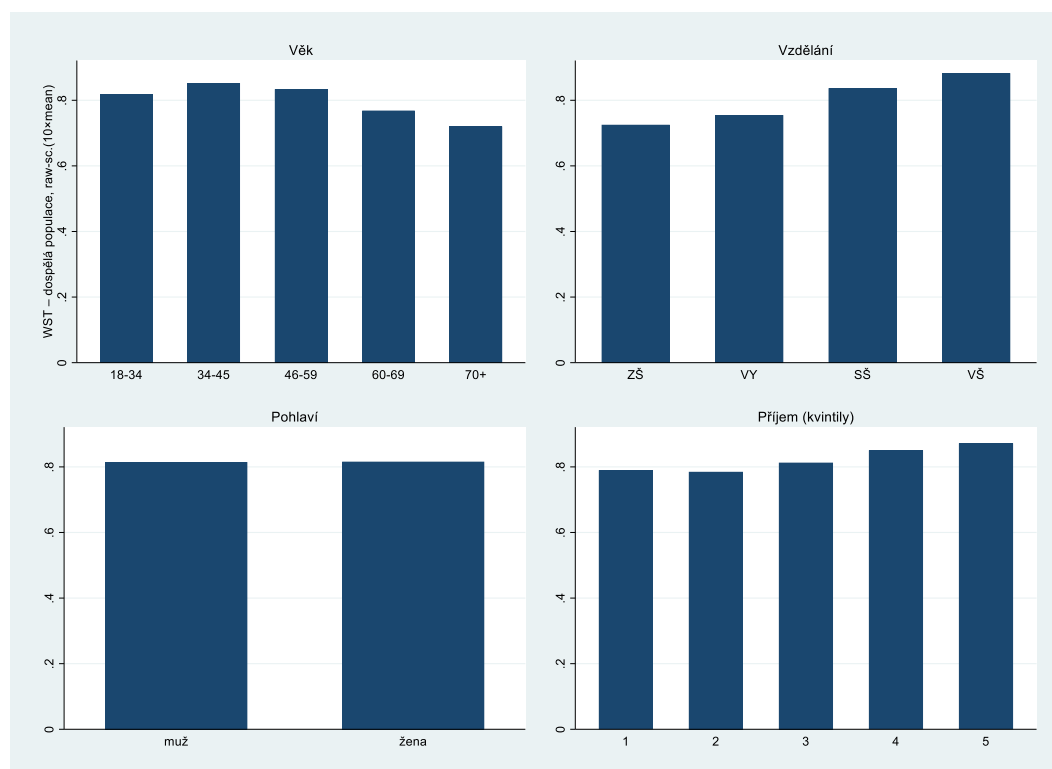
osobnostním rysem – schopnost ovládat svůj život tzv. vnitřní těžiště řízení (interní locus of control) a frekvencí čtení knih. Na druhou stranu neočekáváme rozdíly podle pohlaví a rozdíly podle věku by měly být v souladu s poznatky psychologického výzkumu nelineární – kognitivní schopnosti dosahují vrcholu v rané dospělosti a pak s věkem postupně klesají. Jak ukazuje tabulka 3. všechny tyto korelace jsou v souladu s očekáváním, s výjimkou frekvence čtení knih, s níž testové skóre WST-10A nevykazuje žádný vztah. Korelace s dosaženým vzděláním je vyšší než v případě známek na konci ZŠ a především platí, že vysokou korelaci lze pozorovat se sub-testy IST i kulturními znalostmi. Také jak ukazuje graf 3 ženy a muži se v kognitivních schopnostech neliší a vztah s věkem je nelineární, hodnota WST-10A je nejvyšší zhruba ve věku 34–45, posléze pomalu klesá.

Tabulka 3. Vztah skóre z testu WST-10A a vybraných proměnných (sociodemografické charakteristiky a paralelní testy poznávacích schopností), korelační koeficienty.

vzd9	0.380
znamkyZS	-0.203
SES-ISEI	0.334
prijemTis	0.202
Velobce	0.189
Locus-Control	0.293
Cteni knih	-0.007
Kult.znalosti	0.294
IST-Obrazce	0.371
IST-analogie	0.268

Zdroj: CVVM 2017 pilot II. (pairwise)

Graf 3. Rozdíly v průměrných hodnotách testu WST v dospělé populaci ČR podle sociodemografických charakteristik.



Zdroj: CVVM 2017 pilot II.

Doporučení ohledně české verze testu WST pro dospělou populaci

Na základě pilotní studie vznikla první verze testu WST-10A s 10 položkami (fakticky se jedná o tzv. beta verzi) adaptovanou pro dospělou populaci České republiky. Vyznačuje se dobrou validitou a relativně nízkou položkovou reliabilitou, ale také poměrně velkým vychýlením doprava, což je způsobeno převahou lehkých položek aktuálně zařazených v testové baterii. Domníváme se, že v její kalibraci by proto bylo vhodné dále pokračovat, potenciálním uživatelům zejména doporučujeme zařazení navíc položek (přibližně tří až čtyř) s vysokou a střední úrovní obtížností. K podobnému závěru dochází i autoři, kteří se pokusili aktualizovat americkou verzi WST (Cor, Haertel, Krosnick 2012).

