

Mezinárodní výzkumy výsledků vzdělávání.

Metodologie, přínosy,
rizika a příležitosti

Jana Straková



Univerzita Karlova v Praze
Pedagogická fakulta
2016

Dedikace:

Kniha je výstupem projektu GA ČR: Vztahy mezi dovednostmi, vzděláváním a výsledky na trhu práce: longitudinální studie (číslo P402/12/G130).

Recenzovali:

Prof. PhDr. Arnošt Veselý, Ph. D.

PhDr. Ing. Petr Soukup

Jazyková korektura:

PhDr. Helena Justová

Obsah

Úvod	5
1. Výzkumy výsledků vzdělávání: historie a používané postupy	9
1.1 Využití standardizovaných testů k hodnocení výsledků vzdělávání	9
1.2 Aktivity Mezinárodní asociace pro hodnocení výsledků vzdělávání.....	12
1.2.1 Začátky mezinárodních výzkumů – iniciativa akademické obce	12
1.2.2 Periodická šetření matematických a přírodovědných vědomostí a dovedností	13
1.2.3 Periodická šetření čtenářské gramotnosti	14
1.2.4 Méně tradiční oblasti hodnocení zahrnuté do studií IEA	15
1.3 Aktivity OECD na poli měření výsledků vzdělávání.....	16
1.3.1 Přípravné aktivity pro výzkum OECD PISA – klíčové a mezipředmětové kompetence	16
1.3.2 Výzkum PISA.....	19
1.4 Výzkum studentů vysokých škol	20
1.5 Výzkumy v dospělé populaci	21
1.6 Realizace mezinárodních výzkumů výsledků vzdělávání žáků.....	22
1.6.1 Jednotlivé kroky realizace mezinárodních výzkumů	22
1.6.2 Rozdíly ve výzkumech IEA a OECD	31
1.7 Prezentace výsledků	32
2. Vývoj v konceptualizaci hodnocených oblastí.....	35
2.1 Výzkumy Mezinárodní asociace pro hodnocení výsledků vzdělávání	35
2.1.1 Matematika a přírodovědné předměty ve výzkumu TIMSS	35
2.1.2 Čtenářská gramotnost ve výzkumu PIRLS.....	38
2.2 Konceptuální rámce oblastí hodnocených ve výzkumu PISA	39
2.2.1 Čtenářská gramotnost ve výzkumu PISA	40
2.2.2 Matematická gramotnost ve výzkumu PISA.....	43
2.2.3 Přírodovědná gramotnost ve výzkumu PISA	45
2.2.4 Postoje k přírodním vědám a životnímu prostředí ve výzkumu PISA	46
2.2.5 Metakognitivní dovednosti ve výzkumu PISA.....	47
2.2.6 Řešení problémů ve výzkumu PISA	48
2.2.7 Finanční gramotnost ve výzkumu PISA	54
2.3 Konceptuální rámce oblastí hodnocených ve výzkumu PIAAC	55
3. Metodologie mezinárodních výzkumů	59
3.1 Konstrukce testů a dotazníků	59
3.1.1 Konstrukce testů	59
3.1.2 Konstrukce dotazníků	62
3.2 Konstrukce výběrů a vážení dat	68
3.2.1 Konstrukce výběrů	68
3.2.2 Konstrukce vah.....	72
3.2.3 Replikační váhy	73
3.3 Postupy škálování a prezentace výsledků	78

3.3.1	Teorie odpovědi na položku.....	78
3.3.2	Vícenásobná imputace.....	79
3.3.3	Měření trendů.....	81
3.3.4	Úrovně způsobilosti.....	82
3.4	Práce s daty.....	87
3.4.1	Datové soubory.....	87
3.4.2	Použití žákovských vah.....	87
3.4.3	Práce s plausibilními hodnotami a replikačními váhami.....	89
3.4.4	Další nástroje umožňující jednoduchou práci s daty.....	92
4.	Vývoj pohledů na cíle, možnosti a využití výzkumů.....	95
4.1	Ohlédnutí za metodologickým vývojem.....	95
4.1.1	Vývoj v oblasti konstrukce testu a dotazníku.....	95
4.1.2	Vývoj v technologické oblasti.....	97
4.1.3	Ověřované vzdělávací cíle.....	100
4.1.4	Moderní systémy kombinující formativní a sumativní hodnocení.....	104
4.2	Využití dat z mezinárodních výzkumů – historie a současnost.....	107
4.2.1	Hledání faktorů ovlivňujících výsledky vzdělávání.....	107
4.2.2	Mezinárodní výzkumy a výzkumy efektivity školního vzdělávání.....	111
4.2.3	Rozvoj metodologických postupů.....	114
4.2.4	Využití dat českými výzkumníky.....	121
5.	Kritika mezinárodních výzkumů a její reflexe	123
5.1	Pohledy na užitečnost a škodlivost dopadů mezinárodních výzkumů.....	123
5.1.1	Diskuse v mezinárodní komunitě.....	123
5.1.2	Kritika mezinárodních výzkumů v českých pedagogických časopisech.....	131
5.2	Výhled do budoucna – možné cesty k překonání nedostatků.....	133
5.2.1	Úvahy o budoucím směřování mezinárodních výzkumů výsledků vzdělávání.....	133
5.2.2	Využití mezinárodních výzkumů na národní úrovni – inspirace z Německa.....	138
6.	Shrnutí poznatků z mezinárodních výzkumů týkajících se českého vzdělávacího systému	143
6.1	Silné a slabé stránky v měřených oblastech a dlouhodobé trendy.....	143
6.1.1	Kognitivní výsledky.....	143
6.1.2	Nekognitivní výsledky.....	152
6.2	Rozložení výsledků v populaci.....	155
6.2.1	Rozložení výsledků a socioekonomické souvislosti.....	155
6.2.2	Poznatky týkající se rozdílů mezi skupinami.....	161
6.3	Další poznatky z mezinárodních výzkumů výsledků vzdělávání žáků.....	164
	Závěrem: Poznatky a příležitosti pro českou odbornou komunitu.....	167
	Literatura.....	177
	Seznam zkratk.....	187
	Příloha 1: Přehled realizace mezinárodních výzkumů výsledků vzdělávání žáků v ČR.....	189
	Příloha 2: Ilustrativní úlohy.....	191
	Summary.....	197

Úvod

Tato publikace využívá dvacátého výročí české realizace mezinárodních výzkumů výsledků vzdělávání k ohlédnutí se za jejich vývojem ve zmíněném období a za jejich uplatněním v České republice. V roce 2015 uplynulo 20 let od okamžiku, kdy byly v českých školách poprvé administrovány testy a dotazníky výzkumu TIMSS. Od té doby již u nás mezinárodní výzkumy zdomácněly. Jejich výsledky se stávají standardní součástí výročních zpráv MŠMT, jsou používány jako argumenty ve strategických dokumentech, data jsou čím dál tím častěji zpracovávána výzkumníky bádajícími v oblasti společenských věd. Některé koncepty našly uplatnění i ve školách a uvolněné úlohy čím dál tím častěji vstupují do vzdělávání učitelů i do školních tříd. Jsou realizovány návazné výzkumy sloužící k tomu, aby byly informace, které výzkumy poskytují, lépe zhodnoceny. Bohaté datové zdroje z mezinárodních výzkumů jsou však využity pouze částečně. Zároveň se objevují pochybnosti o užitečnosti těchto výzkumů a o validitě poznatků, které produkují, i o účelnosti využití nemalých finančních prostředků na jejich realizaci.

Tato publikace přináší informace o historickém vývoji mezinárodních výzkumů, jejich metodologii a zpracování získaných dat. Zároveň hledá odpovědi na otázky, do jaké míry je kritika výzkumů oprávněná, co užitečného výzkumy přinesly a kde mohou potenciálně škodit, v čem je možno se z nich poučit a kde je naopak třeba se mít na pozoru. Jejím hlavním cílem je však přispět ke zvýšení užitečnosti implementace mezinárodních výzkumů výsledků vzdělávání v České republice.

Autorka vychází z přesvědčení, že lepšího využití zdrojů, které mezinárodní výzkumy výsledků vzdělávání nabízejí, nelze docílit bez komplexního a důkladného poučení o jejich historii, deklarovaných cílech, metodologických aspektech, vývoji i o kritice, které čelí. Výzkumník, který pracuje s daty z těchto výzkumů, musí rozumět jejich povaze a metodám, které je možno použít k jejich zpracování, a zároveň si musí být vědom omezení, která jsou dána metodami jejich sběru. Detailní poučení potřebuje i odborník, který interpretuje zveřejněné výsledky v kontextu českého vzdělávacího systému, i ten, kdo hledá v mezinárodních výzkumech inspiraci pro národní šetření a analýzy, nebo pro svůj další profesní rozvoj.

Mnoho informací o mezinárodních výzkumech výsledků vzdělávání již bylo v ČR publikováno. Dosud však zde nevyšla monografie, která by pojednávala o mezinárodních výzkumech výsledků vzdělávání komplexně a zabývala se všemi výše uvedenými aspekty. Naše publikace si klade za cíl tuto mezeru zaplnit. Detailní informace o cílech výzkumů a jejich metodologii jsou uvedeny na začátku publikace, neboť bez těchto informací nelze plně docenit obrovský vývoj, který mezinárodní výzkumy v uplynulých 20 letech prodělaly a kterému se podrobně věnuje druhá polovina publikace. Bez poučení o metodologických aspektech výzkumů je také obtížné plně porozumět kritice, které výzkumy čelí, a vytvořit si na ni vlastní fundovaný názor.

Text se zaměřuje na čtyři aspekty mezinárodních výzkumů: na jejich obsah – tedy na testované oblasti, na metodologii jejich provedení, na zpracování dat a na využití výsledků. Zároveň věnuje velký prostor zasazení mezinárodních výzkumů do širšího kontextu pedagogického výzkumu a pojednává o diskusi, která výzkumy provází u nás i v zahraničí.

Sledovat obsah je užitečné z toho důvodu, že na něm je velmi dobře vidět, jak se v průběhu času měnily názory mezinárodní komunity na to, co je důležité měřit a co je možné měřit. Testy v mezinárodních výzkumech jsou velmi dobře dokumentovány – koncepční rámce velmi precizně definují testované vědomosti a dovednosti a také vysvětlují, proč právě tyto vědomosti a dovednosti jsou natolik důležité, že má smysl v mezinárodních výzkumech ověřovat, nakolik

si je žáci osvojili. Na měnícím se obsahu testů je tedy možno pozorovat, jak se v mezinárodní komunitě vyvíjejí pohledy na cíle vzdělávání a na jejich význam.

Metodologie výzkumů je důležitá pro porozumění výsledkům a jejich správné interpretaci a pro řádné zpracování dat v mezinárodních i národních analýzách. Datové soubory z mezinárodních výzkumů jsou mimořádně kvalitní, obsahují množství zajímavých dat získaných na pečlivě kontrolovaných reprezentativních výběrech. Soubory jsou veřejně k dispozici. Jejich komplexnost však vede k tomu, že data mnohdy nejsou analyticky plně vytěžena. Například čeští výzkumníci často sbírají vlastní (mnohdy nekvalitní) data, místo toho, aby k zodpovězení svých výzkumných otázek využili již existující datové soubory. Dalším důvodem pro seznámení s metodologií mezinárodních výzkumů je její užitečnost pro národní aktivity, které by mohly být využitím některých moderních metodologických postupů významně obohaceny.

Důležité je se poučit nejen o metodologii získávání dat, ale také o metodách používaných k jejich zpracování pro zodpovězení různorodých výzkumných otázek. Mezinárodní výzkumy mají velký význam pro formování vzdělávacích politik. Podle jejich výsledků je posuzována úspěšnost vzdělávacích systémů a významně formují obsahy vzdělávání i způsoby ověřování výsledků vzdělávání na národních úrovních. Je velmi užitečné se seznámit s typem analýz, ke kterým mohou být data z mezinárodních výzkumů použita, a s vývojem analytických metod, který umožňuje jejich čím dál tím sofistikovanější vytěžení. Je také zajímavé dozvědět se více o tom, jak jsou efekty mezinárodních výzkumů na vzdělávání reflektovány rozmanitými aktéry a jaká opatření jsou volena ke zvýšení užitečnosti výzkumů a naopak k eliminaci jejich možných negativních dopadů.

Metodologie výzkumů, interpretace získaných dat a dopady na vzdělávací politiku jsou předmětem bouřlivých diskusí v mezinárodní i v české odborné komunitě. Kritika jednotlivých aspektů výzkumů ovlivnila jejich dosavadní vývoj a určuje, jaké bude jejich směřování v budoucnu. Z toho důvodu jsou v publikaci podrobně představeny kritické hlasy a jsou pojednány i možné cesty k eliminaci kritizovaných nedostatků.

Členění publikace odpovídá výše uvedeným tématům. První část publikace podává přehled historie mezinárodních výzkumů v průběhu 65 let jejich existence. Nejprve popisuje aktivity Mezinárodní asociace pro hodnocení výsledků vzdělávání, která tyto výzkumy zahájila a zásadně ovlivnila jejich podobu. Následuje popis přemýšlení, které utvářelo koncept mezinárodních výzkumů výsledků vzdělávání OECD, jež jsou aktuálně nejvýznamnějšími hybateli vzdělávacích politik ve vyspělých zemích, a přehled výzkumů OECD v žákovské i dospělé populaci. Kapitola dále popisuje stručně jednotlivé kroky provedení výzkumů a zveřejňování jejich výsledků.

Druhá část publikace je věnována vývoji koncepcí hlavních měřených oblastí od devadesátých let do současnosti ve výzkumech obou realizátorů. Ukazuje, jak se změnily konceptuální rámce čtenářské, matematické a přírodovědné gramotnosti od 90. let do současnosti a čím byly změny motivovány. Zároveň popisuje vývoj v oblasti řešení problémů, která se stala nedílnou a důležitou součástí výzkumu PISA, a zmiňuje se i o dalších oblastech, které jsou v rámci výzkumů ověřovány.

Třetí část publikace se zaměřuje na vybrané metodologické aspekty mezinárodních výzkumů. Je motivována zejména vědomím, že metody používané v mezinárodních výzkumech nejsou v české odborné komunitě běžně používány a jejich komplexnost často brání výzkumníkům pracovat s daty z těchto výzkumů. Kapitola si klade za cíl napomoci tomu, aby čeští výzkumníci dat více využívali a aby s nimi pracovali řádně. Věnuje se koncepci testů a dotazníků, výběru respondentů a vážení dat, způsobu odhadu výsledků a jejich prezentace. Na příkladech ukazuje, jakých chyb se můžeme snadno dopustit, pokud nezohledníme povahu získaných dat. Závěr

kapitoly popisuje, jak by měli výzkumníci s daty řádně pracovat s využitím nástrojů, které jsou pro práci s daty volně dostupné.

Čtvrtá kapitola přehledně pojednává o vývoji mezinárodních výzkumů, pokud jde o konstrukce výzkumných nástrojů, používané metody, ověřované cíle a využití výsledků. Druhá část kapitoly zasazuje mezinárodní výzkumy do širšího kontextu výzkumů efektivit vzdělávání, zabývá se využitím dat z mezinárodních výzkumů pro vzdělávací politiku a představuje hlavní typy aktuálně využívaných analýz. Pojednává i o zpracování dat z mezinárodních výzkumů českými výzkumníky.

Pátá kapitola se zamýšlí nad výzvami, kterým mezinárodní výzkumy aktuálně čelí. Podrobně představuje diskusi, která se o silných a slabých stránkách mezinárodních výzkumů a jejich dopadech na vývoj vzdělávacích systémů vede ve světě i v České republice. Ve druhé části naznačuje cesty k překonání kritizovaných nedostatků a nastiňuje budoucí vývoj z pohledu předních odborníků, kteří se mezinárodními výzkumy výsledků vzdělávání zabývají.

Šestá kapitola představuje vybraná důležitá zjištění, která přinesly mezinárodní výzkumy za dobu své dvacetileté realizace v České republice. Zaměřuje se zejména na zjištění, kterým by podle názoru autorky měla být věnovaná další pozornost ze strany výzkumníků i tvůrců vzdělávací politiky.

Závěr publikace diskutuje silné a slabé stránky mezinárodních výzkumů výsledků vzdělávání z pohledu informací uvedených v publikaci. Zamýšlí se nad přijetím konceptu výzkumů českou pedagogickou komunitou a nad možnostmi jejich lepšího využití v České republice. Ukazuje, ve kterých konkrétních oblastech by se mohli čeští výzkumníci mezinárodními výzkumy inspirovat.

Autorka děkuje Jaroslavě Simonové, Ph.D., ing. PhDr. Petrovi Soukupovi, doc. PhDr. Arnoštovi Veselému, Ph.D., a prof. PhDr. Elišce Walterové, CSc., za cenné připomínky k textu.

1. Výzkumy výsledků vzdělávání: historie a používané postupy

Tato kapitola podává stručnou a přehlednou formou obecné informace o mezinárodních výzkumech výsledků vzdělávání¹. Popisuje historii vzniku výzkumů a jejich prvotní rozvoj v polovině 90. let v aktivitách Mezinárodní asociace pro hodnocení výsledků vzdělávání. Navazuje popis přemýšlení, které vedlo k zahájení cyklu výzkumu PISA, a přehled hlavních charakteristik tohoto výzkumu. Informaci o výzkumech v žákovské populaci doplňuje stručná informace o výzkumech v dospělé populaci. Další části kapitoly informují o hlavních krocích realizace mezinárodních výzkumů výsledků vzdělávání v žákovské populaci a o postupech při zveřejňování výsledků.

1.1 Využití standardizovaných testů k hodnocení výsledků vzdělávání

Pro přemýšlení o mezinárodních výzkumech výsledků vzdělávání je užitečné zasadit je do obecného kontextu hodnocení ve vzdělávání. Hodnocení probíhá ve vzdělávacích systémech na několika úrovních – na úrovni systému, případně nějakého menšího správního celku (např. regionu), na úrovni školy, na úrovni ředitele, učitele a na úrovni žáka. V centru pozornosti stojí tradičně hodnocení žáka. O hodnocení žáka hovoříme v situaci, kdy jsou cíleně a systematicky shromažďovány doklady o učení žáka s cílem vyslovit nějaký závěr o žakově učení a jeho výsledcích. Ve vztahu k hodnocení rozlišujeme zpravidla mezi sumativním hodnocením (hodnocením učení) a formativním hodnocením (hodnocení pro učení). Cílem sumativního hodnocení je sumarizovat výsledky učení s cílem žáka oznámkovat nebo certifikovat. Cílem formativního hodnocení je zmapovat vývoj učení s cílem naplánovat následující učební postup. V případě formativního hodnocení se setkáváme také s hodnocením diagnostickým, jehož cílem je identifikovat žákovy potřeby na začátku nějaké učební etapy a navrhnout vhodný program (např. Morris 2011).