4. Vybrané testy k měření pre-gramotnosti a kognitivních schopností v populaci dětí na prvním stupni základní školy

4a. Testy k měření pre-gramotnosti – kognitivních a jazykových schopností u dětí na počátku školní docházky (6–7 let)

Pre-gramotnostní dovednosti (tj. ve fázi nečtení) navrhujeme sledovat ve dvou konceptuálně odlišných sférách a tudíž na sobě relativně nezávislých prediktorech gramotnosti: *kognice* reprezentované rychlým (automatizovaným) pojmenováváním jednoduchých podnětů zobrazených na kartě a rozvoje *jazyka* konceptualizované jako rozsah (pasivní) slovní zásoby. Určitou úroveň v těchto sférách schopností jsou v nárocích stanovených v (KOV) RVP PV a učitelé je u žáka přicházejícího do prvního ročníku očekávají. K jejich operacionalizaci byla použita následující měřítka – screeningové nástroje.

K optimalizaci testů využíváme data z longitudinálního panelového výzkumu přechodu žáků do ZŠ a vývoje na 1. stupni. Do studie bylo na s pomocí náhodného víceúrovňového stratifikovaného výběru vybráno 25 ze všech veřejných základních škol ve Středočeském kraji (15) a v Praze (10). Strata představují odlišné residenční území resp. typ školy (venkov, město, Praha a její suburbánní zázemí). Cílovou skupinu tvoří čeští žáci na počátku prvního stupně ZŠ, kteří nastoupili do 1. třídy ve školním roce 2013/2014 bez odkladu školní docházky a jedná se o děti vyvíjející se v pásmu normy, tj. bez diagnostikovaných poruch. Monitoring rozvoje čtenářských a pisatelských dovedností v prvních letech školní docházky optikou psycholinguistiky totiž vyžaduje kritérium výběru žáků a žákyň z monolingvního prostředí rodiny, v němž dítě vyrůstá. Do výzkumu bylo na základě informovaného souhlasu rodičů zařazeno celkově 439 žáků, 247 ze škol pražských a 192 ze Středočeského kraje. V prvních dvou vlnách byla testovací sezení realizována individuálně s každým dítětem. K optimalizaci pro další využití ve výzkumech jsme vybrali následující dva testy pre-gramotnostní dovedností. *Schopnost rychlého automatizovaného* (někdy též vizuálně sériového) *pojmenování* (RAN – „rapid automatised naming“) je kapacita rychle a snadno si vybavovat fonologické informace uložené v dlouhodobé paměti (Denckla, Rudel, 1976). Použity byly dvě varianty téhož testu – jednou s obrázky a podruhé s číslicemi. Pro měření této schopnosti byla v případě obrázků využita česká verze subtestu (Seidlová Málková, Caravolas, 2013). Dětem byly vždy předloženy zácvičné karty s pěti obrázky objektů, resp. číslicemi tak, abychom se jednak ujistili, že děti objekty/ číslice velice dobře uznají a druhá se seznámí se samotnou úlohou (rychlost správně pojmenovávat). Následně děti dostanou samostatnou kartu se 40 objekty/ číslicemi a ty by si už samy měly vybavovat a nahlas co nejrychleji a bez chyb pojmenovat. Test byl přitom dítěti předložen dvakrát (objekty, číslice) bezprostředně po sobě s cílem zvýšit reliabilitu

měření. Výsledkem testu je naměřený čas v jednotlivých oblastech pojmenování. Průměrná celková délka jmenování se u slov a číslic mírně lišila (v první vlně děti zvládly jak čísla, tak slova přečíst přibližně za 50–55 vteřin, v druhé vlně jim slova trvala v průměru 42 sec., ale čísla jen 29 sec.) a tudíž při prostém sečtení času by tak test číslic měl v druhé vlně nižší váhu. Tento problém navrhujeme řešit pomocí určení výsledného vektoru, který lze získat ze čtyř naměřených časů pomocí metody hlavních komponent v podobě latentní proměnné z první komponenty (PCA potvrdila jednodimenzionalitu testu).

Slovní zásoba. Při tvorbě testu slovní zásoby byly využity podklady z amerického neverbálního testu pro měření řečového vývoje – tzv. receptivní slovník (Peabody Picture Vocabulary Test – IV; Dunn & Dunn, 2007³). Administrátor testu předložil dítěti vždy kartu se čtyřmi obrázky a řekl slovo vystihující jeden z nich, dítě má pak ukázat na daný obrázek. Použili jsme 31 položek, a ačkoliv jsme v další vlně výzkumu nejsnazší položky vyřadili a současně počet podnětů navýšili (celkem na 47), tak ve zde prezentované úloze používáme pouze jejich průnik za obě vlny tak, aby bylo možno věcně interpretovat změnu. U každého žáka byl výsledek testu měřen pomocí sumačního skóre, resp. zprůměrováním odpovědí, což na rozdíl od prosté sumace zajišťuje větší výtežnost dat, pokud se v nich vyskytují chybějící hodnoty (při stanovení určitého minima platných odpovědí).

Standardizace rozpětí

U obou testů lze hrubé hodnoty dále přeškálovat tak, aby odpovídaly metrice jiných testů. Doporučujeme hodnoty převádět na rozpětí 0 až 100 na principu tzv. percent in range, kdy nejhorší výsledek odpovídá 0 a nejlepší 100. Vedle takto přeškálovaných testových skóre při popisu úrovně žáků je však vhodné uvádět také hodnoty hrubých testových skóre, neboť při interpretaci výsledků poskytují principiálně věcnou představu o obsahu měřeného indikátoru rané gramotnosti. To, že schopnost vybavování si podnětů z dlouhodobé paměti a úroveň pasivní slovní zásoby představují odlišné složky pre-gramotnosti, dokumentuje poměrně nízká korelace mezi testy (v první vlně 0,17 resp. 0,13 v druhé vlně).

Popisné charakteristiky všech použitých proměnných lze nalézt v tabulce A1 v příloze, kde jsou uvedeny pro původní hrubé (metrické) skóre v testech (u testu RAN je pro jednoduchost uveden jen součet délky obou dílčích subtestů v sekundách).

Je třeba zdůraznit, že dosažené testové skóre samy o sobě nedávají „objektivní“ výpověď o stavu dovedností prvňáčků, neboť jde o testy, které dosud nebyly v českém prostředí normalizovány a standardizovány pro diagnostické – poradenské účely.

Popis počáteční gramotnosti žáků 1. třídy

Dále uvádíme popisné statistiky na základě hrubých skóre z testů odpovídající původní substantivní metrice měření a také pro snadnější porovnání mezi testy jejich standardizovanou verzi. Uvádíme jaké jsou rozdíly v jazykových a kognitivních dovednostech dětí, které přicházejí do první třídy ZŠ ve Středočeském kraji a Praze (pro děti z monolingvních rodin, bez vývojových obtíží a bez odkladu školní docházky).

Základní představu poskytují střední hodnoty v grafech fousatých-krabiček v grafu 8 s výsledky testů ve standardizovaném rozpětí (0–100) na počátku a na konci první třídy, ale i histogramy v grafech 4-7.

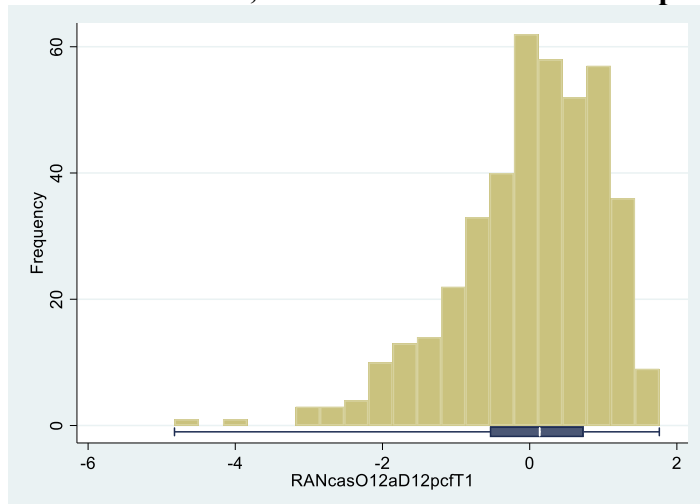
Metalingvistické dovednosti v kognitivní sféře sleduje test schopnosti *rychlého automatizovaného pojmenování (RAN)*. Žáci se během prvního školního roku zlepšili v celkovém standardizovaném skóru z průměrných 73 na 79 bodů, zmenšil se i rozptyl hodnot

³ Původní předlohu sestavil Filip Smolík z PsÚ AV ČR.

(std z 15,5 na 10,7).⁴ Věcně zajímavější informaci však poskytují rozdíly v hrubém skóre, které reprezentuje časový údaj v sekundách navíc s rozlišením pro subtesty obrázků a číslovek. Konkrétně při čtení slov–obrázků průměrná doba čtení se snížila o 9 sec, v první vlně dosahovala 51 sec, ve druhé klesla na 42 sec ($t = 20,05$; $p = 0,000$),⁵ což odpovídá zlepšení o cca 18 % (std se zmenšila z 11,3 na 8,8). Jistě nepřekvapí, že k dramatičtějšimu zlepšení dochází u čtení čísel, zatímco v první vlně dětem trvalo průměrně 55 sec, tak v druhé již jen 29 sec (std 29,4 na 7,8). 26vteřinový pokles doby potřebné k přečtení zadaných číslic tak představuje zlepšení o 47 % ($t = 19,02$; $p = 0,000$).

Ke zlepšení během první třídy dochází také v případě *testu slovní zásoby (VOC)*, i když z věcného hlediska se nejedná o výraznější změnu, neboť na počátku školního roku děti rozpoznaly význam v průměru 22 slov a na konci to bylo, pro srovnatelný počet (31) položek, 23 slov ($t = 7,29$; $p = 0,000$), přičemž ani rozptyl hodnot se víceméně nezměnil (std 3,3 resp. 3,4). Relativní zlepšení slovní zásoby tak představuje jen zhruba 6 procent. Dodejme, že hodnota standardizovaného skóre (0-100) pro plnou verzi v druhé vlně, která byla 73 při příchodu do první třídy stoupla na jejím konci na 92 (std 15,4 a 11,9) a ukazuje tak, že ve slovní zásobě došlo v relativním srovnání vůči rychlému pojmenování (RAN) k mnohem většímu posunu.