Hodnocení může mít interní nebo externí povahu. O interním hodnocení hovoříme tehdy, když je vytvářeno a prováděno ve škole, nejčastěji učitelem žáka. Externí hodnocení je oproti tomu vyvinuto a hodnoceno mimo školu a je u něj pečlivě dbáno o to, aby hodnotící úlohy, podmínky administrace, hodnocení a interpretace výsledků byly pro všechny žáky identické tak, aby výsledky byly srovnatelné a hodnocení objektivní. Externí sumativní hodnocení má nejčastěji podobu standardizovaného testu, který je v české literatuře často označován jako test didaktický. O standardizovaném testu hovoříme v situaci, kdy se jeho příprava, administrace a interpretace výsledků řídí předem danými pravidly zajišťujícími jeho kvalitu a objektivitu (podrobněji např. Chvál a kol. 2015). Standardizované testování je často používáno v situacích, které jsou pro žáka důležité. Zpravidla ovlivňuje rozhodování o žakově další vzdělávací dráze. Jedná se například o situace, kdy žák skládá závěrečnou zkoušku a úspěch v testu rozhodne, zda získá certifikát (například maturitní test), nebo skládá přijímací zkoušku, jejíž výsledek

¹ V angličtině jsou tyto výzkumy označovány jako *International Large Scale Assessments (ILSAs)*. Výstupy, které hodnotí, jsou označovány různě: jako vědomosti, dovednosti, kompetence nebo gramotnosti. Vzhledem k nejednotnosti terminologie v textech, které se těmito výzkumy zabývají a popisují jejich cíle a konceptuální rámce, není ani v této publikaci terminologie zcela jednotná. Nejčastěji jsou však všechny tyto výstupy označovány souhrnně jako výstupy kognitivní. Ty jsou zde odlišovány od výstupů nekognitivních (jako jsou např. postoje, motivace, chování a hodnoty a sociální a personální dovednosti, respektive kompetence). Některé nekognitivní výstupy, o kterých publikace pojednává, jsou v rámci mezinárodních výzkumů výsledků vzdělávání zjišťovány prostřednictvím dotazníků, jiné nekognitivní výstupy ještě zjišťovány nejsou, přestože jsou výzkumníky považovány za velmi důležité.

rozhodne o přijetí na zvolenou školu. V těchto situacích je výsledek zkoušky pro žáka důležitý, má pro něj nějaké důsledky (v angličtině jsou tyto zkoušky označovány jako *high stakes*, z anglického *stakes* – sázky).

Idea objektivního testování se zrodila v psychologii na přelomu 19. a 20. století, a to s rozvojem inteligenčních testů a dalších nástrojů na zjišťování schopností. Ve své původní podobě sestávaly objektivní testy z položek, ve kterých respondenti netvořili vlastní odpovědi, ale vybírali správnou odpověď z několika nabízených možností. Objektivní testy našly rozsáhlé použití v armádě: v obou světových válkách byly používány pro klasifikaci vojenských adeptů. Uplatnily se rovněž v průmyslu a obchodu, kde začaly být hojně využívány při nábořech zaměstnanců. Ve Spojených státech pronikly záhy také do škol a staly se hlavním nástrojem hodnocení žáků. Bylo na nich oceňováno zejména to, že velmi levně a v krátkém časovém období ohodnotí velké množství respondentů. Umožňují rovněž zařadit mnoho testových položek a tak, přinejmenším povrchně, pokrýt rozsáhlé učební celky. Americká zkušenost se rozšířila do celého světa pomocí ukázkových testů, odborné literatury a kurzů organizovaných americkými univerzitami; testy postupně našly uplatnění i v řadě dalších vzdělávacích systémů (Nevo 1995). S tím, jak testování nabývalo na popularitě, začínalo se mu věnovat více a více odborníků, stalo se předmětem výzkumu a vývoje a bylo zařazeno i do programu vzdělávání učitelů. Vývoj testu se postupně stal velmi komplikovanou záležitostí, která se řídila řadou složitých pravidel a teorií a stala se pro učitele mnohdy značně nesrozumitelnou. Vývoj testů tak začal probíhat zcela odtrženě od výuky, jejíž výstupy měly testy hodnotit.

Po období masivního nástupu testů do škol nastalo v 80. letech v řadě zemí severní Ameriky a západní Evropy období, ve kterém se staly testy předmětem velké kritiky zejména ze strany pedagogů a školských odborníků. Vyčítali jim hlavně odtrženost od procesu vzdělávání (tj. od školy i od života) a zaměření na nedůležité aspekty vzdělávání (tj. ty, které je možno zjišťovat pomocí objektivních testů). Někteří odborníci označovali v této souvislosti objektivní testy jako „urážku inteligence“ (Nevo 1995). Důležitým argumentem proti využívání testů byl jejich negativní vliv na výuku. Odborníci upozorňovali na to, že nikoli aktivní myšlení, ale zaměření se na faktické vědomosti a rutinní postupy vede k podpoře výuky, která je moderními vzdělavateli velice kritizována. Předmětem kritiky byl i vliv testování na roli učitelů, kteří byli zbavováni zodpovědnosti za to, co děti učí, a stávali se jen pasivními vykonavateli cizí vůle (tedy tvůrců testů, kteří získávali v systému neúměrný vliv).

Diskusi o negativních důsledcích plošného testování se od samého počátku dostalo v pedagogických kruzích velké pozornosti a v systémech, které plošné testování využívají, probíhá s nezměněnou intenzitou stále. Tato diskuse neměla za následek zastavení plošného testování, neboť to se mezitím stalo důležitým prostředkem sloužícím k informování veřejnosti a posuzování práce školy i efektivity celého vzdělávacího systému. Je však kontinuálním prostředkem k hledání alternativních nástrojů objektivního hodnocení a k eliminaci negativních důsledků plošného testování.

Externě administrované standardizované testy se ovšem nepoužívají pouze v situacích, kdy rozhodují o dalším osudu žáka. Používají se i pro účely monitoringu vzdělávacích výsledků na úrovni školy, správních celků a celých systémů nebo pro účely evaluace vzdělávacích programů nebo dopadů opatření vzdělávací politiky. V případě monitoringu sice testy nemají dopad na osudy žáků, ale často mají dopad na osudy škol, případně učitelů. V tomto kontextu se setkáváme jak s testy, které jsou administrovány všem žákům ve vzdělávacím systému v nějakém ročníku školní docházky (hovoříme o plošném testování), nebo s testy, které jsou administrovány pouze na výběrových souborech žáků.

V současné době již testy zpravidla neobsahují pouze úlohy s výběrem jediné správné z několika nabízených odpovědí, ale umožňují žákům vytvářet odpovědi vlastní, které jsou vyhodnocovány jednotným způsobem tak, aby hodnocení bylo co nejobjektivnější.

Morris (2011) identifikovala několik důvodů pro používání standardizovaných testů: Nejčastěji se s testy setkáváme v souvislosti s *new public managementem*, který spočívá v zavádění postupů ze soukromého sektoru do veřejných služeb. Jeho cílem je zvýšit efektivitu a snížit náklady prostřednictvím větší decentralizace a autonomie, s tím, že se aktéři (v našem případě školy, správní celky nebo celé vzdělávací systémy) musejí zodpovídat ze svých výsledků občanům, kteří jejich fungování platí ze svých daní. Součástí tohoto přístupu bývá často zveřejňování žebříčků škol nebo financování škol podle výsledků žáků. V některých případech nedochází ke vzájemnému porovnávání výkonů jednotlivých žáků, ale výsledky žáků jsou posuzovány vzhledem k nějakému standardu, který představuje očekávanou úroveň vědomostí a dovedností v daném věku, a oblasti vzdělávání a výsledky testů mají ukázat, zda žáci tohoto standardu dosahují a jak se míra dosahování vyvíjí v čase. Důležitým impulsem pro používání standardizovaných testů jsou mezinárodní výzkumy, které ukazují rozdíly mezi zeměmi a motivují vlády k tomu, aby se snažily udržet krok s mezinárodními standardy. Motivem pro zavádění standardizovaných testů je také snaha hodnotit dovednosti pro 21. století, která se nyní projevuje zejména v zavádění informačních technologií do hodnocení s tím, že v budoucnosti technologie pomohou tyto dovednosti měřit. Důležitým činitelem při používání standardizovaných testů je testovací průmysl. V mnoha zemích jsou společnosti, které produkují testy, schopny přesvědčit vládu, že se má spolehnout na jejich testy při monitorování a evaluaci vzdělávacího systému. Tyto firmy mají velký vliv na rozhodování, zda budou standardizované testy používány či nikoli.

Při použití standardizovaných testů je vždy důležité přesně stanovit účel jejich použití. Zpravidla rozlišujeme mezi čtyřmi účely. Nejčastějším je monitoring a evaluace vzdělávacího systému, který posuzuje, zda žáci dosahují očekávaných výsledků a zda se výsledky žáků průběžně zlepšují. Zde jsou sice implicitně hodnoceni učitelé, ale také práce tvůrců vzdělávacích politik, úředníků a všech ostatních aktérů zodpovědných za chod celého vzdělávacího systému. Jak již bylo uvedeno výše, často jsou standardizované testy administrovány s cílem přimět školy, aby skládaly účty ze své práce. Někdy je prioritním cílem testování poskytnout informace veřejnosti. V tomto případě bývají zveřejňovány výsledky jednotlivých škol s odůvodněním, že informace o výsledcích žáků jsou získávány za veřejné prostředky a veřejnost má právo být s těmito výsledky seznámena. V některých vzdělávacích systémech nejsou výsledky škol zveřejňovány a školy nejsou za výsledky žáků poháněny k zodpovědnosti; výsledky slouží učitelům a vedení škol jako zpětná vazba o výsledcích jejich žáků.

Tato publikace se zabývá standardizovanými externími testy, které jsou administrovány na výběrových souborech a slouží k monitoringu vzdělávacích systémů, případně nižších správních celků. Nemají tedy přímé důsledky ani pro žáky ani pro zúčastněné školy. Jejich specifikum spočívá v tom, že jsou vyvíjeny v mezinárodní spolupráci a zadávány jednotně ve všech zúčastněných zemích. Jejich rozvoj byl stimulován nástupem masivního využívání testů k hodnocení výsledků vzdělávání, který byl popsán výše. V dalších částech této kapitoly se podíváme blíže na historii aktivit dvou nejznámějších tvůrců mezinárodních standardizovaných šetření.

1.2 Aktivity Mezinárodní asociace pro hodnocení výsledků vzdělávání

1.2.1 Začátky mezinárodních výzkumů – iniciativa akademické obce²

Mezinárodní asociace pro hodnocení výsledků vzdělávání (*International Association for the Evaluation of Educational Achievement – IEA*) byla právně ustanovena v roce 1967, ale její práce ve skutečnosti začala již v roce 1958. Vznikla z iniciativy vzdělavatelů, psychologů, sociologů a psychometriků, kteří se v *Institutu pro vzdělávání organizace UNESCO* zabývali problematikou hodnocení práce školy a výsledků žáků. Zakladatelé, mezi které patřili například Benjamin Bloom a Torsten Husén, se shodovali v tom, že efektivní evaluace vzdělávacích systémů vyžaduje nejenom do té doby běžnou analýzu vstupů, ale též důkladnou analýzu výstupů. Svět považovali za přirozenou laboratoř, ve které jednotlivé vzdělávací systémy hledají své vlastní cesty k zajištění optimálního vzdělávání pro mladou generaci. Byli přesvědčeni o tom, že pokud se podaří získat dostatečné množství informací o dostatečně rozmanité skupině zemí, bude možno jejich analýzou odhalit důležité zákonitosti, které v rámci jednotlivých systémů není možno odhalit. Jejich cílem bylo identifikovat faktory, které mají konsistentní a významný dopad na výsledky vzdělávání. Zakladatelé IEA se rekrutovali z akademické obce. Výzkumy IEA proto měly vždy spíše akademický než politický charakter a snažily se postihnout realitu vzdělávacího procesu v celé jeho složitosti a šíři. Situace se změnila až v posledních letech, kdy díky rostoucím nákladům a nutnosti koordinovat mezinárodní výzkumy různých organizátorů jsou jednotlivé země i v IEA stále častěji reprezentovány tvůrci vzdělávacích politik, kteří mohou zajistit jejich financování, a nikoli akademiky.

První, pilotní šetření IEA známé jako *Pilot Twelve-Country Study*, bylo uskutečněno v roce 1960 a kladlo si za cíl ověřit možnosti mezinárodního měření výsledků vzdělávání. V rámci tohoto šetření byli testováni třináctiletí žáci ze 12 zemí z matematiky, čtenářské gramotnosti, zeměpisu, přírodovědných předmětů a studijních předpokladů (*non-verbal ability*). Pilotní šetření potvrdilo, že je možné, aby spolu při sběru dat efektivně spolupracovalo několik výzkumných center z různých zemí, a že je možno zkonstruovat takové testy a dotazníky, které budou fungovat v různých vzdělávacích systémech a kulturách.

Na pilotní studii navázal první výzkum matematického vzdělávání – *First International Mathematics Study (FIMS)*, který zahrnoval sofistikovanější metody výběru respondentů a obsáhlejší test. Data byla získána od dvou žákovských populací (13letí a žáci v posledním ročníku středoškolského studia). Výzkum poprvé začal operovat s pojmem „*opportunity to learn*“: snažil se postihnout míru, v jaké mají žáci příležitost získat ve výuce vědomosti a dovednosti, které jsou v rámci výzkumu zjišťovány. Tento koncept byl následně součástí všech realizovaných výzkumů IEA (podrobněji se tímto konceptem zabýváme dále). Výzkum dále ukázal, že všechny vzdělávací systémy vykazují určitou míru vzdělanostních nerovností. Sledování nerovností týkajících se vzdělávacích příležitostí a výsledků se od té doby stalo jedním z ústředních témat mezinárodních pedagogických výzkumů.

Následující výzkum *Six Subject Survey (1970–1971)* je někdy nazýván také *First International Science Study (FISS)*, neboť zdokonalil koncept měření výsledků v přírodovědných předmětech. Kromě přírodovědných vědomostí a dovedností zjišťoval ještě čtenářskou gramotnost a znalost literatury, znalost angličtiny a francouzštiny jako cizích jazyků a vědomosti, dovednosti a postoje

² Výborným zdrojem informací o historii IEA, z kterých čerpá i tato kapitola, jsou webové stránky asociace: http://www.iea.nl/brief_history.html.

v občanské výchově. Cílová populace byla posunuta ze 13 na 14 let, neboť v té době bylo ve všech zúčastněných zemích vzdělávání povinné do 14 let. Zároveň byly testovány populace desetiletých žáků (tedy pro většinu zemí poslední ročník primárního vzdělávání) a žáků v posledním ročníku středoškolského studia. Výzkum pomohl identifikovat nové prediktory výsledků, jmenovitě zájmy, motivaci a postoje, dále výukové metody a školní organizaci a postupy.

1.2.2 Periodická šetření matematických a přírodovědných vědomostí a dovedností

Na konci 70. let se zrodil nápad začít organizovat výzkumy periodicky, aby bylo možno měřit změny v průběhu času. Ve 20, respektive ve 24 zemích se v první polovině 80. let uskutečnily výzkumy *Second International Mathematics Study (SIMS)* a *Second International Science Study (SISS)*. Strategie opakování výzkumu po několika letech poskytla zúčastněným zemím důležitou informaci o trendech v matematických a přírodovědných vědomostech a dovednostech.

Výzkum *Third International Mathematics and Science Study (TIMSS)*, jenž byl realizovaný v roce 1995, byl prvním z výzkumů, ve kterých se testovala matematika a přírodovědné předměty společně. Zahájil čtyřletý cyklus nyní známý pod názvem *Trends in International Mathematics and Science Study*. Výzkumu v roce 1995 se zúčastnilo 46 zemí a byly testovány všechny 3 populace (10letí, 14letí a žáci v posledním ročníku středních škol) (např. Beaton a kol. 1997a, 1997b, Martin a kol. 1998, Mullis a kol. 1998). Do šetření se poprvé zapojily země střední a východní Evropy, které se na počátku 90. let staly členskými zeměmi asociace IEA. Tak tomu bylo i v případě ČR. Výjimku tvořily Maďarsko a Polsko, které s asociací spolupracovaly i v socialistické éře (Maďarsko bylo dokonce zakládajícím členem IEA)³.

V šetření žáků středních škol se poprvé objevil koncept matematické a přírodovědné gramotnosti. Tento koncept byl zaveden proto, že různé střední školy mají velmi různé matematické a přírodovědné kurikulum. Jakýsi základ – matematická a přírodovědná gramotnost – by měl však být podle tvůrců výzkumu rozvíjen v rámci všech vzdělávacích programů. Matematická a přírodovědná gramotnost zde byla chápána jako orientace v matematické a přírodovědné problematice a schopnost používat naučených vědomostí a postupů v situacích běžného života. Některé země testovaly také žáky specializující se na úrovni vyššího sekundárního studia (ISCED 3) na matematiku a na fyziku. Na tuto zkušenost navázal později cyklus šetření *TIMSS Advanced*, který byl zahájen v roce 2008.

Součástí výzkumu TIMSS 1995 byla rovněž rozsáhlá analýza učebnic a kurikula v matematice a přírodovědných předmětech (Schmidt a kol. 1997a, Schmidt a kol. 1997b). Cílem analýzy bylo zmapovat zamýšlené kurikulum (*intended curriculum*), přičemž vlastní výzkum zjišťoval kurikulum realizované (*implemented curriculum*) a kurikulum dosažené (*attained curriculum*). I když tak rozsáhlá analýza učebnic a kurikula jako v roce 1995 v dalších letech již provedena nebyla, koncept tří úrovní kurikula je charakteristický pro všechny studie IEA, jak již bylo zmiňováno dříve.

³ Zdrojem informací o nejvýznamnějších výzkumech asociace IEA – výzkumu matematického a přírodovědného vzdělání TIMSS a výzkumu čtenářské gramotnosti PIRLS – jsou stránky mezinárodního koordinačního centra na Boston College <http://timssandpirls.bc.edu/>. Na stránkách jsou vystaveny publikace výsledků všech kol výzkumů, zároveň se tam nacházejí informace o metodologii výzkumů, použité dotazníky a mezinárodní databáze. Informace o české realizaci TIMSS a PIRLS jsou spolu s národními analýzami, českými výzkumnými nástroji a českými datovými soubory k dispozici na stránkách České školní inspekce <http://www.csicr.cz/Prave-menu/Mezinarodni-setreni/TIMSS> a <http://www.csicr.cz/Prave-menu/Mezinarodni-setreni/PIRLS>.

Zajímavou součástí výzkumu TIMSS v roce 1995 bylo měření experimentálních dovedností 10letých a 14letých žáků. V rámci tohoto měření řešila podmnožina žáků, kteří se zúčastnili testování, několik krátkých experimentálních úloh. V úlohách měli žáci za úkol navrhnout a provést jednoduché matematické a přírodovědné experimenty, zaznamenat výsledky a tyto výsledky vyhodnotit. Výkon žáků byl hodnocen na základě jejich záznamů podle jednotných návodů. Experimentální vybavení bylo jednoduché a pro všechny žáky jednotné. Žáci měli například za úkol zjistit, zda teplota vody ovlivňuje rychlost rozpouštění šumivého celaskonu nebo jak se mění tep při provádění jednoduchého cvičení (viz např. Palečková, Mandíková 1997). Ověřování experimentálních dovedností již nebylo do dalších cyklů výzkumů zařazováno, neboť kladlo zvýšené nároky na administraci a ukázalo se, že měří velmi podobné vědomosti a dovednosti jako písemný test. Realizátoři tedy usoudili, že jeho přidání hodnota neopravňuje navýšení nákladů na realizaci.