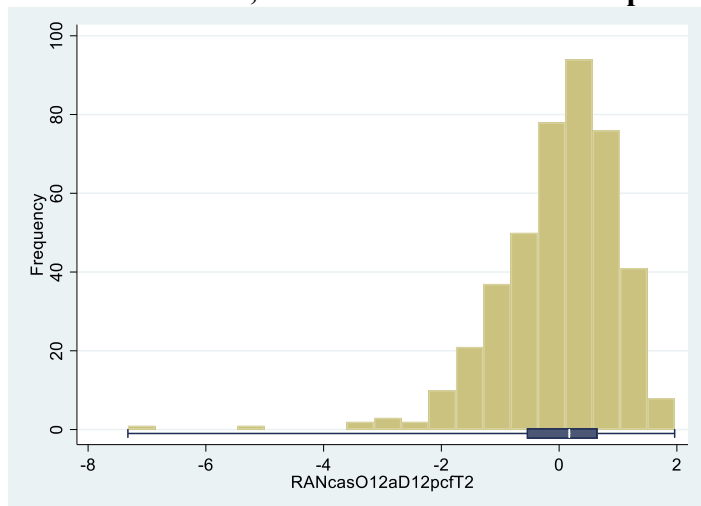
Graf 4. Test RAN, rozložení hodnot na začátku první třídy.



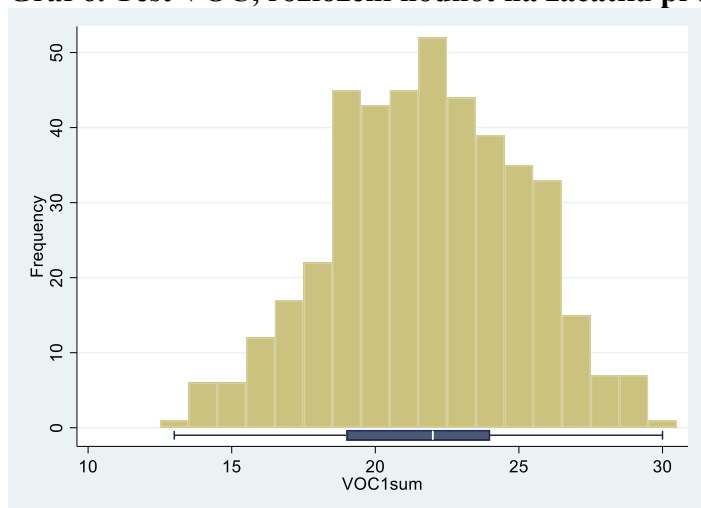
⁴ U tohoto testu ovšem není horní hranice rozpětí jasně daná, při standardizaci na % z rozpětí jsme ji stanovili na 200 sekund u číslic a 100 sekund u obrázků.

⁵ Jedná se vždy o průměr z obou pokusů v testu.

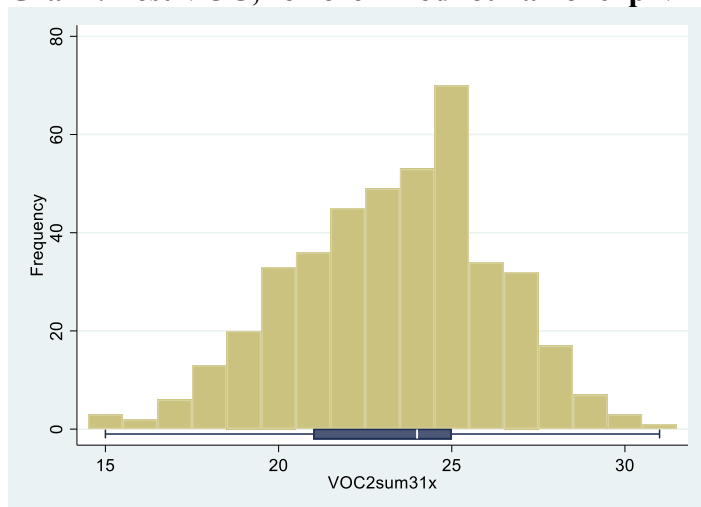
Graf 5. Test RAN, rozložení hodnot na konci první třídy.



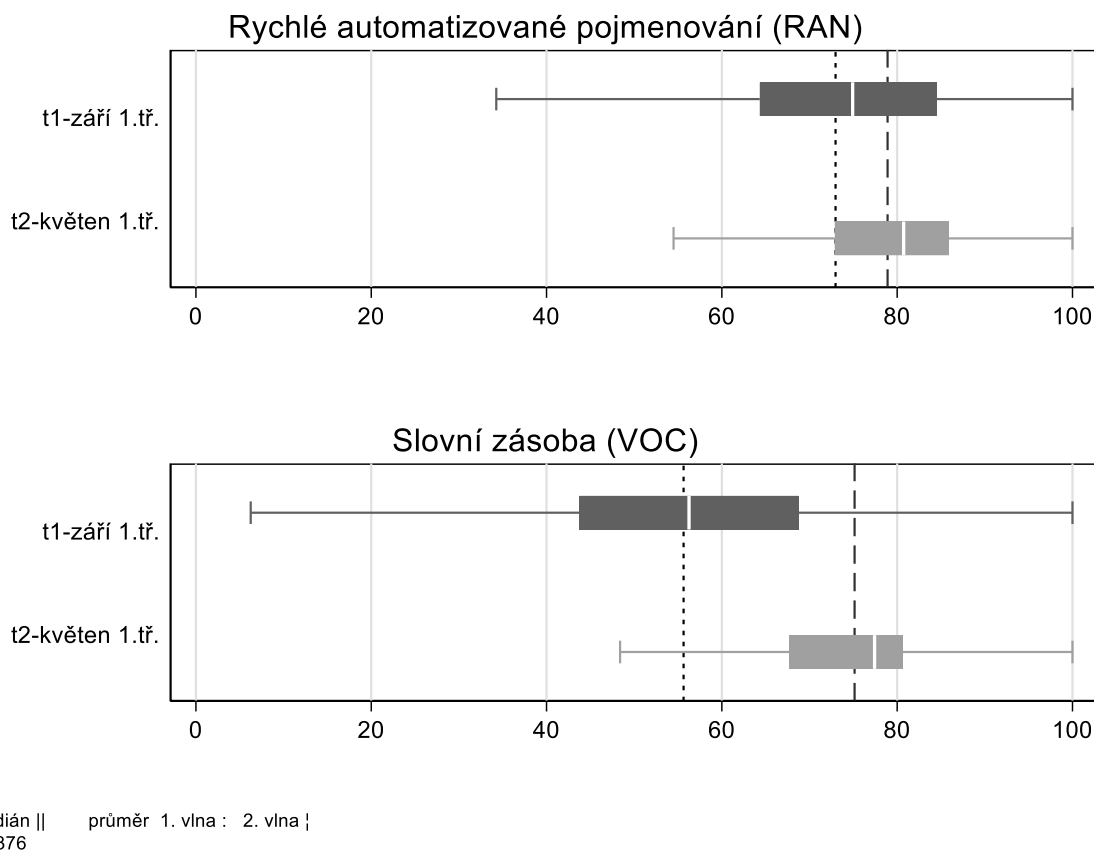
Graf 6. Test VOC, rozložení hodnot na začátku první třídy.



Graf 7. Test VOC, rozložení hodnot na konci první třídy.



Graf 8. Skóre testů na počátku a na konci první třídy, standardizované rozpětí skóre (percentage range).



4b. Test kognitivních schopností Wordsum (dWST-12A) – verze pro děti (10-11 let)

Test WST-10A – Wordsum je určen proměření kognice dospělých, neboť kromě schopnosti číst vyžaduje rozsáhlejší slovní zásobu a schopnost abstrakce. Není proto použitelný pro děti ve věku mladší školní docházky, nicméně v rámci panelové studie ZŠ jsme se ho pokusili adaptovat pro žáky na přelomu druhého stupně základní školy (dWST-12A). K výběru položek jsme využili data z pilotáže v subpopulaci studentů druhého stupně ZŠ (dvě třídy jednoho pražského víceletého gymnázia). Finálně použitá verze testu v Panelu ZŠ obsahovala 12 položek, celkem test obsahoval 4 podstatná jména, 5 sloves a 3 přídavná jména. Šest položek bylo stejných jako ve verzi pro rodiče, jedna položka byla podobná ale s pozměněnou lehčí variantou odpovědi. Rozložení v úspěšnosti správných odpovědí je na rozdíl od verze pro rodiče relativně rovnoměrné a ukazuje na smysluplnost použití tohoto testu pro měření kognitivních dovedností u dětí mladšího školního věku. Podíl úspěšnosti v jednotlivých položkách uvádí tabulka 4, kde z důvodu zachování unikátnosti, resp. anonymity znění testu uvádíme jednotlivé položky pouze s označením pomocí slovních druhů.

Celková položková reliabilita testového skóre pro všech 12 položek vyjádřená koeficientem KR-20 je 0,44 pomocí koeficientu Ordinal Alpha 0,60. Zdůrazněme, že se jedná o test s velmi malým počtem položek, kdy vhodnější metodou pro skórování by bylo využití metody IRT. Výsledky rozložení hodnot pro součtovou verzi hrubého skóre ukazuje graf 9. Minimální hodnota dosaženého hrubého skóre byla 2 správné odpovědi, ale pouze u jednoho dítěte, minimálně 3 správné odpovědi mělo 94 % žáků. Průměrné hrubé skóre z 12 položek bylo u dětí

zúčastněných v panelu ve 4. vlně, tj. ve věku 10-11 let bylo 6.7(SD 1.99). Dále byla také zkonstruovaná standardizovaná verze na percentil-rank (viz tabulku 5).

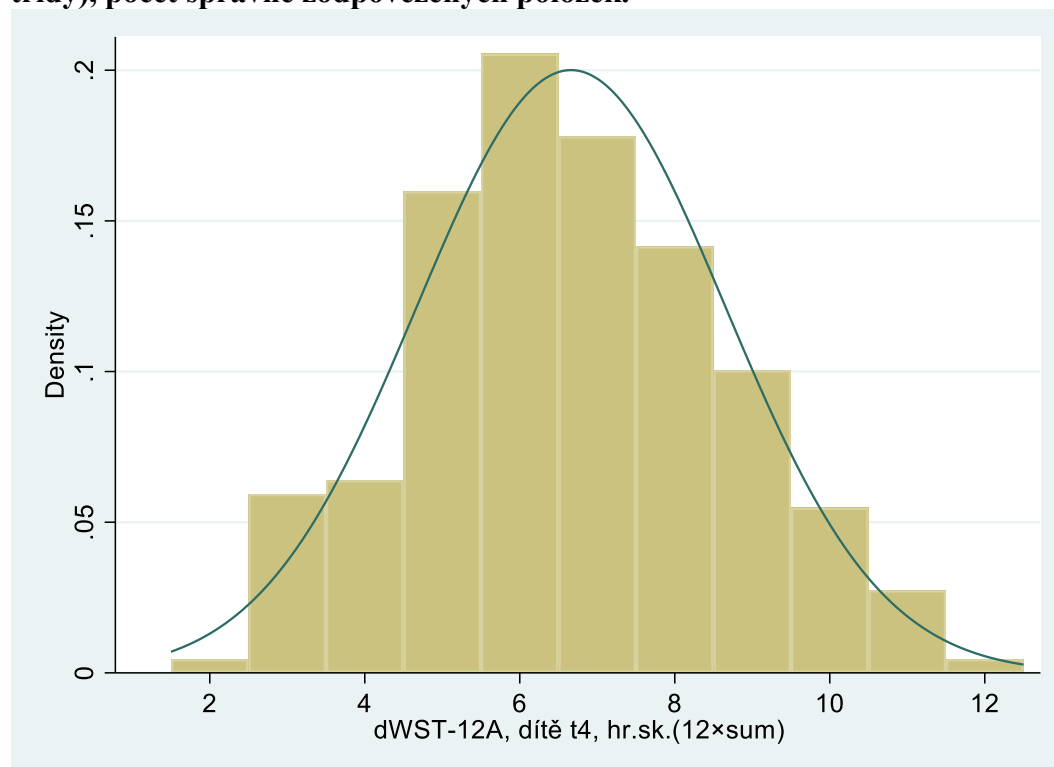
Tabulka 4. Test dWST-12A pro děti t4 (pololetí 5. třídy), úspěšnost u jednotlivých položek.