Šetření TIMSS se opakuje v pravidelných čtyřletých intervalech již jen v populacích 10letých (4. ročník) a 14letých (8. ročník). Výběr žáků probíhá dvoustupňově s tím, že se nejprve vyberou školy a následně jedna až dvě třídy v příslušném ročníku. Výzkum proběhl v letech 1999, 2003, 2007, 2011 a 2015. ČR se zúčastnila všech šetření s výjimkou šetření v roce 2003. Testování žáků je doplňováno dotazníkovým šetřením mezi žáky a jejich učiteli a řediteli. Vývojem pojetí matematické a přírodovědné oblasti ve výzkumu TIMSS se podrobně zabývá kapitola 2, koncepcí dotazníků kapitola 3.

1.2.3 Periodická šetření čtenářské gramotnosti

V letech 1990–1991 byl zahájen cyklus šetření ve čtenářské gramotnosti. Šetření *Reading Literacy Study* bylo charakteristické svým důrazem nejen na výsledky vzdělávání, ale i na vzdělávací kontext (v tomto případě čtenářské zájmy žáků). Šetření se zúčastnilo 32 zemí a proběhlo opět v populaci žáků 4. a 8. ročníků. Čtenářská gramotnost byla pro účely výzkumu RLS definována jako *dovednost porozumět formám psaného textu požadovaným společností a/nebo ceněným jedincem a používat je*. V rámci výzkumu bylo zjišťováno porozumění 3 typům textů: vyprávění, výkladový text a dokument (nesouvislý text ve formě grafů, obrázků, seznamů, tabulek a podobně). Tyto texty byly zakotveny ve čtyřech typech kontextů: domov, škola, společnost, práce. Úlohy, které byly zahrnuty do mezinárodního srovnání, byly uzavřené, žáci v nich měli za úkol vybrat jedinou správnou z nabízených odpovědí. Ve čtyřech úlohách měli za úkol provést jednoduché doplnění (Elley 1992).

Od roku 2001 probíhají návazné studie čtenářské gramotnosti v pětiletých intervalech pod názvem *Progress in International Reading Literacy Study (PIRLS)*⁴: 2001, 2006, 2011, další studie proběhla v roce 2016. Tyto studie probíhají ve 4. a 8. ročnících. Jejich pojetí se od roku 1991 významně změnilo. Velký podíl úloh je otevřený, tedy žáci mají za úkol vypracovat vlastní odpověď; zároveň jsou úlohy seskupovány kolem delších textů, které jsou pro účely výzkumu velmi pečlivě voleny z textů přístupných příslušné věkové kategorii žáků. Vývoji konceptu čtenářské gramotnosti ve výzkumech IEA se podrobněji věnuje následující kapitola.

Design výzkumu PIRLS je obdobný jako design výzkumu TIMSS s tím, že součástí výzkumu PIRLS je rovněž dotazník pro rodiče testovaných žáků, který zkoumá rozvoj raných čtenářských dovedností a postoje rodičů ke čtenářským aktivitám.

Vzhledem k tomu, že cyklus matematických a přírodovědných studií je odlišný, nejsou pra-

⁴ Informace na <http://timssandpirls.bc.edu/> <http://www.csicr.cz/Prave-menu/Mezinarodni-setreni/PIRLS> (viz poznámka pod čarou 3).

videlně sledovány všechny 3 oblasti souběžně. K tomu došlo v roce 2011, kdy se pětiletý cyklus čtenářských studií a čtyřletý cyklus matematických a přírodovědných studií setkaly. Řada zúčastněných zemí využila této skutečnosti k tomu, že provedla testování ve všech oblastech u stejných žáků, což umožnilo analyzovat vzájemné vztahy mezi nimi.

1.2.4 Méně tradiční oblasti hodnocení zahrnuté do studií IEA

I když je organizace IEA spojována především s výzkumy TIMSS a PIRLS, uskutečnila rovněž řadu dalších inspirativních šetření. Opakovaně byl realizován výzkum občanské výchovy. Výzkum *IEA Civic Education Study* (1998-1999) se zaměřil na 3 oblasti: demokracie (určující rysy, demokratické instituce a aktivity, práva a povinnosti občana), národní identita (včetně regionálních a mezinárodních vztahů) a sociální soudržnost a diskriminace. V rámci těchto tří tematických okruhů byly u žáků zjišťovány jejich znalosti a dovednosti pomocí testů, které zjišťovaly kromě faktických vědomostí i schopnost rozlišit mezi faktem a názorem a interpretovat sdělení s politickým obsahem. Dotazníky zjišťovaly, jak žáci rozumějí pojmům, jako jsou demokracie a občanství, zjišťovaly postoje žáků a jejich očekávání ohledně vlastní účasti na politických aktivitách. Učitelé dotazníky zjišťovaly, jakým oblastem a tématům věnují učitelé občanské výchovy největší pozornost, jaké považují za důležité a jak se cítí pro výuku těchto témat připraveni. Výzkum byl zaměřen na žáky 8. ročníků a žáky na konci středoškolského studia (Tourney-Purta a kol. 2001, Křížová 2001). Výzkum *International Civic and Citizenship Education Study (ICCS)* realizovaný v roce 2009 byl koncipován podobně, tedy umožňoval sledování trendů. Zahrnoval čtyři obsahové oblasti: občanská společnost a systém, občanské principy, občanská angažovanost a občanská identita. Zároveň obsahoval čtyři oblasti afektivní: hodnoty, postoje, zamýšlené chování a faktické chování. Byl zaměřen na žáky 8. ročníků, tedy neumožnil sledovat vývoj v populaci středoškoláků. Další sběr dat výzkumu ICCS proběhl v roce 2016⁵.

V 90. letech se IEA začala věnovat rovněž výzkumu informačních technologií. Výzkum *Second Information on Technology in Education Study* (1998-1999) se zaměřil na mapování vybavení škol informačními technologiemi a jejich využití ve výuce (Pelgrum; Anderson 2011, Plomp a kol. 2003). V roce 2006 potom navázal zkoumáním dopadu investic do vybavení škol informačními technologiemi na kvalitu výuky (Carstens; Pelgrum 2009). V návaznosti na zájem o dovednosti spojené s využíváním informačních technologií jako jedné ze skupiny „dovedností pro 21. století“ byl v roce 2010 zahájen výzkum *International Computer and Information Literacy Study (ICILS)*. Tento výzkum zaměřený na žáky 8. ročníků hodnotil prostřednictvím autentických počítačových úloh porozumění počítačům a dovednost s nimi pracovat. Byly hodnoceny dvě oblasti dovedností. Oblast 1 je zaměřena na získávání informací a jejich zpracování a obsahuje tyto aspekty: znalosti o počítačích a jejich používání, získávání a hodnocení informací a zpracování informací. Oblast 2 je zaměřena na tvorbu a výměnu informací a obsahuje tyto aspekty: transformování informací, tvorba informací, sdílení informací a bezpečné používání informací. Výzkum dále obsahoval dotazníky pro žáky, učitele, ředitele a koordinátory ICT, a sledování kontextu pro výuku ICT v zúčastněných zemích. Další cyklus výzkumu ICILS se uskuteční v roce 2018. Koncepti výzkumů občanské výchovy i informačních technologií podrobněji představuje kapitola 2⁶.

⁵ Podrobnější informace o výzkumech občanské výchovy jsou k dispozici na stránkách Mezinárodní asociace pro hodnocení výsledků vzdělávání <http://www.iea.nl/cived.html> a http://www.iea.nl/iccs_2009.html.

⁶ Informace o výzkumech IEA v oblasti informačních technologií jsou k dispozici na stránkách Mezinárodní asociace pro hodnocení výsledků vzdělávání: <http://www.iea.nl/sites-m1.html>, <http://www.iea.nl/sites-m2.html>, http://www.iea.nl/sites_2006.html, http://www.iea.nl/icils_2013.html. Informace o české realizaci výzkumu ICILS, výsledky

Pro rok 2017 připravuje IEA výzkum v oblasti anglického jazyka *English Teaching and Learning Study*. Výzkum bude obsahovat adaptivní počítačový test, který bude zjišťovat čtení, poslech, písemný projev a mluvený projev v angličtině u žáků, kteří se učí angličtinu jako cizí jazyk. Zároveň poskytne zúčastněným zemím informaci o příležitostech žáků učit se angličtině ve škole i mimo ni, spolu s informací o postojích žáků a učitelů k výuce angličtiny a o používaných výukových metodách.

1.3 Aktivity OECD na poli měření výsledků vzdělávání

1.3.1 Přípravné aktivity pro výzkum OECD PISA – klíčové a mezipředmětové kompetence⁷

V devadesátých letech minulého století došlo v Organizaci pro ekonomickou spolupráci a rozvoj (OECD) k rozhodnutí zahájit hodnocení výsledků školního vzdělávání, které by zjišťovalo, jak vzdělávací systémy vyspělých zemí připravují své mladé lidi pro uplatnění na trhu práce a v osobním a společenském životě. OECD vydává každoročně publikaci *Education at a Glance (EaG)*, do které vstupují nejen statistické indikátory, ale rovněž další indikátory charakterizující rozmanité aspekty vzdělávacích systémů⁸. Ukazatele charakterizující výsledky vzdělávání byly do EaG přejímány z výše zmiňovaných výzkumů asociace IEA. Koncem 90. let však OECD navrhla vlastní výzkumný projekt. Neměl zjišťovat, jak si žáci osvojili učivo školních osnov (tak jako to zjišťovaly výzkumy IEA), ale zda mají takzvané *klíčové kompetence*, které jim umožní vést úspěšný a spokojený život v moderní společnosti. Tento záměr byl reakcí na přesvědčení odborné komunity soustředěné kolem aktivit OECD, že škola čím dál tím více zaostává za potřebami života a že pro úspěch v životě a na trhu práce nejsou aktuální školní znalosti důležité. Důležité jsou právě klíčové kompetence. I když se tvůrci vzdělávacích politik zapojení v práci OECD shodovali v tom, že je třeba mezinárodní měření zacílit na klíčové kompetence, neměli k dispozici jejich vymezení ani sdílené pochopení. Konceptualizaci výzkumu PISA tedy předcházela řada aktivit, jejichž cílem bylo tyto kompetence vymezit, tedy popsat vědomosti, dovednosti a postoje, které jsou důležité pro úspěch v životě a na moderním pracovním trhu.

K vymezení klíčových kompetencí měl významně přispět projekt *DeSeCo (Defining and Selecting Key Competencies)*, který byl zahájen v roce 1996. Jeho cílem bylo vymezit klíčové kompetence prostřednictvím syntézy pohledů předních světových odborníků z různých oborů lidského konání, kteří byli v rámci projektu osloveni. Projekt shromáždil několik desítek rozmanitých pohledů a přístupů k uchopení problematiky klíčových kompetencí. Na základě syntézy všech pohledů vydali autoři projektu DeSeCo doporučení pro práci OECD v oblasti hodnocení dosažené úrovně klíčových kompetencí.

Vymezení pojmu kompetence provedl Franz E. Weinert ve svém pojednání *Concepts of Competence* (Weinert 2001). Uvádí, že se výraz kompetence používá v běžném jazyce i ve vědecké terminologii v mnoha rozmanitých významech. Chceme-li s konceptem kompetencí pracovat, nemůžeme se spolehnout na jeho sdílené pochopení, ale musíme jej úžeji definovat. Kompetence pro účely školního vzdělávání autor doporučuje vymezit pomocí úkolů, ke kterým má vzdělání

národních analýz, české verze výzkumných nástrojů a české datové soubory jsou k dispozici na stránkách České školní inspekce <http://www.csicr.cz/Prave-menu/Mezinarodni-setreni/ICILS>.

⁷ Tyto aktivity byly popsány rovněž publikacích Straková 2008 a Straková 2013.

⁸ Poslední vydání ročenky je k dispozici na stránkách <http://www.oecd.org/edu/education-at-a-glance-19991487.htm>.

připravovat. Tato vymezení jsou dostupná v kurikulu, testových nástrojích, popisech profesních kvalifikací a typických životních situací (ekonomické, administrativní a politické aktivity, interakce s médii, volnočasové aktivity, cestování, společenské chování, apod.) a v popisech specifických činností (např. obchodních činností, her – např. šachů a podobně). Na základě dostupných vymezení je možno popsat kognitivní kompetence, které jsou nezbytné pro dobré zvládnutí každodenních školních a pracovních požadavků. Tento přístup sebou nese eliminaci složky intelektových předpokladů a omezení konceptu kognitivních kompetencí na naučené vědomosti, dovednosti a odpovídající metakognitivní znalosti.

Některé z dovedností a kompetencí jsou užitečné v mnoha učebních situacích a pomáhají při řešení řady problémů. Sem patří například znalost mateřského jazyka, cizích jazyků, čtenářské kompetence a písemné vyjadřování, matematické dovednosti a téměř všechny metakompetence. Weinert zdůrazňuje, že tyto klíčové kompetence nemohou nicméně nahradit specifické vědomosti, a to zejména v případech řešení komplikovanějších úloh. Utopická je podle jeho názoru i představa, že člověk může pomocí několika klíčových kompetencí získat tyto specifické vědomosti v elektronických médiích. Pouze ti, kteří mají dostatečné počáteční vědomosti, jsou schopni efektivně využívat nových vědomostí. Ve vztahu ke kompetencím hraje velmi důležitou úlohu motivace, a to nejen při řešení aktuálního problému, ale i při dlouhodobém získávání odbornosti. Všechna hodnocení kompetencí by měla obsahovat hodnocení motivace, která se k dané kompetenci váže.

Výstupy projektu DeSeCo navrhuji definovat kompetenci jako *dovednost dostat požadavkům nebo úspěšně splnit úkol, který má kognitivní i nekognitivní složky*. Kompetence podle této definice zahrnují kognitivní a praktické dovednosti, vědomosti, motivace, hodnotové orientace, přístupy, emoce a další sociální a behaviorální složky, které mohou být společně mobilizovány pro efektivní akci (např. kooperativní kompetence sestává z vědomostí, kognitivních dovedností, praktických dovedností, přístupů, emocí, hodnot a motivací, které se k této kompetenci vztahují). Autoři doporučují se soustředit na kompetence, které jsou důležité v různých oblastech života a které přispívají k celkově úspěšnému životu a dobře fungující společnosti. Podle jejich závěrů jsou kompetence pozorovatelné pouze v konkrétních akcích, které realizují osoby v konkrétních situacích a kontextech. Jsou získávány a rozvíjeny v průběhu celého života v různých institucích a prostředích.

Volba klíčových kompetencí závisí na tom, co uznává ta která společnost (odsud volíme ty, které posilují společenský, ekonomický a osobní úspěch). Normativní bázi pro výběr by měly tvořit lidská práva, demokratické hodnoty a snaha o udržitelný rozvoj. Mezi cíle na individuální úrovni patří: přístup k ekonomickým zdrojům, politická participace, přístup k intelektuálním zdrojům, zapojení do sociálních sítí, blízké vztahy s druhými, fyzické a psychické zdraví, osobní bezpečnost a pocit úspěchu a radosti. Na úrovni společenské pak: mír a bezpečnost, rovnost bez diskriminace, ekonomická produktivita, demokratické procesy, společenská soudržnost, lidská práva a udržitelný rozvoj.

Podle názorů autorů vyžaduje dnešní doba kritické myšlení a přemýšlivý, holistický přístup k životu, který se vyznačuje vzrůstající komplexitou a vzájemnými souvislostmi mezi aktuálními jevy (vzrůstající nerovnosti, „digital divide“⁹, environmentální problémy, ekonomické změny, politická nestabilita, fragmentace komunit, nové prostředky komunikace apod.) Člověk musí být schopen pohlédnout s odstupem na požadavky, které na něj klade prostředí, učinit samostatný úsudek, řídit své vlastní konání a poměřovat je svými city, myšlenkami, hodnotami, aktivně

⁹ Rozdělení společnosti podle přístupu k informačním technologiím.

jednat a nenechat se vláčet okolnostmi a jinými osobami, určovat svůj život. Kompetence, které byly vybrány v rámci projektu DeSeCo, směřují k takovému jednání. Jejich výběr je založen na demokratických principech a lidských právech, jsou mezinárodně platné a jsou nezbytné k potýkání se s komplexností moderního života (Rychen; Salganik 2001):

Jednat autonomně

- Schopnost bránit a uplatňovat svá práva, zájmy, potřeby, zodpovědnosti a omezení
- Schopnost vytvářet a realizovat životní plány a osobní projekty
- Schopnost jednat v rámci širšího kontextu

Používat interaktivně¹⁰ nástroje

- Schopnost interaktivně používat jazyk, symboly a text
- Schopnost interaktivně používat vědomosti a informace
- Schopnost interaktivně používat (nové) technologie

Fungovat v sociálně heterogenních skupinách¹¹

- Schopnost udržovat dobré vztahy s druhými
- Schopnost spolupracovat
- Schopnost zvládnout a vyřešit konflikt

Autoři uvádějí, že kompetence se zpravidla uplatňují v různých situacích nikoli jednotlivě, ale v různých kombinacích, kde nabývají dle situace různý význam (např. pro udržitelný rozvoj je nejdůležitější kompetence jednat v rámci širšího kontextu a používání vědomostí a informací, zatímco pro pocit úspěchu může být nejdůležitější tvorba a realizace vlastního plánu, práce s technologiemi a kooperace).

Systém kompetencí vytvořených v rámci projektu DeSeCo byl v mnoha ohledech inspirativní, většina z nich však nebyla operacionalizovatelná pro objektivní měření. Výjimkou byly pouze kompetence v prostředním oddílu (interaktivní používání nástrojů), které byly následně rozpracovány do testovaných oblastí.

V rámci přípravných prací pro výzkum PISA došlo v polovině 90. let v rámci pracovní skupiny *Network A* projektu OECD INES ještě k pokusu hodnotit takzvané mezipředmětové kompetence v rámci projektu *Cross Curricular Competencies (CCC)* (OECD 1997). V té době nebyly k dispozici žádné standardizované nástroje a pojetí mezipředmětových kompetencí nebylo v odborné komunitě ustáleno. Odborníci působící v rámci pracovní skupiny *Network A* vyvinuli vlastní nástroje pro měření čtyř oblastí, které podle jejich názorů spadaly do okruhu mezipředmětových kompetencí: 1) politika, ekonomika a občanská výchova (zde byly sledovány kromě znalostí i postoje – postoje k ženským právům, kritický přístup k vládě, důvěra v budoucnost, politické sebevědomí, tolerance, angažovanost), 2) sebepojetí (např. vytrvalost, přesvědčení o vlastních schopnostech, sebepřijetí, úzkost), 3) řešení problémů a 4) komunikace. Na základě pilotáže v 9 zemích byly přijaty koncepty politiky, ekonomiky a občanské výchovy a koncept

¹⁰ Interagujeme se světem pomocí kognitivních, sociálních a fyzických nástrojů. Do této interakce se promítá, jak rozumíme světu, jak jsme kompetentní ve vzájemné interakci s ním, jak snášíme změnu a jak se porovnáváme s dlouhodobými výzvami.

¹¹ Život v multikulturních společnostech a setkávání se s rostoucí individuální a společenskou růzností vyžaduje schopnost se účastnit a fungovat v sociálně heterogenních skupinách a dosáhnout efektivní interakce i s těmi, jejichž osobnost a zázemí jsou odlišné od našich.

sebepojetí. Měření v konceptu politiky, ekonomiky a občanské výchovy probíhala i nadále ve výzkumech IEA (viz 1.2.4), tedy do aktivit OECD nebyla začleněna. Měření některých aspektů sebepojetí bylo zpracováno v rámci výzkumu OECD PISA. V rámci výzkumu PISA došlo také k vývoji konceptu kompetencí k řešení problémů (podrobněji viz kapitola 2).

1.3.2 Výzkum PISA¹²

I když přípravné práce, které předcházely konceptualizaci výzkumu *OECD PISA (Programme for the International Student Assessment – Program pro mezinárodní hodnocení žáků)*, přinesly mnoho zajímavých podnětů a ovlivnily kurikulární politiku mnoha zúčastněných zemí, koncept výzkumu PISA nijak zásadně neovlivnily v tom smyslu, že stejně jako ve výzkumech IEA i ve výzkumu PISA se staly hlavními hodnocenými oblastmi matematika, přírodovědné předměty a čtenářské dovednosti. Na rozdíl od výzkumů IEA zde nebyly koncepční rámce, které definují testovanou oblast a jednotlivé měřené aspekty, vymezovány na základě průniků kurikulů zúčastněných zemí, ale byly vytvářeny mezinárodní expertní skupinou sestávající z odborníků na dané oblasti, které byly označovány jako matematická, přírodovědná a čtenářská gramotnost.

První šetření PISA se uskutečnilo v roce 2000 a od té doby je výzkum realizován pravidelně ve tříletých intervalech (dosud tedy v letech 2000, 2003, 2006, 2009, 2012 a 2015), přičemž jsou vždy ověřovány všechny tři výše zmiňované oblasti, ale na jednu z nich je kladen větší důraz. S každou hlavní hodnocenou oblastí se pak pojí ještě nějaké přídatné zkoumání: se čtenářskou gramotností zkoumání metakognitivních dovedností, s matematickou gramotností zkoumání řešení problémů a finanční gramotnost, s přírodovědnou gramotností zkoumání postojů žáků k vědě a technice a k životnímu prostředí (OECD 2001, 2004a, 2007, 2010a, 2013b).

Konceptuální rámce, které charakterizují hodnocené oblasti, se průběžně vyvíjejí s tím, jak se vyvíjí poznání v těchto oblastech. Ke změnám dochází zpravidla v kole, kdy je daná oblast v hodnocení dominantní. Změny v konceptuálních rámcích a tedy v hodnocení, které je z nich odvozeno, však musejí být prováděny uvážlivě tak, aby byla zachována kontinuita a mohlo dojít k porovnávání výsledků v jednotlivých kolech. Následující kapitola věnovaná vývoji hodnocených oblastí popisuje, jak byly vymezeny hlavní hodnocené oblasti v první realizaci výzkumu v roce 2000 a v posledním šetření v roce 2015, tedy po ukončení celého druhého cyklu. Změny v konceptuálních rámcích odrážejí skutečnost, že od roku 2006 postupně docházelo k zavádění elektronického testování tak, že byly zúčastněným zemím nabízeny elektronické moduly jednotlivých oblastí jako volitelná součást měření. V roce 2015 se již stává elektronická administrace preferovanou alternativou s tím, že tištěné testy slouží pouze pro země, které z nějakých důvodů nejsou schopny realizovat elektronickou administraci.

Výzkum PISA je v každé ze zúčastněných zemí realizován na reprezentativním výběru patnáctiletých žáků (ti, kteří v roce realizace dosáhnou 16 let¹³). Stejně jako ve výzkumech IEA jsou nejprve náhodně vybírány školy ze seznamu všech škol, které žáky v tomto věku vzdělávají. Ve škole však nejsou vybíráni žáci z určitého ročníku a náhodně volené třídy, ale je náhodně

¹² Informace o všech kolech výzkumu PISA včetně odkazů na publikace výsledků, sekundární analýzy, metodologické informace, originální verze testových nástrojů a mezinárodní datové soubory jsou k dispozici na <http://www.oecd.org/pisa/>. Informace o české realizaci výzkumu PISA jsou spolu s národními analýzami, českými výzkumnými nástroji a českými datovými soubory k dispozici na stránkách ČSÍ: <http://www.csicr.cz/Prave-menu/Mezinarodni-setreni/PISA>.

¹³ Definice je v jednotlivých zemích mírně modifikována podle doby testování tak, aby byli všude testováni stejně staří žáci (podrobněji viz oddíl 3.2.1).

vybráno 30 žáků ze všech žáků školy, kteří splňují věkovou definici. I ve výzkumu PISA je testování doplněno dotazníkovým šetřením žáků a ředitelů škol. Neuplatňují se zde rodičovské dotazníky ani dotazníky pro učitele. Ty nejsou zařazovány z toho důvodu, že se testování žáci nenacházejí v jedné třídě a zpravidla nejsou vyučováni stejnými učiteli. Šetření učitelů by v tomto výzkumném designu bylo tedy velmi náročné. Podrobněji o výběru a dotazníkových šetřeních pojednává kapitola 3 věnovaná metodologickým aspektům mezinárodních výzkumů.

1.4 Výzkum studentů vysokých škol

Na setkání ministrů školství zemí OECD v roce 2006 došlo k rozhodnutí pokusit se zrealizovat mezinárodní výzkum výsledků vzdělávání v populaci studentů na konci bakalářského studia pod názvem *International Assessment of Higher Education Learning Outcomes (AHELO)*. Rozhodnutí bylo motivováno snahou vyspělých zemí zajistit kvalitu a relevanci vysokoškolského studia v situaci, kdy dochází k jeho masifikaci. Deklarovanou ambicí bylo koncipovat výzkum jako zpětnovazební a poskytnout vysokým školám informaci o přidané hodnotě, tedy posunu studentů během vysokoškolského studia.

Při vývoji výzkumných nástrojů se museli výzkumníci vyrovnat s problémem, že v zemích OECD neexistuje žádné sdílené vymezení výstupů z vysokoškolského vzdělávání. Rozhodli se pro zařazení přenositelných obecných dovedností a dále dvou disciplín, které mají stabilní sdílené metodologie: ekonomie a inženýrství. V oblasti obecných dovedností zamýšleli organizátoři použít k měření existující nástroje, tedy nevyvíjet pro účely AHELO nový koncept. Nástroj *Collegiate Learning Assessment*, který obsahoval otevřené úlohy, byl doplněn otázkami s výběrem odpovědí z jiných testů obecných dovedností. V průběhu vývoje testu se nicméně ukázalo, že mezi zúčastněnými zeměmi neexistuje shoda na tom, jaké obecné dovednosti by měly být zjišťovány, tedy že bez hledání shody nad nově vyvinutým konceptuálním rámcem se nemůže mezinárodní měření realizovat. Pro účely měření v ekonomice a inženýrství byly vytvořeny expertní skupiny, které vyvinuly konceptuální rámce. Vývoj rámců ukázal, že je možno docílit mezinárodní shody na cílech, které budou ověřovány v těchto oblastech. Následně byly vytvořeny a pilotovány úlohy, které byly vyvinuty ve shodě se schválenými konceptuálními rámci.

V roce 2012 se uskutečnila studie proveditelnosti, která měla za úkol ověřit realizovatelnost výzkumu v různých národních, kulturních, jazykových a institucionálních kontextech. Studie se zúčastnilo 17 zemí. Každá zúčastněná země měla za úkol vybrat 10 vysokých škol, které měly co nejlépe demonstrovat rozmanitost vysokoškolského systému. Studenti a učitelé byli v těchto institucích vybírání pomocí pravděpodobnostního výběru. Studie se zúčastnilo celkem 249 vysokých škol a v nich 4 900 vysokoškolských učitelů a 23 000 studentů, kteří končili bakalářské studium. Testování a dotazování probíhalo elektronicky. Česká republika se do šetření nezapojila.

Studie proveditelnosti ukázala, že je v principu možné vyvinout mezinárodně přijatelné hodnocení v oblastech ekonomie a inženýrství. Hodnocení obecných dovedností se ukázalo jako problematičtější, neboť zaměření na obecné dovednosti na úkor dovedností specializovaných není vlastní všem kulturám.

V průběhu realizace se objevila řada problémů s financováním a organizací, které upozornily na potřebu eventuální další šetření lépe zafinancovat, stanovit jasněji kompetence a komunikaci mezi jednotlivými partnery a vymezit pro přípravu a realizaci delší časové rozmezí. Jako největší problém se ukázala nízká návratnost, zejména ze strany studentů. Řada zúčastněných

zemí navrhovala v budoucnosti nahradit pravděpodobnostní výběr testováním a dotazováním všech studentů v rámci vybraných univerzit nebo jejich částí. I když přání řady zemí OECD realizovat měření v populaci vysokoškoláků stále trvá, nedošlo zatím k rozhodnutí výzkum dopracovat a přikročit k jeho standardní realizaci¹⁴.

1.5 Výzkumy v dospělé populaci

Významný přínos v měření výsledků vzdělávání představovaly nejen výzkumy žáků a studentů, ale i výzkumy v dospělé populaci. Šetření v dospělé populaci jsou zaměřena na podobné oblasti jako šetření v populacích žákovských, oblasti jsou však definovány poněkud odlišně. Zatímco v žákovské populaci jsou alespoň částečně zjišťovány akademické vědomosti, výzkumy dospělé populace se zcela striktně zaměřují na situace praktického života. Rozdíl je asi nejmarkantnější v matematice, kde se vůbec neobjevují „školské“ úlohy, tedy úlohy bez autentické vazby na situace běžného života.

Největší rozdíly však spočívají v provedení výzkumů, kde jsou respondenti vybíráni náhodně z populace dospělých ve věku 16-65 let a testy a dotazníky jsou jim administrovány v domácnostech.

První mezinárodní výzkum vědomostí a dovedností dospělých *International Adult Literacy Survey (IALS)* se uskutečnil v letech 1995–2000. Zjišťoval numerickou gramotnost (*numeracy* – schopnost pracovat s čísly), dokumentovou gramotnost (*document literacy* – schopnost pracovat s nesouvislými texty, např. se seznamy, grafy, mapami, formuláři) a literární gramotnost (*prose literacy* – schopnost pracovat se souvislými texty) (OECD 2000).

Následoval výzkum *ALL (Adult Literacy and Lifeskills Survey)*, který se uskutečnil v letech 2003–2008¹⁵. V tomto výzkumu byla zjišťována numerická gramotnost, dále literární gramotnost, která kombinovala dokumentovou a literární gramotnost z výzkumu IALS, a kompetence k řešení problémů (Binkley a kol. 1999).

Oba výzkumy byly zaměřeny na dospělou populaci ve věku 16-64 let a probíhaly v domácnostech. Sestávaly z tištěných testů a dotazníků. Testy obsahovaly úlohy s výběrem odpovědí i úlohy, ve kterých měli respondenti doplnit vlastní odpověď, případně ji vyznačit v textu. Odpovědi byly následně vyhodnocovány školenými hodnotiteli. Výzkumu IALS se zúčastnilo 22 zemí, výzkumu ALL se zúčastnilo 10 zemí.

V roce 2009 zahájila OECD další výzkum vědomostí a dovedností dospělých *PIAAC (Programme for the International Assessment of Adult Competencies)*. U respondentů byly zjišťovány matematické dovednosti (*numeracy*), čtenářské dovednosti (*literacy*) a dovednosti řešit problémy v prostředí informačních technologií (*problem solving skills in technology rich environments*) (OECD 2012, OECD 2013c)¹⁶.

Výzkum PIAAC přinesl do mezinárodních srovnávacích výzkumů mnoho nových prvků. Asi nejdůležitější inovací byla administrace testů a dotazníků prostřednictvím počítačů. Respondentům, kteří nebyli schopni absolvovat testování na počítači, byly administrovány tištěné

¹⁴ Podrobnější informace o studii proveditelnosti jsou k dispozici na <http://www.oecd.org/edu/skills-beyond-school/testingstudentanduniversityperformancegloballyoecdshelo.htm>.

¹⁵ Více informací např. na: <http://nces.ed.gov/surveys/all/>.

¹⁶ Informace o výzkumu PIAAC, výsledky, popis metodologických postupů, originální výzkumné nástroje a mezinárodní databáze jsou k dispozici na <http://www.oecd.org/site/piaac/>. Informace o české realizaci včetně výsledků, českých verzí nástrojů a českých datových souborů jsou k dispozici na www.piaac.cz.

testy, které byly ovšem k dispozici pouze pro numerickou a čtenářskou gramotnost. Výhodou počítačové administrace je možnost sledovat nejen výsledek, ale celý postup řešení úlohy, což je důležité zejména pro hodnocení v oblasti řešení problémů.

Elektronické testy numerické a čtenářské gramotnosti byly koncipovány jako adaptivní, což umožňuje přesnější měření, protože obtížnost řešených úloh je možno lépe přizpůsobit schopnostem respondenta. Pro adaptivní testování je ovšem nezbytné, aby byly úlohy okamžitě vyhodnocovány a na základě výkonu v jednom bloku byl respondentům přidělen další blok s úlohami odpovídající obtížnosti. To vyžadovalo koncipovat úlohy tak, aby bylo možno je vyhodnotit počítačově, tedy eliminovat článek hodnotitelů. Zároveň však nebylo možno řešit tento problém pouze zařazováním úloh s výběrem odpovědi, neboť tyto úlohy umožňují testovat jen omezenou škálu vědomostí a dovedností. Proto byly v rámci výzkumu ověřovány různé formáty odpovědí. V některých úlohách tak měli respondenti za úkol vyznačit odpověď v textu, v jiných doplnit číselnou odpověď, jinde měli za úkol učinit nějaké rozhodnutí a pro toto rozhodnutí zvolit zdůvodnění¹⁷.

Dalším úkolem bylo koncipovat úlohy tak, aby měřily stejné dovednosti v tištěné i v počítačové podobě, tedy aby bylo možno srovnávat vědomosti a dovednosti osob, které se zúčastnily počítačového testování a které pracovaly s tištěnými sešity, na stejné škále.

Výzkumu PIAAC se zúčastnilo v prvním kole 24 zemí. Stejně jako ve výzkumu PISA vyjádřila řada zemí zájem výzkum zrealizovat v dalších kolech. Výzkum PIAAC bude stejně jako výzkum PISA realizován periodicky. V případě výzkumu PIAAC by měl být cyklus desetiletý, tedy další šetření by mělo proběhnout v roce 2021.

1.6 Realizace mezinárodních výzkumů výsledků vzdělávání žáků

1.6.1 Jednotlivé kroky realizace mezinárodních výzkumů

Tento oddíl přehledně informuje o všech krocích realizace mezinárodních výzkumů výsledků vzdělávání. Vybranými metodologickými aspekty, které jsou důležité pro práci s daty, se podrobně zabývá kapitola 3.

Východiska

Mezinárodní výzkumy výsledků vzdělávání jsou definovány jako výzkumy vědomostí, dovedností, postojů a chování ve vybrané oblasti vzdělávání v určité definované populaci. Zahrnují a) výběr vědomostí, dovedností, postojů a chování, které budou ve výzkumu sledovány – ty jsou definovány v koncepčních rámcích, b) výběr relativně velkého množství úloh a dotazníkových položek, které tyto vědomosti, dovednosti, postoje a chování reprezentují, c) relativně početné výběry zástupců populace, jejíž vědomosti, dovednosti, postoje a chování zjišťujeme.

Výsledky jsou prezentovány za určité skupiny. Nikdy netestují všechny osoby z cílové populace, tj. například všechny žáky 8. ročníků v daném vzdělávacím systému, ale pouze jejich výběr, z kterého je usuzováno na vědomosti a dovednosti celé skupiny. Výsledky jsou tedy nutné

¹⁷ Tyto postupy kladou vysoké nároky na technickou realizaci. Správné odpovědi musí být vyhodnocovány pro každý z jazyků testu samostatně, neboť rozmístění textu na obrazovce se liší. Stejně tak bylo například nutno upravit správné odpovědi v oblasti matematické gramotnosti národním zvyklostem.

zatíženy výběrovou chybou. Měření vědomostí a dovedností je dále nutně zatíženo chybou měření. Použité metody a postupy jsou odvozeny od tradice výběrových šetření a používají osvědčené metody efektivního sběru dat, které se snaží výše uvedené chyby minimalizovat, ale nikdy je nemohou zcela eliminovat. Poznatky z výzkumů tedy musí být interpretovány v kontextu neurčitosti, která tato šetření provází.

Mezi chyby spojené s realizací, které by měly být vzaty v úvahu při interpretaci výsledků, patří například inkonsistence při překladu nástrojů, způsobu administrace výzkumných nástrojů, vyhodnocování žákovských odpovědí, pořizování a zpracování dat a podobně. Tyto zdroje chyb jsou zpravidla minimalizovány postupy, které jsou podrobně popsány níže. Cílem těchto postupů je standardizovat podmínky napříč všemi zúčastněnými zeměmi.

Nad rámec nekonzistencí ve výše uvedených postupech musíme počítat s dalšími dvěma zdroji chyb, které jsou kontrolovány nikoli standardizací postupů, ale prostřednictvím statistických metod. Tyto zdroje chyb se vztahují k výběru respondentů a výběru úloh, které tito respondenti řeší, neboť ve výběrových šetřeních vědomostí a dovedností je zpravidla používán maticový design, při němž řeší různé skupiny respondentů různé soubory úloh s cílem pokrýt širší rozsah hodnocených vědomostí a dovedností. Těmto aspektům šetření se podrobně věnujeme v kapitole 3.

Při realizaci výzkumů je kladen důraz na následující 4 faktory (OECD 2014b):

- Konzistentnost – data musejí být sbírána stejným způsobem ve všech zemích a ve všech cyklech, aby bylo možno provádět srovnání. Výsledky musejí odrážet rozdíly ve vědomostech a dovednostech, nikoli rozdíly související s jiným způsobem provedení výzkumu.
- Přesnost – procedury musejí být koncipovány takovým způsobem, aby se minimalizovaly systematické i náhodné chyby.
- Zobecnitelnost – data jsou získávána od vybraných jedinců ve specifické situaci v určitém čase. Respondenti i výzkumné nástroje musejí být voleni tak, aby výsledky byly zobecnitelné a nevztahovaly se pouze k této specifické skupině a situaci. Tomu musí odpovídat konstrukce výběru respondentů i výzkumných nástrojů.
- Včasnost – data musejí být shromážděna ve všech zúčastněných zemích souběžně a zpracována tak, aby byly výsledky k dispozici co nejdříve (stejně to však bývá nejdříve 1,5 roku po ukončení sběru).