		n	p
d4wst01	podstatné jméno	219	0.91
d4wst03	sloveso *	219	0.81
d4wst10	přídavné jméno	219	0.79
d4wst06	přídavné jméno *	219	0.75
d4wst04	sloveso *	219	0.73
d4wst11	podstatné jméno *	219	0.55
d4wst05	sloveso	219	0.47
d4wst12	přídavné jméno	219	0.41
d4wst09	sloveso *	219	0.34
d4wst02	sloveso	219	0.30
d4wst07	podstatné jméno *	219	0.30
d4wst08	podstatné jméno	219	0.28

Zdroj: Panel ZŠ, N (t4) = 219

Poznámka: * stejné změni položky jako ve verzi pro rodiče

Graf 9. Rozložení hodnot hrubého skóre testu dWST-12A pro děti t4 (1. pololetí páté třídy), počet správně zodpovězených položek.



Zdroj: Panel ZŠ, N (t4) = 219

Tabulka 5. Výsledky v testu dWST-12A pro děti t4 (pololetí 5. třídy), percentilové pořadí (percentile rank).

Pct-rank	Freq.	%	Cum.%
0,0	1	0,46	0,46
0,5	13	5,94	6,39
6,4	14	6,39	12,79
12,8	35	15,98	28,77
28,9	45	20,55	49,32
49,5	39	17,81	67,12
67,4	31	14,16	81,28
81,7	22	10,05	91,32
91,7	12	5,48	96,80
97,3	6	2,74	99,54
100,0	1	0,46	100,00
Total	219	100,00	

Zdroj: Panel ZŠ, N (t4) = 219

Výkon v testu WST se neliší mezi chlapci a dívkami (6,8 a 6,6 $t = 0,902$ $p = 0,368$) ani podle typu školy (s rozlišením na 4 typy škol; $F = 0,59$ $p = 0,672$). Korelace s ostatními testy pre-gramotnosti uvádí tabulka 6. Validitu testu WST dokumentuje poměrně vysoká korelace s testem pasivní slovní zásoby VOC, i schopnost predikovat školní prospěch v páté třídě (korelace s průměrem ze 4 školních předmětů je $r = 0,244$, přičemž nejvyšší je se známkou z českého jazyka (0,264) a nejnižší se známkou z matematiky (0,098; hodnoty známek otočeny na min-max.). Na validitu testu WST ukazuje i korelace se vzděláním rodičů ($r = 0,289$) a také s výsledkem rodičů v obdobném testu WST pro dospělé (u prvního rodiče 0,222 ($n = 161$) a u druhého rodiče 0,264 ($n = 56$)).

Použitý test dWST-12A pro děti mladšího školního věku vykazuje dobré vlastnosti, nicméně s ohledem na to, že jde o testovou baterii s velmi malým počtem položek, bude třeba ještě jeho psychometrické vlastnosti dále ověřit v nezávislém výběru, ideálně v odlišných věkových kohortách, a také dále provést zhodnocení vlastností jednotlivých položek, nejlépe pomocí metody IRT.

Tabulka 6. Korelace testů VOC (vlna t1-t4) a RAN (vlna t1-t3) s testem WST-12 A pro děti t4.

VOC1sum*	0.449
VOC2sum*	0.364
VOC3sum*	0.181
VOC4sum*	0.496
VOCx23sum1	0.443
VOCx23sum2	0.333
VOCx23sum3	0.197
RANcas012aD12pcfT1	0.175
RANcas012aD12pcfT2	0.137
RANcas012aD12pcfT3	0.019

Zdroj: Panel ZŠ, N (t1-t4) = 171

Poznámka: * plná verze testu v příslušné vlně sběru dat (odlišný počet položek mezi vlnami)