V této kapitole jsou obecně popsány hlavní kroky realizace největších mezinárodních výzkumů TIMSS, PIRLS (Olson a kol. 2008, Martin; Mullis 2015) a PISA (OECD 2014a, 2014b). Ostatní výzkumy v žákovské populaci se řídí velmi podobnými principy.

Výzkumy v dospělé populaci mají velmi podobnou organizaci a používají stejných metod škálování a zpracování výsledků jako výzkumy v populaci žákovské. Zcela zásadně se nicméně liší výběr respondentů, kteří jsou vybíráni buď prostým náhodným výběrem z registrů obyvatelstva, nebo víceetapovým výběrem (nejčastěji prostřednictvím volebních okrsků). Šetření pak probíhá v domácnostech respondentů, tedy vyžaduje zapojení zkušených administrátorů, kteří mají zkušenosti s dotazováním v domácnostech. Vzhledem k tomu, že tato publikace je zaměřena na výzkumy v žákovské populaci, věnuje se specifikům šetření v dospělé populaci pouze pro dokreslení – okrajově a nesystematicky.

Realizace výzkumů – rozdělení úkolů a zodpovědností

Jak již bylo uvedeno dříve, výzkumy TIMSS a PIRLS organizuje Mezinárodní asociace pro hod-

nocení výsledků vzdělávání, která tradičně sdružovala výzkumné instituce. Má sekretariát se sídlem v Amsterdamu v Nizozemsku, který zajišťuje chod organizace a jednotlivých výzkumů. Součástí organizace je i datové centrum se sídlem v Hamburku (*IEA DPC – IEA Data Processing Center*), které připravuje software na pořizování dat, shromažďuje a čistí data a připravuje finální datové soubory. IEA DPC zároveň organizuje řadu vzdělávacích aktivit pro realizátory výzkumu a výzkumníky v zúčastněných zemích. Za reprezentaci organizace IEA a koncepční práci nesou zodpovědnost výkonný ředitel a předseda, kteří jsou zpravidla představiteli výzkumných center a univerzit¹⁸.

Na plánování dalšího směřování organizace a výzkumných aktivit se podílejí i jednotlivé členské země prostřednictvím svých zástupců. Každá členská země je v IEA zastoupena jednou výzkumnou organizací, zástupci těchto organizací se každoročně setkávají na valném shromáždění (*IEA General Assembly*). Realizaci výzkumů TIMSS a PIRLS zajišťuje *TIMSS & PIRLS Study Center* na Boston College v USA ve spolupráci se zmiňovaným datovým centrem v Hamburku.

V ČR byly výzkumy IEA realizovány ve Výzkumném ústavu pedagogickém (do roku 1995), poté v Ústavu pro informace ve vzdělávání (v letech 1995-2011) a v současné době je jejich realizací pověřena Česká školní inspekce. Příslušné organizace vždy také zastupovaly ČR v IEA.

Zadavatelem a hlavním supervizorem výzkumu PISA je sekretariát OECD. Organizace pro hospodářskou spolupráci a rozvoj sdružuje ministerstva členských zemí OECD¹⁹. Zástupci ministerstev školství reprezentují členské země v tzv. *Education Policy Committee (EDPC)*, který rozhoduje o tom, jaké aktivity na poli vzdělávání budou prioritně realizovány. Důležitým orgánem OECD v oblasti vzdělávání je rovněž *Centrum pro pedagogický výzkum a inovace (Centre for Educational Research and Innovation – CERI)*²⁰, v němž jsou členské státy zastoupeny odborníky. Centrum provádí vývojovou práci, tedy reflektuje aktuální poznatky týkající se vzdělávání a rozvoje vzdělávacích systémů, předjímá budoucí potřeby a navrhuje výzkumné a koncepční studie reagující na aktuální vývoj a poznatky²¹.

O cílech a podobě jednotlivých šetření rozhodují zástupci ministerstev školství prostřednictvím řídicího výboru (*Board of Participating Countries BPC*), který je zřizován pro každou výzkumnou aktivitu. Výstupy aktivit jsou koncipovány tak, aby z nich čerpali primárně tvůrci vzdělávací politiky. To například znamená, že interpretace jsou v případě výzkumu PISA daleko směřlejší a jednoznačnější než v případě výzkumů IEA a data jsou daleko více vytěžována.

Na realizátora výzkumu PISA jsou v každém kole vypisována výběrová řízení. V prvních kolech byl výzkum realizován Australskou radou pro vzdělávací výzkum (*ACER – the Australian Council for Educational Research*) v konsorciu s *Netherlands National Institute for Educational Measurement (CITO)*, *the National Institute for Educational Research in Japan (NIER)* a význam-

¹⁸ V současné době je předsedkyní IEA Anne-Berit Kavli z Norského direktorátu pro vzdělávání (Norwegian Directorate for Education and Training) a výkonným ředitelem je Dirk Hastedt, který působil jako ředitel IEA DPC.

¹⁹ V roce 2015 měla OECD 32 členských zemí.

²⁰ Více informací o činnosti řídicího výboru CERI lze získat na www.oecd.org/edu/ceri/.

²¹ OECD např. realizuje v členských zemích tzv. *Policy reviews*, v jejichž rámci se shromažďují informace o zvoleném aspektu vzdělávacího systému v zúčastněných zemích (každá země vypracuje studii podle podrobné osnovy o aktuální situaci ve sledované oblasti, vývoji a problémech, kterým čelí). V zemích, které o to projeví zájem, pak proběhne hloubkové expertní posouzení komisí sestavenou ze 2 expertů OECD a dvou expertů z jiné země. Na jeho základě je vypracována národní zpráva s podrobnými doporučeními. Ze získaných informací pak vznikne zpráva popisující situaci ve všech zúčastněných zemích, formulující obecná doporučení. V posledních letech takto bylo například v ČR analyzováno vysokoškolské vzdělávání, střední odborné vzdělávání, hodnocení ve vzdělávacím systému, spravedlivost v povinném vzdělávání. CERI na základě podrobné znalosti problematiky doporučuje oblasti, které by měly být těmto detailním analýzám podrobeny.

nou americkou organizací se sídlem v Princetonu, která se specializuje na měření výsledků vzdělávání *Education Testing Service (ETS)*. V pozdějších letech dbala OECD o to, aby byl zvyšován počet organizací, které budou členy realizačního konsorcia tak, aby reprezentovaly co nejširší spektrum znalostí, zkušeností a pohledů na sledovanou problematiku z co nejširšího spektra členských zemí (zapojily se například výzkumné instituce z Německa). Postupně došlo i ke změně vedení konsorcia, které je aktuálně vedeno americkou organizací ETS (*Education Testing Service*).

V jednotlivých zemích je ve všech zmiňovaných výzkumech realizace zajišťována národními projektovými týmy vedenými národními projektovými manažery, kteří se pravidelně setkávají na mezinárodních instruktážních setkáních a výzkumy ve svých zemích realizují podle jednotných pečlivě promyšlených a vypracovaných pokynů. V některých zemích pracují národní koordinační týmy při univerzitách, jinde při národních ústavech pro pedagogický výzkum nebo při agenturách, které jsou zodpovědné za organizaci národního testování. Některé země vypisují na realizátory jednotlivých kol výzkumů výběrová řízení, tedy jednotlivá kola výzkumu realizují různá pracoviště, jinde zůstává realizátor stále stejný. Projektové týmy musí zajistit rozmanité odbornosti. Jejich součástí musí být odborník zodpovědný za překlad výzkumných materiálů, odborník schopný práce s daty zodpovědný za přípravu opory pro výběr, čištění dat a zpracování finálního datového souboru, dále odborník schopný vyškolit administrátory, získat ke spolupráci školy a supervidovat administraci, a odborník, který rozumí testované oblasti a je schopen vyškolit hodnotitele a koordinovat vyhodnocování úloh.

Príslušní odborníci se zúčastňují mezinárodních školení ve své oblasti a někteří z nich následně školí ve svých zemích pracovníky, kteří provádějí specifické úkony, například administraci testu nebo vyhodnocování otevřených odpovědí. Národní školení probíhají na základě přeložených a adaptovaných tréninkových manuálů a jejich průběh je pečlivě dokumentován.

Všechny práce musí být odvedeny ve vysoké kvalitě. Realizace výzkumů je v jednotlivých zemích pečlivě dokumentována do předem připravených formulářů a monitorována externími pracovníky mezinárodního centra. Výsledkům je přikládána velká váha, a z toho důvodu musí být mezinárodní centrum schopno doložit, že data byla sebrána řádně a výsledky věrně vypovídají o vědomostech a dovednostech žáků v příslušné věkové kategorii v dané zemi a o faktorech, které tyto vědomosti a dovednosti ovlivňují.

V ČR je výzkum PISA realizován stejným pracovním týmem jako výzkumy IEA, tedy aktuálně je za jeho provedení v ČR zodpovědná Česká školní inspekce. Výběrová řízení na realizaci výzkumů zde nikdy vypisována nebyla.

Příprava konceptuálních rámců a testových nástrojů

V první etapě přípravy mezinárodního výzkumu výsledků vzdělávání jsou vymezeny hodnocené oblasti prostřednictvím konceptuálních rámců, které jsou představeny v následující kapitole. Součástí konceptuálních rámců je i rozvaha týkající se složení testů. Důležitou složkou výzkumů jsou dotazníky pro žáky, učitele, ředitele škol a případně rodiče žáků. Zjišťují doplňující informace týkající se žáků (mimoškolní činnost, domácí zázemí, plány do budoucna) a výuky v dané oblasti vzdělávání, včetně podmínek a obtíží, které ji komplikují (jak ji vnímá žák, učitel, ředitel, rodič). Konceptuální rámce dotazníkových šetření budou představeny v kapitole 3.1.2. Na základě konceptuálních rámců jsou pak vytvořeny testové úlohy a dotazníkové položky a sestaveny testy a dotazníky pro pilotní šetření.

Již bylo uvedeno, že pro testování je používán takzvaný maticový design, při kterém různí žáci

řeší různé úlohy; díky sofistikovaným metodám škálování je však možno výsledky žáků mezi sebou porovnávat, tedy prezentovat je na stejné škále. Tento systém umožňuje ohodnotit širší rozsah vědomostí a dovedností, neboť celkový test, který výrazně přesahuje časové možnosti jednoho žáka, je rozložen mezi větší množství žáků. Tuto metodu je možno použít zejména proto, že v mezinárodních výzkumech nejde o porovnávání jednotlivců, ale skupin, v našem případě jednotlivých zemí, respektive vzdělávacích systémů, případně nějakých větších územních celků uvnitř těchto zemí (např. krajů, provincií a podobně).

Testy používané ve výzkumech IEA jsou tvořené z velké části úlohami, ve kterých žáci vybírají z několika nabízených odpovědí jedinou správnou. Dále jsou zastoupeny úlohy, ve kterých žáci odpovídají jednoslovně či několika slovy a kde poskytují obsáhlejší odpověď nebo úplné řešení dané úlohy. Vlastní odpovědi žáků jsou následně vyhodnocovány hodnotiteli podle velmi podrobných návodů.

Testy výzkumu PISA jsou na první pohled podobné testům z výzkumů IEA. Odlišnost (která je patrná zejména v matematických a přírodovědných úlohách) spočívá v tom, že ve výzkumu PISA jsou více zastoupeny úlohy, ve kterých žáci odpovídají vlastními slovy. Úlohy jsou uvozeny nějakým autentickým textem, grafem nebo obrázkem, za kterým následuje sada otázek (výskyt izolovaných úloh je zde daleko menší než ve výzkumu TIMSS). Žáci tak pracují delší dobu s jedním tématem a mohou se tak na ně plně soustředit. Z hlediska měření je však tento přístup nevýhodný v tom, že na sobě jednotlivé úlohy nejsou úplně nezávislé.

Pro účely elektronické administrace byly typy odpovědí modifikovány tak, že žáci označují správné odpovědi na obrazovce (buď kliknutím na určité místo na obrazovce nebo vyznačením nějaké části obrazovky), nebo doplňují číselné nebo jednoslovné odpovědi. V některých případech také píší otevřené odpovědi, které jsou následně hodnoceny hodnotiteli stejně jako u tištěných testů.

Jak bylo uvedeno výše, obsah testu není určován probíraným učivem, ale předem vypracovanými rámcovými koncepcemi, na jejichž tvorbě se podílejí přední světoví odborníci na jednotlivé testované oblasti. Koncepce oblastí se průběžně inovují spolu s tím, jak se vyvíjí stav poznání a pohled na jednotlivé sledované oblasti. Přitom je kladen větší důraz na provázanost vědomostí a dovedností se situacemi, s nimiž se lze setkat v běžném životě. Koncepce zohledňují zejména důležitost důležitých vědomostí a dovedností pro úspěšné fungování v životě osobním a společenském a na trhu práce.

Velká péče je v mezinárodních výzkumech věnována tomu, aby úlohy byly voleny tak, aby žáky z žádných zemí nevýhodňovaly, respektive neznevýhodňovaly. Proto jsou všechny zúčastněné země vybízeny k tomu, aby dodávaly vlastní úlohy, aby skladba úloh byla typově a kontextově dostatečně pestrá a neodpovídala zvyklostem pouze některých vzdělávacích systémů. K tvorbě úloh jsou organizovány mezinárodní dílny. Dále jsou úlohy posuzovány zúčastněnými zeměmi z hlediska možného znevýhodnění. Pokud by byl kontext nějaké úlohy pro žáky některé země natolik cizí, že by jim mohl bránit ve zdárném vyřešení úlohy, je tato úloha z testu vypuštěna.

Překlady a adaptace výzkumných nástrojů

Důležitým aspektem přípravy výzkumných nástrojů je jejich překlad do jazyků zúčastněných zemí. Pro překlad jsou vypracovávány podrobné pokyny, které zajišťují, aby výrazy důležité pro řešení úlohy byly překládány jednotně a aby byl zachován smysl sdělení. Při překladu je doporučováno, aby byly pořízeny dva překlady od dvou nezávislých překladatelů, a ty byly potom sjednoceny další nezávislou osobou. Ve výzkumu PISA jsou předlohy pro překlad dokon-

ce vyhotovovány v angličtině a ve francouzštině, tedy jeden překlad je z angličtiny a druhý z francouzštiny. Výsledné národní překlady jsou verifikovány nezávislým překladatelem z mezinárodního koordinačního centra. V počátcích výzkumů se kontrola prováděla zpětným překladem do angličtiny. Součástí překladu je i adaptace reálií (jmen, zápisů čísel, cen zboží apod.) podle národních zvyklostí. Řada zemí musí překládat testy do více jazyků tak, aby mohly být administrovány příslušníkům jazykových menšin. Všechny adaptace jsou zaznamenávány a schvalovány mezinárodním centrem.

Při počítačové administraci komplikuje situaci různá délka jednotlivých jazyků, neboť je důležité, aby se na obrazovku ve všech jazycích vešly vždy všechny potřebné informace. Komplikací je také skutečnost, že některé jazyky (např. hebrejšтина) se píšou zprava doleva nikoli zleva doprava, tedy rozložení textu a odpovědí na obrazovce se výrazně liší. Při automatickém vyhodnocování žákovských odpovědí je třeba pro každou zemi modifikovat i formát správných odpovědí (např. zápis čísla – desetinná tečka, čárka, oddělování tisíců, zápis cen apod.). Mají-li žáci na obrazovce vyznačit nějakou textovou pasáž, je důležité, aby byla v každé národní mutaci přesně vymezena, neboť její délka i umístění na obrazovce jsou zcela specifické.

V rámci pilotního ověřování jsou sledovány parametry úloh v jednotlivých zemích, které ukazují, zda úlohy fungují podobným způsobem. Někdy se ukáže, že tomu tak není právě kvůli posunu v překladu, který nebyl rozpoznán. Tedy pilotáž slouží i jako jakási finální kontrola překladu.

Důležité je zachování znění testových úloh a dotazníkových položek, které slouží ke srovnávání, i pokynů k jejich administraci tak, aby bylo možno jejich srovnávání v jednotlivých kolech šetření. Země proto nesmějí provádět změny v žádných dokumentech, aniž by tyto změny byly schváleny mezinárodním centrem.

V případě tištěných výzkumných nástrojů jsou ukázky všech testových sešitů a dotazníků zasílány do mezinárodního centra ke kontrole. Je důležité zajistit, aby grafická podoba nástrojů byla ve všech zemích identická a aby jejich tisková kvalita dodržovala předepsaný standard.

Výběr respondentů

Ve výzkumech IEA probíhá výběr respondentů ve všech případech dvoustupňově: nejprve je náhodně vybrán určitý počet škol z databáze všech škol, které jsou navštěvovány žáky v daném ročníku (zpravidla 4. a 8. ročník²²). Výběr je realizován jako systematický podle velikosti cílové populace na jednotlivých školách²³, v rámci těchto škol pak je náhodně vybrána jedna nebo dvě třídy v příslušném ročníku²⁴. Tímto postupem je zajištěno, že všichni žáci příslušného ročníku v dané zemi mají stejnou pravděpodobnost výběru (větší školy mají vyšší pravděpodobnost, že budou vybrány, jednotliví žáci však mají vyšší pravděpodobnost výběru v malých školách). V rámci výběru jsou pro každou školu vybírány 2 školy náhradní, které jsou osločovány v případě, že původně vybraná škola účast ve výzkumu odmítne. V rámci pravidel může každá země vyloučit maximálně 5 % cílové populace, která je obtížně testovatelná – některé země vylučují například školy v obtížně dostupných oblastech, jiné země, jako například ČR, vylučují žáky ve zvláštních (u nás nyní základních praktických) školách.

²² Ročníky jsou voleny podle věku: jedná se o ročníky, které navštěvuje většina desetiletých a čtrnáctiletých žáků, což je ve většině vzdělávacích systémů 4. a 8. ročník.

²³ Jako míra velikosti (*measure of size*) slouží počet žáků v cílovém ročníku.

²⁴ Dvě třídy jsou vybírány tehdy, je-li cílem kromě rozdílů mezi školami postihnout rovněž rozdíly mezi žáky jednotlivých tříd v dané škole. To je zajímavé zejména v situaci, kdy se setkáváme s existencí výběrových tříd.

V rámci výzkumu PISA je cílová skupina definována nikoli prostřednictvím ročníku školní docházky, ale prostřednictvím věku. Zde je cílem, aby výsledky nebyly ovlivněny případnými rozdíly ve věku respondentů, kteří navštěvují jednotlivé ročníky školní docházky. Cílovou skupinu tvoří všichni žáci, kteří v kalendářním roce testování dosáhnou 16 let (s drobnými modifikacemi podle období testování). To znamená, že do ní patří nejen žáci z různých ročníků, ale dokonce různých stupňů vzdělání: v ČR se zhruba polovina testované populace nachází v posledních ročnících povinného vzdělávání (tj. v 9. ročníku a v malé míře rovněž v 7. a 8. ročníku) a zhruba polovina v prvním ročníku střední školy (10. ročník školní docházky). I zde probíhá výběr dvoustupňově. Z databáze škol, která obsahuje všechny školy, které jsou navštěvovány žáky narozenými v daném kalendářním roce (tj. v ČR základní školy a všechny typy škol středních), je vybírán náhodně příslušný počet škol (minimum 150 škol v každé zemi) a v každé škole náhodně 30 žáků (žáci se logicky nacházejí v různých třídách a ročnících).

Ve výzkumu PISA je rovněž možno vyloučit školy v nedostupných oblastech nebo školy, které mají malý počet potenciálních respondentů. Výzkum PISA však neumožňuje na rozdíl od výzkumů IEA plošně vyloučit žáky, kteří se vzdělávají mimo hlavní vzdělávací proud v praktických školách. Pro tyto žáky jsou k dispozici jednodušší testy, které nicméně umožňují vyhodnotit vědomosti a dovednosti těchto žáků na stejné škále jako vědomosti a dovednosti ostatních žáků, kteří se testování zúčastnili. V opoře pro výběr jsou tedy kromě běžných základních škol zařazeny i základní školy praktické.

Stejně jako ve výzkumech IEA, i ve výzkumu PISA dochází také k vylučování žáků z testování v jednotlivých školách – jedná se například o žáky, kteří neznají jazyk testu natolik, aby mohli být testováni, o žáky s postižením, které jim účast v testu neumožňuje, a podobně. Podíl vyloučené populace na úrovni systému i na úrovni škol nesmí překročit 5 %.

V mezinárodním organizačním týmu je vždy zastoupena organizace nebo specializovaný tým, který je zodpovědný za výběr respondentů v zúčastněných zemích. Stalo se běžnou praxí, že země zasílají seznam škol (oporu pro výběr) s požadavkem na náležitosti výběru (reprezentativita za jednotlivé typy škol, regiony a podobně) do mezinárodního centra, které pro ně provádí výběr škol. Paralelně probíhá zpravidla výběr i pro pilotní šetření, aby nedošlo k tomu, že školy budou zařazeny do pilotního vzorku i do hlavního šetření. Výběr žáků (ve výzkumu PISA), respektive tříd (ve výzkumech IEA) je pak prováděn již v národním centru na základě aktuálních seznamů tříd a žáků, které školy zasílají do národního centra v situaci, kdy svolí s účastí ve výzkumu. Podrobněji se výběru věnujeme v kapitole 3.2.1.

Administrace výzkumných nástrojů

Jak již bylo uvedeno výše, při realizaci mezinárodních výzkumů je věnována velká pozornost jejich srovnatelnosti. Ta je zajišťována nejen podílem jednotlivých zemí na tvorbě výzkumných nástrojů, pečlivou kontrolou překladu a adaptace nástrojů, ale též standardizací administrace. Administrace probíhá podle jednotného časového plánu (například ve výzkumu PISA ve všech zemích v průběhu předem stanovených 6 týdnů, vždy přesně 3 roky po začátku administrace v předchozím kole) a podle jednotného scénáře, který je opět překládán do národních jazyků z anglické předlohy. Již bylo uvedeno, že žáci vyplňují různé bloky testových úloh, aby bylo zajištěno pokrytí širšího rozsahu učiva, než je v možnostech individuálních žáků. Je zcela zásadní, přidělit testové sešity žákům správně podle systému, který je dopředu stanoven.

Průběh administrace je pečlivě zaznamenáván do připravených záznamových archů. Postup administrace a dokumentace je přesně popsán v manuálech pro školního koordinátora a pro

administrátora. Manuály jsou rovněž překládány z anglické předlohy. Administrátoři jsou také proškoleni k tomu, aby zajistili utajení testových úloh, neboť některé z těchto úloh jsou používány opakovaně (pro sledování trendů) a není žádoucí, aby byly ve školách běžně k dispozici a aby je učitelé používali ve výuce.

Administrátoři by dále měli garantovat anonymitu žáků tím, že přiřadí k jednotlivým žákům kódy (např. na základě třídní knihy), ale jejich jména nebudou nikde figurovat. Zároveň je důležité, aby administrátoři byli schopni zajistit, aby všechny výzkumné materiály od jednoho žáka (různé testové sešity, dotazník) byly označeny stejným kódem tak, aby bylo možno data následně propojit.

Důležité je také řádně zaznamenat, kteří žáci byli, respektive nebyli při testování přítomni, případně z jakých důvodů se neúčastnili. Tato informace je důležitá při výpočtu žákovských vah (viz oddíl 3.2.2). V případě, že chybí v nějaké škole více žáků například z důvodů nemoci, je možno provést náhradní testování, kdy administrátor přijde do školy v době, kdy bude většina chybějících žáků přítomna, a výpovědi těchto žáků doplní.

Administrátoři by neměli mít k testovaným žákům žádný osobní vztah, aby nebyli motivováni jim s řešením úloh pomáhat. V žádném případě by to tedy neměli být jejich učitelé, zejména ne učitelé testovaných předmětů. Aby byla zaručena objektivita administrátorů, jsou v ČR testy administrovány pracovníky České školní inspekce.

Průběh administrace je kontrolován kontrolory kvality, kteří jsou zaměstnaní mezinárodními centry a pečlivě vyškoleni. Mají za úkol navštívit školy, které jsou k tomu účely vybrány mezinárodním centrem, a pečlivě zdokumentovat, jak v nich administrace proběhla. Kontroloři zpravidla navštěvují 5-10 % zúčastněných škol.

Pilotní ověřování

Pilotní ověřování je součástí každého kola šetření. Probíhá zpravidla rok před hlavním sběrem dat. Primárním cílem pilotáže je ověřit funkčnost testových úloh a testu jako celku. Velikost pilotního výběru je tedy určována minimálním počtem respondentů, který je nezbytný pro spolehlivý výpočet parametrů jednotlivých úloh v každém národním jazyce. Pilotováno je zpravidla více úloh, než kolik jich nakonec vstupuje do hlavního šetření. Na základě pilotních výsledků jsou vybrány úlohy tak, aby kvalita výsledného testu byla co nejvyšší.

V rámci pilotáže je rovněž ověřován proces administrace a časový rozvrh; ty mohou být na základě připomínek z pilotáže modifikovány.

Pokud země umožňuje šetření ve více jazycích, provádí se pilotáž ve všech jazycích, které reprezentují testovací jazyky pro více než 5 % testované populace. Pro hlavní testovací jazyk (více než 50 % respondentů) musí v rámci pilotáže být každá testová úloha řešena minimálně 200 respondenty, pro vedlejší jazyky je minimum 100 respondentů na úlohu.

Oslovování škol a návratnost

V době, kdy se všude ve světě množí výzkumné aktivity a zároveň je zaváděno v mnoha zemích povinné testování, je čím dál tím obtížnější získat školy ke spolupráci ve výzkumu. Mezinárodní výzkumné centrum poskytuje národním koordinátorům náměty, jak školy ke spolupráci získávat. Zároveň poskytuje inspiraci týkající se školních zpráv, které mohou národní koordináční centra školám zaslat jako zpětnou vazbu k testování, a tím je k účasti motivovat.

Jak bylo uvedeno výše, v rámci výběru respondentů jsou vybírány školy, ve kterých by měl

výzkum proběhnout, plus dvě školy náhradní. Z toho, jaké školy se nakonec výzkumu zúčastní a kolik žáků bude v daných školách testováno, je vypočítána návratnost (*response rate*). Ve výzkumu PISA určuje hodnota návratnosti, zda bude daná země zařazena do řádného mezinárodního srovnání nebo zda bude umístěna „pod čarou“, tedy její výsledky budou prohlášeny za nespolehlivé z důvodů nedodržení mezinárodních standardů. Ve výzkumu PISA musí být podíl zúčastněných škol 85 % z původního výběru a podíl žáků 80 % ze zúčastněných škol. U zemí, jejichž návratnost se blíží hraniční hodnotě, je z mezinárodního koordinačního centra pečlivě posuzována shoda mezi strukturou výběru a skladbou skutečných respondentů (je například sledováno, jestli někteří respondenti v konečném vzorku systematicky nechybějí – například žáci z hlavního města nebo naopak z malých venkovských škol a podobně). V některých případech jsou pak výsledky uznány i v situaci, kdy je návratnost mírně nižší než stanovená minimální hodnota.

Vyhodnocování otevřených odpovědí

Důležitou povinností národních koordinačních center je získat kvalifikované hodnotitele žákovských odpovědí a vyškolit je tak, aby k posuzování odpovědí přistupovali jednotným předepsaným způsobem. Návod na hodnocení žákovských odpovědí je vytvářen s využitím autentických odpovědí žáků. Na základě přesného popisu hodnocené dovednosti a znaků, podle kterých poznáme úroveň jejího osvojení, je posuzována úroveň zvládnutí, ale často také typ odpovědi. Typy špatných odpovědí jsou například užitečné pro posouzení miskoncepcí nebo nejčastějších chyb, kterých se žáci dopouštějí. Typy správných odpovědí například ukazují, zda jsou žáci ve škole vedeni k řešení úloh jednotným způsobem nebo zda jsou zvyklí hledat vlastní řešení. Jednotlivé úrovně a typy odpovědí jsou doprovázeny řadou autentických příkladů žákovských odpovědí.

Národní hodnotitelé jsou školeni pracovníky národního centra, kteří absolvovali mezinárodní školení, podle tréninkového manuálu a s využitím připravených ukázek žákovských odpovědí. V rámci školení hodnotitelé společně hodnotí ukázkové sporné příklady odpovědí. Následně hodnotí úlohy v testových sešitech s tím, že sporné případy projednávají s přítomným supervizorem a ten s nimi seznamuje ostatní hodnotitele, aby při jejich řešení volili identický postup. Část žákovských odpovědí je hodnocena nezávisle dvěma hodnotiteli a je stanovena míra shody mezi nimi. V některých výzkumech vybraní hodnotitelé ještě vyhodnocují anglické žákovské odpovědi, aby se ukázalo, zda se celý národní hodnotitelský tým nějakým zásadním způsobem neodchýlil od mezinárodního standardu.

Pořizování, kompletace a čištění dat

Mezinárodní koordinační centrum zpravidla jednotlivá národní koordinační centra vybaví specializovaným softwarem, do kterého mají přenést všechna data. V případě tištěných materiálů jsou data přepisována tak, že jsou zpravidla vyhotovovány dva nezávislé přepisy s tím, že nesrovnalosti jsou rozhodovány třetí osobou. V případě elektronického sběru přepis odpadá, ale je třeba dopředit hodnocení žákovských odpovědí a některých dotazníkových položek. Například ve výzkumu PISA je zjišťováno povolání rodičů dvojicí otázek: Jaké má Tvá matka/Tvůj otec zaměstnání? Co v tom zaměstnání dělá? Na základě odpovědí je dle mezinárodní klasifikace *ISCO (International Standard Classification of Occupations)* kódováno zkušeným hodnotitelem povolání každého rodiče.

Když jsou data zkompletována, nastává jejich čištění podle předepsaného návodu, přičemž zjištěné nedostatky musejí být dohledány a doplněny nebo opraveny (chybějící data, logické neshody v datech). Následuje další čištění v mezinárodním datovém centru a vyjasňování dalších nesrovnalostí s národním datovým manažerem. Vše probíhá podle předepsaného harmonogramu. Ve výzkumu PISA jsou národní centra povinna zaslat data do mezinárodního centra do 12 týdnů od posledního testování.

Zveřejnění výsledků

Jednotlivé země nesmějí zveřejnit výsledky dříve, než dojde ke zveřejnění výsledků na mezinárodní úrovni. Národní koordinační centra zpravidla připravují k datu zveřejnění tzv. národní zprávu, která obsahuje hlavní výsledky výzkumu v dané zemi. Autonomie národních center při zveřejňování výsledků se v jednotlivých zemích výrazně liší. Ve většině zemí jsou výsledky zveřejňovány v kooperaci s ministerstvem školství. Podrobněji se zveřejňování a základnímu zpracování výsledků věnuje oddíl 1.7.

1.6.2 Rozdíly ve výzkumech IEA a OECD

Jak bylo uvedeno výše, výzkum TIMSS představoval v 90. letech nejvýznamnější mezinárodní výzkum výsledků vzdělávání a po roce 2000 převzal jeho úlohu výzkum PISA, který se zaměřoval na stejné oblasti vzdělávání a částečně také na podobnou věkovou skupinu. Od té doby probíhají oba výzkumy paralelně a řada odborníků se zabývá otázkou, do jaké míry se výsledky obou výzkumů liší a do jaké míry jsou rozdíly skutečně zapříčiněny odlišným pojetím testovaných oblastí (školní vzdělání vs. příprava pro život).

Rozdíly v koncepci obou výzkumů popsal krátce po zveřejnění výsledků prvního kola výzkumu PISA tehdejší ředitel Direktorátu OECD pro vzdělávání Barry McGaw takto: Zatímco TIMSS se snaží zjistit, *jaké učivo se žák učil a co si z něho osvojil*, PISA chce zjistit schopnost žáků použít osvojené vědomosti a dovednosti k řešení problémů reálného života, tedy odpovídá na otázku, *co je žák schopen udělat s tím, co se naučil* (Hutchison; Schagen 2006). Přestože se výzkumy TIMSS zaměřují na měření vědomostí a dovedností zahrnutých do školních osnov a výzkum PISA na aplikace vědomostí a dovedností do reálného života, jsou výsledky na úrovni vzdělávacích systémů silně korelovány. Například korelace mezi výsledky žáků ve výzkumu TIMSS 2003 v 8. ročníku a výzkumu PISA 2003 u patnáctiletých žáků činí pro 19 zemí, které se účastnily obou výzkumů, 0,87 v matematice a 0,97 v přírodovědných předmětech. Korelace pro výzkumy TIMSS 1999 a PISA 2000 činila 0,86 pro všech 21 zemí, které se zúčastnily obou šetření v matematice i přírodovědných předmětech. Vysoká korelace byla zjištěna i mezi výsledky výzkumu TIMSS a výsledky výzkumu vědomostí a dovedností dospělých IALS. Skutečnost, že testy s deklarovaným velmi rozdílným zaměřením jsou na úrovni jednotlivých zemí silně provázány, naznačuje, že navzdory záměru měří stejné dimenze dovedností (Hanushek a Woessman 2013).

Přestože se výzkumníci shodují v tom, že vědomosti a dovednosti měřené oběma výzkumy jsou velmi podobné, přicházejí zároveň s rozmanitými vysvětleními pro zjištěné odlišnosti. Například Hutchison a Schagen (2006) srovnávali výsledky ve výzkumech TIMSS 1999 a 2003 a PISA 2000 a 2003 a uskutečnili dotazníkové šetření mezi koordinátory výzkumů. Na základě získaných informací usoudili, že nejrozvinutější země dosahují lepších výsledků ve výzkumu PISA, zatímco země bývalého východního bloku dosahují lepších výsledků ve výzkumu TIMSS, což vysvětlovali odlišným přístupem k výuce sledovaných předmětů. Vysvětlení odlišností ve

výsledcích zároveň hledali v kurikulárních odlišnostech spojených s odlišnými věkovými kategoriemi žáků (žáci 8. ročníků vs. patnáctiletí), kdy dochází k situaci, že v některých vzdělávacích systémech není v 8. ročníku probráno učivo hodnocené v rámci výzkumu TIMSS. Na základě informací získaných od národních koordinátorů zamítli hypotézu, že rozdíly ve výsledcích výzkumů jsou způsobeny odlišnými metodologickými standardy pro výběr respondentů, pro administraci výzkumu či vyhodnocování žákovských odpovědí.

Wu (2009) analyzovala testové úlohy z matematiky z výzkumů TIMSS 2003 a PISA 2003. Porovnávala úspěšnost v úlohách výzkumu TIMSS, které odpovídají a které naopak neodpovídají konceptuálnímu rámci výzkumu PISA. Zjistila, že v úlohách, které neodpovídají konceptuálnímu rámci PISA a zahrnují více formální matematiky, dosahují vyšší úspěšnosti asijské a východoevropské země než země „západní“. Tento poznatek odpovídá vyšší úspěšnosti „západních zemí“ ve výzkumu PISA než ve výzkumu TIMSS, protože PISA obsahuje více úloh, které jsou zasazeny do reálného kontextu a neobsahují formální matematiku.

Někteří autoři (např. Burdett; Strman 2013) však docházejí k závěru, že koncepty obou výzkumů jsou natolik podobné, že je obtížné rozdíly ve výsledcích jakkoli vysvětlit. Jako vysvětlující faktor zpravidla nemůže sloužit ani vyšší čtenářská náročnost testů PISA, neboť se neuplatňuje v matematice a v přírodovědných předmětech konsistentním způsobem.

Vzhledem k tomu, že účast v mezinárodních výzkumech je pro vzdělávací systémy velmi nákladná a že přínos paralelní účasti ve výzkumu TIMSS a PISA v populaci žáků ve věku odpovídajícímu zhruba závěru povinného vzdělávání je nejasný, setkáváme se stále častěji s tím, že se vyspělé země účastní výzkumů IEA pouze ve 4. ročnících. Informace o výsledcích žáků na konci povinného vzdělávání čerpají pouze z výzkumu PISA a výzkum TIMSS volí pouze jako informaci o relativní úspěšnosti žáků na konci primární školy, přičemž mezi koncepty přírodovědných předmětů a matematiky ve výzkumu TIMSS a ve výzkumu PISA příliš nerozlišují.

1.7 Prezentace výsledků

Jak již bylo uvedeno, zatímco výzkumy IEA byly vždy koncipovány výzkumníky a sloužily primárně pedagogickému výzkumu, výzkum PISA si klade za cíl ovlivňovat vzdělávací politiku v jednotlivých zemích. Odlišné záměry vyplývají i z odlišné podstaty obou organizací a odlišného řízení výzkumů.

Výsledky výzkumu TIMSS jsou vždy zhruba rok a půl po ukončení sběru dat publikovány ve dvou publikacích: jedna je pro přírodovědné předměty, druhá pro matematiku. Výsledky jsou v nich prezentovány v členění podle kontextuálních proměnných, publikace však má víceméně popisný charakter. Autoři (pracovníci *TIMSS & PIRLS Study center* na Boston College) se zdráhají publikovat sofistikovanější analýzy, neboť se domnívají, že pro ně v získaných datech není dostatečná opora. Podobným způsobem jsou prezentovány výsledky výzkumu PIRLS.

I když Mezinárodní asociace pro hodnocení výsledků vzdělávání není příliš aktivní v produkci sofistikovanějších analytických zpráv založených na získaných datech, podporuje využití dat, která jsou zpřístupněna odborné veřejnosti ihned po prezentaci prvních výsledků, pořádáním pravidelné konference, na které účastníci prezentují analýzy dat získaných ve studiích IEA (*IEA Research Conference*). Tyto práce jsou zveřejňovány na webových stránkách organizace.

Hlubkové analýzy jsou realizovány často i na národní úrovni. Např. Wendt a kol. (2015) publikovali v roce 2015 obsáhlý a inspirativní soubor analýz německých dat získaných v rámci mezinárodních výzkumů TIMSS a PIRLS od roku 2001 do roku 2011. Soubor obsahoval například

analýzy, které se zaměřovaly na vývoj vztahu ke čtení a souvislost mezi výsledky a čtenářským vybavením škol, souvislost mezi formátem úloh a výsledky u různých skupin žáků, úspěšnost žáků v matematických úlohách orientovaných prakticky a teoreticky v závislosti na rodinném zázemí, souvislost výsledků s kvalifikací učitele a s jeho vírou ve vlastní schopnosti, srovnání výsledků celodenních (*Ganztagsschulen*) a běžných škol. Analýzy jsou velmi dobře teoreticky podloženy a výsledky jsou diskutovány také vzhledem k jejich praktické užitečnosti.