5. Použitá literatura

- Alwin, D.F. (2010). Family of Origin and Cohort Differences in Verbal Ability. *American Sociological Review*, 56(5), 625–638.
- Amthauer, R., Brocke, B., Liepmann, D., Beauduce, A. (2015). *Test struktury inteligence IST - 2000 R. 2. české vydání*, Hogrefe – Testcentrum, Praha.
- Cor, M. Ken; Haertel, Edward; Krosnick, Jon A.; Malhotra, Neil. Improving ability measurement in surveys by following the principles of IRT: The Wordsum vocabulary test in the General Social Survey. *Social Science Research*, 41 (5): 1003–1016. doi:10.1016/j.ssresearch.2012.05.007
- Denckla, M. B., Rudel, R. G. (1976). Rapid 'automatized' naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, 14(4), 471-479. [https://doi.org/10.1016/0028-3932\(76\)90075-0](https://doi.org/10.1016/0028-3932(76)90075-0)
- Dorius, S. F., Alwin, D. F., & Pacheco, J. (2016). Twentieth century intercohort trends in verbal ability in the United States. *Sociological Science*, 3, 383-412. <https://doi.org/10.15195/v3.a18>
- Dunn, L. M. a Dunn, D. M. (2007) *Peabody picture vocabulary test, fourth edition (PPVT- 4)*. Bloomington: Pearson Clinical.
- Miner J.B. (1957). *Intelligence in the United States: A Survey*. New York: Springer Press.
- Seidlová Málková, G., Caravolas, M. (2013). *BTFS – Baterie testů fonologických schopností*. Praha: Národní ústav pro vzdělávání.
- Thorndike, R. L. (1942). Two screening tests of verbal intelligence. *Journal of Applied Psychology*, 26(2), 128–135. <https://doi.org/10.1037/h0060053>
- Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging*, 18(2), 332–339. <https://doi.org/10.1037/0882-7974.18>
- Zhu, J., Weiss, L. (2005). The Wechsler Scales. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (p. 297–324). The Guilford Press.

6. Přílohy

6a. Použitá data

Tabulka A1. Charakteristika výzkumného vzorku žáků prvních tříd ZŠ a testů pre-gramotnosti – deskriptivní statistiky.

	N	Průměr nebo %	SD	Min	Max
Věk	376	6,61	0,33	4,08	7,67
Pohlaví	376				
Děvčata		56 %			
Chlapci		44 %			
<i>Hrubé (metrické) skóry v testech pre-gramotnosti:</i>					
Slovní zásoba VOC t1–září 1.tř.	376	21,92	3,27	13	29
Slovní zásoba VOC t2–květen 1.tř.	376	23,24	3,40	0	31
Rychlé automatizované pojmenování RAN (obrázky+čísla, sec.) t1–září 1.tř.	373	53,05	18,06	25,5	145
Rychlé automatizované pojmenování RAN (obrázky+čísla, sec.) t2–květen 1.tř.t	376	35,50	7,39	19	88,75

Panel ZŠ t1–t2, N (listwise) = 376

Tabulka A2. Charakteristika výzkumného vzorku dospělé populace ČR (CVVM 2017 pilot II.) – deskriptivní statistiky.

	N	Průměr nebo %	SD	Min	Max
Věk	107	46,09	15,91	18,00	81,00
Pohlaví	107				
muž (%)	49	45,79 %			
žena (%)	58	54,21 %			
Velikost obce	107				
<5 tis. (%)	48	44,86 %			
5-30 tis. (%)	24	22,43 %			
>30tis. (%)	35	32,71 %			
Vzdělání 9k (ord.)	107	5,05	2,73	1,00	9,00
Vzdělání 3k	107				
ZŠ/VY (%)	38	35,51 %			
SŠ (%)	41	38,32 %			
VŠ (%)	28	26,17 %			
Prospěch ZŠ (studijní průměr)	104	1,94	0,65	1,25	3,90
Příjem v tis. Kč	79	16962	6808,68	5999	33500

Zdroj: CVVM 2017 pilot II., N = 107

6b. Zkrácené verze testu pre-gramotnosti VOC pro žáky 1. třídy

Výše uvedené testy VOC a RAN jsou původně určeny k diagnostickým účelům v pedagogickém poradenství. S ohledem na to pochopitelně vyžadují zajištění spolehlivosti výsledků vyšší počet položek měření. Tyto testy je možné je používat i běžném výzkumu např. v oblasti sociologie vzdělávání, nicméně s ohledem na časovou náročnost při jejich zadávání se jeví pro některé situace a úlohy vhodné jejich zkrácení. Navrhli jsme zkrácení na zhruba polovinu položek.