Data z výzkumů PISA jsou používána principiálně pro tři typy analýz a zpráv. Z každé vlny šetření je Direktorátem pro vzdělávání OECD publikována zpráva hlavních výsledků. Tato zpráva je zveřejňována zpravidla koncem roku, který následuje po sběru dat. Spolu s ní jsou zpravidla zveřejňovány národní výsledky v zúčastněných zemích. Na počátku výzkumu PISA byly hlavní výsledky prezentovány v jedné publikaci. Nyní sestává hlavní zpráva z několika analytických publikací, které se zaměřují specificky na rozmanité aspekty získaných informací. Například zpráva z výzkumu PISA 2012 obsahovala 6 tematických publikací:

- Co studenti znají a umějí udělat: výsledky žáků v matematice, čtení a přírodovědných předmětech.
- Dokonalost prostřednictvím spravedlivosti: poskytnout každému žákovi příležitost vyniknout.
- Připravení se učit: zaujetí žáků, tah na branku a přesvědčení o vlastních schopnostech.
- Co činí školy úspěšnými? Zdroje, zásady a postupy.
- Kreativní řešení problémů. Dovednost žáků poradit si s problémy běžného života.
- Žáci a peníze: finanční gramotnost pro 21. století.

Na hlavní zprávu navazují dílčí tematické analýzy. O tématech, na která budou zaměřeny, rozhoduje Rada zúčastněných zemí. Tematické zprávy se zaměřují na specifické aspekty zkoumané problematiky, například na rozdíly mezi chlapci a dívkami, na výsledky znevýhodněných žáků, na postoje k životnímu prostředí a podobně. Z dat získaných v rámci výzkumu PISA 2012 byly například publikovány analýzy *Students, Computers and Learning: Making the Connection* (Žáci, počítače a učení: hledání souvislostí) a *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence* (Abeceda genderových nerovností ve vzdělávání: postoje, chování, sebevědomí).

S cílem zvýšit využitelnost výstupů PISA pro tvůrce vzdělávacích politik začalo OECD v roce 2011 produkovat v měsíčních intervalech krátká sdělení (*policy notes*) založená na analýzách dat z výzkumu PISA pod názvem *PISA in Focus*²⁵. Do konce roku 2015 vzniklo celkem 55 zpráv, které se řadí do 5 tematických celků:

- výsledky žáků (např. Prospívá více vyučovacích hodin vzdělávacím výsledkům? Je důležité, jakou školu žák navštěvuje? Co zapříčiňuje genderové nerovnosti ve vzdělávání?)
- postoje žáků ke škole a k učení (např. Ovlivňují vztahy mezi žáky a učiteli spokojenost žáků ve škole?, Bojí se žáci matematiky?, Kteří žáci ve škole podvádějí?)
- žákovské výsledky a opatření vzdělávací politiky (např. Posilují domácí úkoly nerovnosti? Opakují znevýhodnění žáci častěji ročník? Dostává se předškolního vzdělávání těm, kteří ho nejvíce potřebují?)
- rodinné zázemí a vzdělávání (např. Mohou být odstraněny rozdíly ve výsledcích žáků imigrantů a žáků z majoritní populace? Mohou výukové strategie odstranit rozdíly ve výsledcích

²⁵ Všechna sdělení jsou k dispozici na: <http://www.oecd.org/pisa/pisaproducts/pisainfocus.htm>.

znevýhodněných a zvýhodněných žáků? Jak se vzdělávací systémy přizpůsobují zvýšenému počtu přistěhovalců?)

- prostředí školní třídy (např. Jsou žáci aktivnější ve školách, které nabízejí volnočasové aktivity? Zvyšuje se ve školách četnost kázeňských problémů? Prospívají žáci lépe ve školách, kde panuje řád a pořádek?)

Zprávy jsou rozesílány podle rozsáhlého distribučního seznamu tvůrcům vzdělávacích politik v členských zemích a osobám zapojeným nějakým způsobem ve výzkumu PISA.

Do doby zveřejnění hlavních výsledků jsou data i výsledky drženy v přísné tajnosti a k jejich utajení jsou přijímána rozmanitá opatření. Poté, co jsou zveřejněny hlavní výsledky, uvolňuje OECD mezinárodní databáze k dalším analýzám. Na webových stránkách výzkumu PISA jsou – stejně jako na stránkách výzkumů TIMSS a PIRLS – kromě datových souborů všech zúčastněných zemí vyvěšeny rovněž technická dokumentace a anglické verze dotazníků. Testy jsou uchovávány v tajnosti z důvodů dalšího použití, vždy je však uvolněno několik úloh, které slouží pro ilustraci hodnocených oblastí.

Jak již bylo uvedeno, indikátory vytvořené na základě analýz dat z výzkumu PISA vstupují pravidelně do ročenky OECD Education at a Glance, kde slouží k porovnávání výkonů zemí OECD a přidružených zemí a k hodnocení různých aspektů fungování jejich vzdělávacích systémů.

2. Vývoj v konceptualizaci hodnocených oblastí

Tato kapitola ukazuje, jak se vyvíjely koncepty hlavních měřených oblastí ve výzkumech obou organizátorů. Popisuje koncepční rámce jednotlivých oblastí na počátku cyklů měření a srovnává je s pojetím při poslední realizaci výzkumů. Konceptuální rámce velmi pěkně demonstují, jak jsou pro účely testování přesně popsány vzdělávací cíle, které jsou testem ověřovány. Na vývoji koncepčních rámců pak můžeme dobře pozorovat, jak se ve všech popisovaných výzkumech rozšiřuje a prohlubuje přemýšlení o měřených oblastech, tedy o tom, co by si měli mladí lidé odnášet ze školy do života v hlavních kurikulárních oblastech. Kromě konceptuálních rámců hlavních měřených oblastí představuje kapitola i koncepty oblastí, které jsou měřeny pouze občas či se objevily ve výzkumech nově.

2.1 Výzkumy Mezinárodní asociace pro hodnocení výsledků vzdělávání

2.1.1 Matematika a přírodovědné předměty ve výzkumu TIMSS

Tabulka 1 přehledně znázorňuje konceptuální rámec pro matematickou a přírodovědnou oblast výzkumu TIMSS 1995, který sloužil jako východisko pro všechny následující výzkumy. Obsahoval tři dimenze: 1. obsah - učivo, které úloha hodnotila, 2. výkon - typ kognitivní operace, která byla vyžadována pro vyřešení úlohy, 3. perspektivy - postoje k dané oblasti a zájem v ní dále pracovat. Úlohy byly koncipovány tak, aby ověřovaly zároveň obsahové a kognitivní aspekty, tedy aby rovnoměrně pokrývaly první dvě dimenze. Třetí, nekognitivní dimenze byla sledována prostřednictvím dotazníků.

Od roku 1995 došlo v konceptu matematických a přírodovědných vědomostí a dovedností měřených ve výzkumu TIMSS k určitým zpřesněním. Z koncepčního rámce zároveň zmizela třetí oblast *Perspektivy*; nahradila ji oblast *Postupy*, jež zahrnuje výukové aktivity, které žáci vykonávají. Hodnocení matematických vědomostí a dovedností bylo obohaceno o tzv. numerickou gramotnost (numeracy), která je pro účely výzkumu TIMSS považována za jakýsi předstupeň matematiky. Hodnocení numerické gramotnosti slouží jako alternativa pro země, ve kterých jsou matematické vědomosti a dovednosti u většiny žáků na takové úrovni, která znemožňuje kvalitní měření pomocí standardního matematického testu.

Tabulky 2 a 3 obsahují obsahové a kognitivní aspekty, které byly ve 4. ročníku hodnoceny v rámci výzkumu TIMSS v matematice a přírodovědných předmětech v roce 2015. Kognitivní oblast je členěna zcela identicky i pro žáky 8. ročníku, testovaný obsah se přirozeně liší (Mullis; Martin 2013).

Tabulka 1: Matematický a přírodovědný konceptuální rámec výzkumu TIMSS 1995²⁶

Matematika	Přírodovědné předměty
Obsah	
Čísla	Vědy o zemi
Měření	Vědy o životě
Geometrie	Fyzikální vědy
Úměrnost	Věda, technika a matematika
Funkce, vztahy a rovnice	Historie vědy a techniky
Reprezentace dat, pravděpodobnost a statistika	Životní prostředí
Elementární analýza	Přírodní vědy a další obory
Výkon	
Znalost	Porozumění
Používání rutinních postupů	Teoretizování, analyzování a řešení problémů
Zkoumání a řešení problémů	Používání nástrojů, rutinní postupy
Matematické zdůvodňování	Zkoumání světa přírody
Komunikace	Komunikace
Perspektivy	
Postoje	Postoje
Volba povolání	Volba povolání
Zapojení	Zapojení
Zájem	Zájem
Myšlenkové návyky	Bezpečnost
	Myšlenkové návyky

²⁶ http://timssandpirls.bc.edu/timss1995i/t95_study.html

Tabulka 2: Konceptuální rámec hodnocení v matematické oblasti, TIMSS 2015, 4. ročník

Obsahové oblasti	
Číslo	Celá čísla Zlomky a desetinná čísla Výrazy, jednoduché rovnice a vztahy
Geometrické tvary a měřítka	Body, přímky, úhly Dvourozměrné a třírozměrné útvary
Zobrazení dat	Čtení, interpretace a reprezentace
Kognitivní oblasti	
Znalost	Vybavit si Rozpoznat Klasifikovat/uspořádat Spočítat Získat informaci (např. z grafu apod.) Měřit
Aplikace	Určit Reprezentovat/modelovat Implementovat
Zdůvodnění	Analyzovat Integrovat/Syntetizovat Hodnotit Vyvodit závěry Zobecnit Odůvodnit

Tabulka 3: Konceptuální rámec hodnocení v přírodovědné oblasti, TIMSS 2015, 4. ročník

Obsahové oblasti	
Vědy o životě	Charakteristiky organismů a jejich životní procesy Životní cyklus, reprodukce, dědičnost Organismy, prostředí a jejich interakce Ekosystémy Lidské zdraví
Fyzikální vědy	Klasifikace látek, jejich vlastnosti a změny Formy energie a její přeměny Síly a pohyb
Vědy o zemi	Struktura země, její fyzikální charakteristiky a zdroje Procesy a historie Země ve Sluneční soustavě
Kognitivní oblasti	
Znalost	Vybavit si/rozpoznat Popsat Uvést příklady
Aplikace	Srovnat/kontrastovat/třídít Dát do souvislosti Použít modely Interpretovat informaci Vysvětlit
Zdůvodnění	Analyzovat Syntetizovat Formulovat otázku/hypotézu, předpovídat Navrhnout výzkum Hodnotit Vyvodit závěry Zobecnit Odůvodnit

2.1.2 Čtenářská gramotnost ve výzkumu PIRLS

Ve výzkumu *Reading Literacy Study* v roce 1991/1992 bylo zjišťováno, jak je porozuměno třem typům textů: vyprávění, výkladovým textům a dokumentům, které byly zakotveny ve čtyřech typech kontextů: domov, škola, společnost, práce.

Konceptuální rámec byl i zde významně zdokonalen. Zasazení textů do rozmanitých kontextů zůstalo zachováno. Tři typy textů byly nahrazeny dvěma typy účelů čtení, kdy práce s vyprávěním a dalšími literárními texty spadá do kategorie zážitků z četby, a práce s výkladovými

a dokumentárními texty patří do kategorie získávání a používání informací. K typům textů přibyla dimenze čtenářských procesů, kde je posuzováno, jak respondent s texty pracuje. Jednotlivé dimenze konceptuálního rámce PIRLS 2016 jsou uvedeny v tabulce 4. Výzkum v roce 2016 nabízí nově též možnost hodnotit uvedené aspekty čtenářské gramotnosti elektronicky (ePIRLS).

Tabulka 4: Konceptuální rámec výzkumu PIRLS 2016

Účely čtení	
	Zážitek z četby Získávání a používání informací
Procesy porozumění	
	Vyhledat a získat explicitně uvedené informace Činit přímé závěry Interpretovat a integrovat myšlenky a informace Hodnotit a kriticky posoudit obsahové a textové prvky

V posledních kolech bylo hodnocení dále obohaceno o dimenzi přístupů a postojů, která je posuzována na základě dat získaných z žákovských dotazníků. Stejně jako ve výzkumu TIMSS byly i ve výzkumu PIRLS administrovány také učitelské a ředitelské dotazníky a byly zjišťovány informace o vzdělávacím systému.

Jak již bylo uvedeno v kapitole 1, konceptuální rámce ze všech cyklů šetření TIMSS a PIRLS jsou k dispozici na webových stránkách realizátora výzkumu TIMSS and PIRLS Study Center v Boston College²⁷. Informace v českém jazyce jsou k dispozici na stránkách České školní inspekce. Na těchto stránkách jsou k dispozici rovněž české překlady některých konceptuálních rámců a ukázky úloh²⁸.

2.2 Konceptuální rámce oblastí hodnocených ve výzkumu PISA

Důležitým počínem výzkumu PISA bylo rozpracování konceptů matematické, přírodovědné a čtenářské gramotnosti do struktury použitelné k mezinárodním měřením a obrácení pozornosti tvůrců vzdělávacích politik od „kurikulárních vědomostí“ ke „kompetencím pro život“. V tomto smyslu se výzkum PISA stal pro řadu zemí (a pravděpodobně i pro Českou republiku) impulsem pro rozmanité kurikulární reformy. Některé systémy dokonce hledaly společenský konsensus na tom, co má být prioritou vzdělávání ve 21. století, identicky jako zmiňovaný projekt DeSeCo, tedy dotaz kladly odborníkům z rozmanitých profesí a klíčové kompetence definovaly jako průnik jejich preferencí (Eurydice 2002).

V prvních dvou úplných cyklech výzkumu PISA v letech 2000 – 2015 bylo s každou z hlavních testovaných oblastí spojeno další specifické hodnocení. Se čtenářskou gramotností bylo

²⁷ <http://timssandpirls.bc.edu/>

²⁸ [http://www.csicr.cz/Prave-menu/Mezinarodni-setreni/TIMSS/TIMSS-%28Trends-in-International-Mathematics-and-Sci-a](http://www.csicr.cz/Prave-menu/Mezinarodni-setreni/TIMSS/TIMSS-%28Trends-in-International-Mathematics-and-Sci-a%20http://www.csicr.cz/Prave-menu/Mezinarodni-setreni/PIRLS) <http://www.csicr.cz/Prave-menu/Mezinarodni-setreni/PIRLS>

spojeno zjišťování metakognitivních dovedností, s matematickou gramotností řešení problémů a s přírodovědnou gramotností to byly znalost environmentální problematiky a šetření postojů k přírodovědným a technickým disciplínám a k životnímu prostředí. Následující text pojednává postupně o všech sledovaných oblastech.

2.2.1 Čtenářská gramotnost ve výzkumu PISA²⁹

Čtenářská gramotnost byla hlavní testovanou oblastí v roce 2000. Její definice byla vytvořena mezinárodní expertní skupinou ustavenou za tím účelem z odborníků členských zemí. Při konceptualizaci čtenářské gramotnosti odborníci vycházeli z výzkumu IEA *Reading Literacy Study* a z výzkumu gramotnosti dospělých *International Adult Literacy Survey* z roku 1994. Dále z teorií čtení, které zdůrazňují jeho interaktivní povahu.

Čtenářská gramotnost ve výzkumu PISA 2000 byla definována jako *schopnost porozumět psanému textu, používat psaný text a přemýšlet o něm za účelem dosažení cílů jedince, rozvoje jeho vědomostí a potenciálu a za účelem jeho aktivní účasti ve společnosti*.

Konceptuální rámec obsahoval 3 dimenze:

- typy úloh, tedy obsah, se kterým práce probíhala;
- čtenářské postupy, které charakterizovaly typy čtenářských úkonů, které čtenář realizoval s daným obsahem;
- čtenářské situace, tedy kontext čtenářského úkolu.

Hlavní dimenze konceptuálního rámce jsou uvedeny v tabulce 5.

Konceptuální rámec čtenářské gramotnosti prošel hlavní revizí v roce 2009 na počátku druhého cyklu šetření, kdy se opět stala hlavní testovanou oblastí. Při této revizi byla zachována podstata konceptuálního rámce z roku 2000, neboť hlavním cílem výzkumu je získat a prezentovat výsledky v testovaných oblastech tak, aby bylo možno sledovat dlouhodobé trendy. Konceptuální rámec však byl zároveň vždy pojmán jako živý dokument, který se bude průběžně vyvíjet tak, aby zahrnul nové poznatky týkající se čtenářské teorie a praxe. Proto v něm můžeme pozorovat značný vývoj, který odráží změny, k nimž dochází ve světě, jenž nás obklopuje, a rostoucí porozumění podstatě čtenářských aktivit.

Nejdůležitějším aspektem při revizi čtenářského konceptuálního rámce bylo zařazení elektronického čtení, tedy čtení textů, které byly vytvořeny pro četbu na obrazovce počítače. Aktuální výzkumy ukazují, že pracovníci v širokém spektru povolání potřebují stále větší měrou využívat počítače a další informační technologie při výkonu své práce. Tento vývoj pozorujeme nejen na pracovištích, ale i v osobním, společenském a občanském životě, kde lidé vyhledávají informace a komunikují prostřednictvím mobilních telefonů, tabletů a dalších informačních technologií. Konceptuální rámec PISA tedy rozlišuje mezi texty „stálými“ (*fixed texts*), které jsou typicky dostupné v tištěné podobě, ale mohou být znázorňovány i elektronicky, a texty „dynamickými“ (*dynamic texts*), které plně využívají možností elektronických médií.

²⁹ Konceptuální rámce ze všech cyklů výzkumu PISA jsou k dispozici na mezinárodních stránkách výzkumu PISA <http://www.oecd.org/pisa/>. České verze konceptuálních rámců jsou k dispozici na <http://pisa2012.cz/?a=vystupy> nebo na <http://www.csicr.cz/Prave-menu/Mezinarodni-setreni/PISA>.

Tabulka 5: Konceptuální rámec čtenářské gramotnosti, PISA 2000

Typy textů Souvislé texty Nesouvislé texty	Vyprávění	
	Výklad	
	Popis	
	Argumentace/Polemika	
	Instrukce	
	Grafy, diagramy	
	Tabulky	
	Obrázky	
	Mapy	
	Formuláře	
	Reklamní texty	
Čtenářské úkoly	Obecné porozumění	porozumění textu jako celku, porozumění účelu, identifikace hlavní myšlenky
	Získávání informací	nalezení informace v textu
	Vytvoření interpretace	integrace informací z různých částí textu, vyvozování závěrů
	Posouzení obsahu textu	propojení a porovnání informací z textu s informacemi z jiných zdrojů nebo s dříve nabytými vědomostmi
	Posouzení formy textu	zhodnocení stavby a jazyka textu vzhledem k danému účelu
Čtenářský kontext (situace, ve které čtení probíhá, nebo jeho účel)	Osobní	
	Vzdělávací	
	Pracovní	
	Veřejné	

Mnoho čtenářských dovedností se v případě obou typů textů neliší, elektronické čtení však vyžaduje, aby čtenáři rozšířili svůj čtenářský záběr o nové dovednosti. Získávání informací na internetu vyžaduje rychlé čtení, skenování velkého množství materiálu a bezprostřední hodnocení jeho věrohodnosti. Kritické myšlení se tak stává stále důležitějším. Někteří odborníci se shodují v tom, že znevýhodnění osob, které nejsou schopny získávat informace prostřednictvím informačních technologií, není pouze otázkou zajištění přístupu k těmto technologiím, ale také zlepšení schopnosti všech občanů integrovat, hodnotit a sdělovat informace.