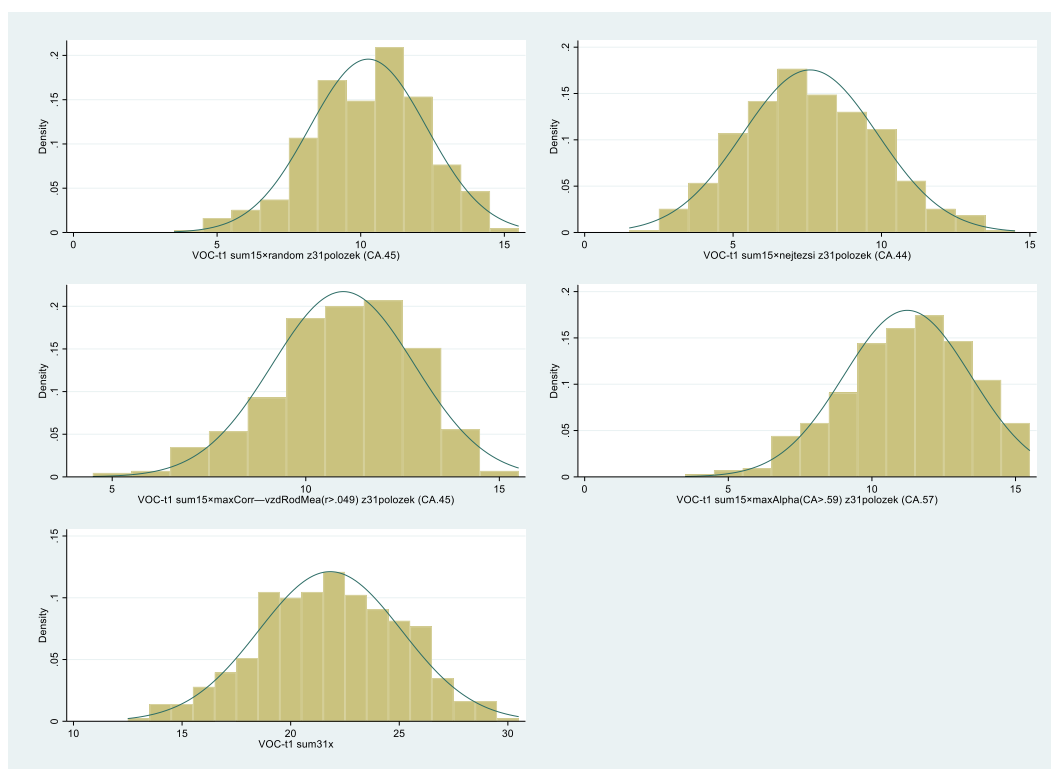
Přitom jsme postupovali tak, že jsme využili několik možných postupů pro jejich krácení a následně vybrali tu variantu, která maximalizovala prediktivní sílu testu. Testovány byly zkrácené verze pomocí: náhodného výběru položek, výběru nejtěžších položek, výběru položek s maximální korelací se vzděláním rodičů a výběrem položek maximalizujícím vnitřní konzistenci výsledného skóre (koeficient reliability KR-20).

Tabulka A3. Popisné statistiky k testu slovní zásoby pro žáky první třídy.

	N	Mean	SD	Min	Max	Reliabilita	Verze testu
VOC1sum31	430	21.82	3.29	13.00	30.00	0.60	VOC t1 sum31x – plná verze testu
VOC1sum15rn	430	10.26	2.04	4.00	15.00	0.45	VOC t1 sum15×random z31polozek
VOC1sum15tez	430	7.59	2.27	2.00	14.00	0.44	VOC t1 sum15×nejtezi z31polozek
VOC1sum15vzd	430	10.97	1.84	5.00	15.00	0.45	VOC t1 sum15×maxCorr—vzdRodMea($r>.049$) z31polozek
VOC1sum15alp	430	11.23	2.22	4.00	15.00	0.57	VOC t1 sum15×maxRelia(KR>.59) z31polozek

Poznámka: reliabilita měřena koeficientem Kuder–Richardson Formula 20 (KR-20) pro dichotomické znaky

Graf A1. Rozložení hodnot hrubého skóre různých verzí krácení testu slovní zásoby VOC.



Validizace zkrácené verze testu VOC.

Při výběru vhodné verze zkrácené varianty testu slovní zásoby uvažujeme závislost na kulturním prostředí rodiny (souběžná validita) a prediktivní sílu testu jak pro pozdější hodnoty testu v následné vlně a také pro pozdější školní výsledky (prediktivní validita). Při tom jsme využili panelovou složku dat i rodičovský dotazník, z něž máme k dispozici informaci o vzdělání rodiny.

Korelace VOC-t1 se vzděláním rodičů a s totožným (nezkráceným) testem VOC-t2 na konci první třídy jednak ve srovnatelné verzi (31 stejných položek), tak v plné verzi zohledňující růst slovní zásoby během první třídy (41 položek) ukazuje následující tabulka.

V případě vzdělání je nejvyšší souvislost pochopitelně v případě varianty krácení pomocí maximalizace korelace právě se vzděláním rodičů.

Tabulka A4. korelace různých verzí krácení testu VOC.

	vzdRodMeanW4i	VOC2sum	VOC2sum31x
VOC1sum31	0.216	0.531	0.500
VOC1sum15alp	0.216	0.463	0.416
VOC1sum15rn	0.212	0.453	0.444
VOC1sum15tez	0.205	0.454	0.445
VOC1sum15vzd	0.330	0.483	0.436
N	426	415	415

Poznámka: Pearson koef. korelace; $p < 0.001$.

Tabulka A5. korelace různých verzí krácení testu VOC.

	rW4znamkyR5t	W4Q11VG
VOC1sum31	0.146	0,207
VOC1sum15alp	0.150	0,232
VOC1sum15rn	0.152	0,213
VOC1sum15tez	0.108	0,124 (ns.)
VOC1sum15vzd	0.229	0,299
N	215	214

Poznámka: Pearson koef. korelace pro studijní průměr a polychorické korelace pro Přechod na VG; $p < 0.1$.

Na základě výše uvedených výsledků analýz jsme jako výslednou variantu optimalizované zkrácené verze testu slovní zásoby VOC pro úroveň na začátku první třídy (přibližně pro věk 6 let) zvolili verzi krácenou na základě maximalizace reliability položek (VOC1sum15alp).

Nicméně v případě využití testu u dětí starších, tj. při přechodu mezi první a druhou třídou doporučujeme spíše využití varianty s obtížnějšími položkami v testu, které se lépe přizpůsobí růstu schopností dětí v čase a ve výsledku poskytne symetričtější rozložení hodnot pro následnou vlnu v pozdějším věku (cca 7let).