Další změnou oproti roku 2000 bylo důkladnější rozpracování konceptu metakognice a zapojení do čtenářských aktivit (*reading engagement*). Mnoho výzkumných studií z posledních let ukázalo, že motivace a zájem o čtenářské aktivity vykazují silnou souvislost s úrovní čtenářské gramotnosti. Zkoumání metakognice, která má dle výzkumů podobnou vazbu na čtenářské kompetence, bylo možné díky vývoji nových nástrojů na její sledování.

V roce 2009 probíhalo hodnocení čtenářské gramotnosti primárně prostřednictvím tištěných textů. Zúčastněné země k tomu mohly zvolit ještě alternativu hodnocení elektronických textů. Obě dovednosti byly však hodnoceny na samostatných škálách, neboť se ukázalo, že měří poněkud odlišné dovednosti.

Tabulka 6: Konceptuální rámec čtenářské gramotnosti, PISA 2015

Typy textů	Zobrazení	tištěné (stálé texty) elektronické (dynamické texty)
	Prostředí (týká se pouze dynamických textů)	určené pro autorské práce (čtenář nemůže do čteného nijak zasahovat) založené na příspěvcích (čtenář může přispět vlastním názorem/zprávou)
	Formát textu	souvislý nesouvislý smíšený soubor textů
	Typ textu	popis vyprávění výklad diskuze/polemika instrukce transakce (typicky emailová komunikace a SMS zprávy, jejichž prostřednictvím se domlouváme o každodenních aktivitách)
Aspekty čtení/ čtenářské úkoly	Vyhledávání a získávání informací Integrace informací z různých částí textu a vytváření interpretace Posuzování obsahu a formy textu a jejich vyhodnocování	
Čtenářský kontext (situace, ve které čtení probíhá, nebo jeho účel)	Osobní Veřejné Pracovní Vzdělávací	

V dosud posledním výzkumu v roce 2015 již byl primární způsob administrace testu elektronický – tištěné testy byly použity pouze zeměmi, které nebyly schopny zajistit elektronickou administraci. Čtenářská gramotnost však byla v tomto roce vedlejší doménou a měla tedy jen omezený prostor. Aby bylo možno sledovat trendy, bylo do výzkumu v roce 2015 zařazeno čtení fixních textů s tím, že administrace testu probíhala prostřednictvím počítačů. Dynamické texty

budou jako povinná součást hodnocení čtenářské gramotnosti administrovány v roce 2018, kdy bude čtenářská gramotnost opět hlavní testovanou oblastí.

Konceptuální rámec z roku 2015 odráží všechny modifikace, které byly provedeny v roce 2009, a zařazuje i některé nové prvky spojené s výlučně elektronickou administrací fixních textů.

V PISA 2015 byla tedy čtenářská gramotnost definována jako *porozumění psaným textům, jejich využívání, posuzování a angažování se v nich za účelem dosažení cílů jedince, rozšíření jeho znalostí, potenciálu a aktivní účasti ve společnosti*.

Hodnocení bylo (stejně jako v roce 2000) zaměřeno na 3 aspekty čtenářské gramotnosti s tím, že vymezení prvních dvou oblastí bylo oproti roku 2000 mírně modifikováno. Texty byly členěny sofistikovanějším způsobem, odrážejícím rovněž potřebu zachytit charakteristiky elektronických textů. Aspekty čtenářských úkonů byly naopak sloučeny do tří širších kategorií. Hlavní charakteristiky konceptuálního rámce čtenářské gramotnosti pro rok 2015 jsou uvedeny v tabulce 6.

2.2.2 Matematická gramotnost ve výzkumu PISA

Matematická gramotnost byla v roce 2000 definována jako *schopnost rozpoznat a pochopit matematiku, zabývat se jí a činit dobře podložené soudy o úloze matematiky v soukromém životě jednotlivce, v zaměstnání, ve společnosti přátel a příbuzných a v životě konstruktivního, zainteresovaného a přemýšlivého občana, a to jak v přítomnosti, tak v budoucnosti*.

Konceptuální rámec byl členěn analogicky jako v případě čtenářské gramotnosti. Figurovaly v něm postupy, matematický obsah a kontext. Konceptuální rámec matematické gramotnosti použitý v prvním cyklu šetření PISA je popsán v tabulce 7.

Tabulka 7: Konceptuální rámec matematické gramotnosti, PISA 2000

Matematické postupy	Reprodukce, definice, výpočty Propojení a integrace při řešení problémů Matematizace, matematické myšlení, zobecňování a pronikání do podstaty matematiky
Obsah	Změna a růst Prostor a tvar Náhodnost Kvantitativní uvažování Neurčitost Závislost a vzájemné vztahy
Situace	Osobní (každodenní) život Život ve škole, v práci a ve sportu Obec a společnost Vědecký kontext

V roce 2012 bylo zúčastněným zemím nabídnuto přídatné testování elektronické (CBAM – *computer based assessment in mathematics*), analogicky, jako tomu bylo v případě čtenářské gra-

motnosti v roce 2009. V roce 2015, kdy byla matematika vedlejší testovanou oblastí, nebylo možno plně navázat na zkušenosti získané v roce 2012, neboť by tím utrpělo sledování trendů. I v případě matematiky byl tedy výběr úloh omezen tím, že úlohy byly administrovány sice primárně na počítači, ale bylo nutno vytvořit odpovídající verzi pro tištěný test pro země, které elektronicky testovat nemohou.

Matematická gramotnost je ve výzkum PISA 2015 definována jako *schopnost jedince formulovat, používat a interpretovat matematiku v různých kontextech. Zahrnuje matematické myšlení, používání matematických pojmů, postupů, faktů a nástrojů k popisu, vysvětlování a předpovídání jevů. Pomáhá jedinci si uvědomit, jakou roli matematika hraje ve světě, a díky tomu správně usuzovat a rozhodovat se tak, jak to vyžaduje konstruktivní, angažované a reflektivní občanství.*

Hodnocení bylo zaměřeno na 3 aspekty matematické gramotnosti, které zajistily, aby výzkum odpověděl na 3 základní otázky:

1. Jaké procesy jedinci používají, když řeší autentické matematické úlohy, a jaké dovednosti očekáváme, že budou schopni demonstrovat, když se bude zlepšovat jejich matematická gramotnost?
2. Jaké znalosti můžeme očekávat od patnáctiletých žáků?
3. V jakých kontextech můžeme pozorovat a měřit matematickou gramotnost?

Tabulka 8: Konceptuální rámec matematické gramotnosti, PISA 2015

Matematické postupy a dovednosti	Matematické postupy	formulování situací matematiky, používání matematických pojmů, faktů, postupů a uvažování interpretování, aplikace a hodnocení matematických výsledků.
	Matematické dovednosti	komunikace matematizace reprezentace uvažování a argumentace navržení strategií řešení problému používání symbolického, formálního a technického jazyka a operací používání matematických nástrojů
Obsah	Změna a vztahy Prostor a tvar Kvantita Neurčitost a data	
Kontext	Osobní Profesní Společenský Vědecký	

Konceptuální rámec byl členěn analogicky konceptuálnímu rámci z roku 2000. Prvním zjišťovaným aspektem byly matematické postupy, jež zachycují, co žáci dělají, když propojují kontext problému s matematikou, aby ho mohli vyřešit. Tato dimenze byla oproti analogické dimenzi, jež byla použita v konceptuálním rámci v roce 2000, značně rozpracována. Kromě postupů do ní byly zahrnuty i základní dovednosti, jež jsou pro tyto postupy nezbytné. Konceptuální rámec zároveň popisoval propojení mezi dovednostmi a matematickými postupy, tedy ukazoval, jak se uplatňují dovednosti s rostoucí náročností matematických postupů.

Další dvě dimenze – matematický obsah a kontext – byly oproti roku 2000 jen mírně modifikovány. Hlavní charakteristiky konceptuálního rámce jsou uvedeny v tabulce 8.

2.2.3 Přírodovědná gramotnost ve výzkumu PISA

Přírodovědná gramotnost byla ve výzkumu PISA 2000 definována jako *schopnost využívat přírodovědné vědomosti, klást otázky a na základě důkazů vyvozovat závěry, které vedou k porozumění a usnadňují rozhodování týkající se světa přírody a změn, které v něm nastaly v důsledku lidské činnosti*.

Tabulka 9: Konceptuální rámec přírodovědné gramotnosti, PISA 2000

Přírodovědné postupy	Rozpoznání otázek, které je možno vědecky zkoumat Stanovení důkazů nezbytných pro přírodovědné zkoumání Vyvozování nebo hodnocení závěrů Sdělování platných závěrů Prokázání porozumění přírodovědným pojmům
Obsah	Struktura a vlastnosti hmoty Atmosférické změny Chemické a fyzikální změny Přeměny energie Síla a pohyb Forma a funkce Biologie člověka Fyziologické změny Biologická různorodost Genetika Ekosystémy Země a její postavení ve vesmíru Geologické změny
Situace	Přírodní vědy v životě a zdraví Přírodní vědy na Zemi a v životním prostředí Přírodní vědy v technice

Základní aspekty konceptuálního rámce jsou uvedeny v tabulce 9. Členění oblasti bylo analogické jako v případě matematické a čtenářské gramotnosti.

V roce 2000 definice přírodovědné gramotnosti (pod označením přírodovědné znalosti) zahrnovala *znalost přírodních věd* a *znalost o přírodních vědách*. V konceptuálním rámci v roce 2006, kdy byla přírodovědná gramotnost hlavní hodnocenou oblastí, byly tyto dvě dimenze odděleny. Koncept přírodovědné gramotnosti se také přestal omezovat jen na přírodovědné disciplíny a byl obohacen o vztah mezi vědou a technikou.

Do roku 2015, kdy byla přírodovědná gramotnost hlavní hodnocenou oblastí podruhé, se její koncept dále vyvinul. Největší rozdíl mezi konceptuálním rámcem z roku 2006 a 2015 spočívá v lepší specifikaci dimenze *znalost o přírodních vědách* a jejím rozdělení na dvě dílčí dimenze: *procedurální znalosti* (znalosti vědeckých postupů) a *epistemické znalosti* (znalost podstaty vědeckého zkoumání). Dalším významným obohacením je zahrnutí další dimenze, která je zaměřena na postoje studentů k vědě a technice. Postoje byly již součástí konceptuálního rámce v roce 2006, v roce 2015 byly však ještě rozpracovávány a modifikovány.

Přírodovědná gramotnost je pro účely PISA 2015 tedy definována jako *schopnost zabývat se vědeckými problémy a vědeckými myšlenkami jako přemýšlivý občan. Přírodovědně gramotná osoba je ochotna se zapojit do diskuse o vědě a technice, což vyžaduje následující kompetence:*

1. *vědecky vysvětlovat jevy: rozpoznat, nabízet a hodnotit vysvětlení pro řadu přírodních a technických jevů;*
2. *hodnotit a navrhovat vědecké zkoumání: popsat a ocenit vědecký výzkum a navrhnout způsoby, jak vědeckým způsobem odpovídat na otázky;*
3. *vědecky interpretovat data a vědecké důkazy: analyzovat a hodnotit různě reprezentovaná data, tvzení a argumenty a vyvozovat správné vědecké závěry.*

Výše uvedené tři okruhy kompetencí jsou zjišťovány ve třech kontextech: osobním, lokálním/národním a globálním. Úlohy zjišťují tři typy znalostí: 1. obsahové (fyzikální systémy, živé systémy, vesmír a Země), 2. procedurální (zahrnuje porozumění následujícím konceptům: proměnná, měření, stanovení a minimalizace chyby měření, opakovatelnost, abstrahování a znázorňování, kontrola proměnných, výzkumný design) a 3. epistemické (týkající se povahy vědeckého zkoumání). V rámci dotazníků jsou rovněž zjišťovány postoje: zájem, respekt k vědeckému zkoumání a ohled na životní prostředí.

Základní aspekty konceptuálního rámce přírodovědné gramotnosti z roku 2015 jsou uvedeny v tabulce 10.

2.2.4 Postoje k přírodním vědám a životnímu prostředí ve výzkumu PISA

Rozšíření přírodovědné oblasti reagovalo na dva aktuální problémy spojené s problematikou vědy a techniky. Prvním je rostoucí znečišťování životního prostředí a s tím spojené problémy a rizika. Zde bylo cílem přídatného šetření zjistit, do jaké míry jsou patnáctiletí žáci obeznámeni s jevy, které ohrožují životní prostředí, s jejich dopady, a do jaké míry je problematika životního prostředí zajímavá a jsou připraveni se v ní nějakým způsobem angažovat.

Druhým problémem je malý a stále se snižující zájem mladých lidí volit povolání v technických a přírodovědných oborech. Zde bylo cílem šetření zjistit, jací žáci o pracovní kariéře v této oblasti uvažují a co je k volbě technické a přírodovědné profílance motivuje.

Tabulka 10: Konceptuální rámec přírodovědné gramotnosti, PISA 2015

Přírodovědné kompetence	Vědecky vysvětlovat jevy	Například aplikace vědeckých poznatků a vysvětlování jejich dopadů pro společnost, používání a vytváření modelů, tvorba hypotéz a předpovědí.
	Hodnotit a navrhnout vědecké zkoumání	Například rozpoznat výzkumnou otázku v předložené studii, rozpoznat, co lze a co nelze vědecky zkoumat, navrhnout vědecké zkoumání a reflektovat jeho průběh, rozumět tomu, jak je ve vědeckém výzkumu zajišťována reliabilita, validita a zobecnitelnost.
	Vědecky interpretovat data a vědecké důkazy	Například analyzovat a interpretovat data a vyvozovat z nich správné závěry, ve vědeckých textech rozpoznat předpoklady, důkazy a zdůvodnění, rozpoznat vědecky podložené argumenty a hodnotit vědecké argumenty a důkazy z různých zdrojů.
Znalosti	Obsahové	Fyzikální systémy Živé systémy Systémy Země a vesmíru
	Procedurální	Znalost konceptů, na kterých je založeno vědecké zkoumání, např. závislá a nezávislá proměnná, typy měření, chyby měření a jejich minimalizace, opakovatelnost, výzkumný design, práce s daty a prezentace dat.
	Epistemické	Znalost náležitostí budování poznatků, např. porozumění konceptu hypotéz, teorií a pozorování a tomu, jak přispívají k našim znalostem.
Kontext	Situace	osobní lokální/národní globální
	Oblasti aplikace	zdraví a nemoc přírodní zdroje kvalita životního prostředí rizika hranice vědy a techniky
Postoje	Zájem o přírodní vědy Oceňování vědeckého přístupu ke zkoumání Úcta k životnímu prostředí	

2.2.5 Metakognitivní dovednosti ve výzkumu PISA

Odborníci zodpovědní za vývoj konceptuálního rámce čtenářské gramotnosti považovali od počátku za důležité zkoumat spolu se čtenářskými dovednostmi i dovednosti metakognitivní.

Vycházeli z poznatků, že povědomí o metakognici a schopnost ji regulovat mají pozitivní dopad na čtenářské dovednosti a že je možno je u žáků rozvíjet. V roce 2000 však bohužel nemohlo být zjišťování metakognitivních dovedností do výzkumu zařazeno, neboť nebyly k dispozici žádné nástroje na jejich měření. Výzkum se proto v tomto roce omezil na zkoumání studijních strategií, které určují, do jaké hloubky a jak systematicky žák zpracovává získané informace. Zahrnoval:

- motivační preference a cílové orientace, které ovlivňují množství času a myšlenkového úsilí investovaného do studia a volbu studijních strategií;
- osobní kognitivní mechanismy, které řídí cíle a postupy;
- strategie ovlivňující činnosti, zvláště úsilí a vytrvalost, které ve studijním procesu zamezují rozptýlení a pomáhají překonat studijní obtíže;
- preference různých typů studijních situací, stylů studia a sociálních dovedností potřebných pro skupinové učení.

V roce 2009 byla metakognice definována jako povědomí o různých strategiích pro zpracování textu a schopnost tyto strategie cíleně a aktivně využívat. Užívání strategických aktivit předpokládá znalost struktury textů a její využití při četbě. Při metakognitivních činnostech je důležité využívat jak znalostí o jazyce a struktuře textu, tak znalosti tématu k identifikaci relevantních informací a k propojení těchto informací s dříve získanými informacemi z textu nebo s informacemi uloženými v dlouhodobé paměti. Koncept hodnocení vycházel z poznatků, že kognitivní a metakognitivní dovednosti umožňují čtenáři využívat text pro dosažení kognitivních cílů a že metakognitivním dovednostem je možno se naučit a cíleně je využívat. Kladení otázek nad textem vede k aktivnějšímu zpracování textu; dalšími strategiemi jsou různé typy zvýrazňování a sumarizace, reflexe porozumění a postupu k překonávání překážek.

Ve výzkumu PISA 2009 byla respondentům předložena řada čtenářských scénářů (*vignettes*). V případě každého scénáře měli respondenti za úkol ohodnotit jeho užitečnost pro dosažení stanoveného čtenářského úkolu. Respondentem stanovené pořadí z hlediska užitečnosti bylo porovnáno s „optimálním“ pořadím stanoveným experty na zpracování textu (výzkumníci, učitelé a psychologové). Na základě shody mezi oběma pořadími byl vypracován výsledek (*metacognition score*), který ukazoval, do jaké míry jsou si žáci vědomi způsobů, jakými je možno nejlépe zpracovat textovou informaci, a jak rozumějí nárokům kladeným na paměť a porozumění. Tvůrci hodnocení vycházeli z toho, že k dosažení dobrého výsledku musí žáci aktivovat vědomosti o kognitivních zdrojích, o podstatě úkolu a strategiích, které usnadňují porozumění informacím, jejich zapamatování a vybavování.

2.2.6 Řešení problémů ve výzkumu PISA

Spolu s matematickou gramotností byla v letech 2003 a 2012 zjišťována dovednost řešit problémové úlohy. V roce 2003 byla tato oblast definována jako *schopnost jednotlivce využívat své kognitivní dovednosti k porozumění problémové situaci, která přímo nespadá do čtenářské, matematické či přírodovědné oblasti, a k jejímu vyřešení v případě, že způsob řešení není bezprostředně zřejmý*. Hlavní prvky konceptuálního rámce z roku 2003 jsou uvedeny v tabulce 11.

Od té doby, co byl vyvinut konceptuální rámec hodnocení pro sběr v roce 2003, byla usku-
tečněna celá řada studií v oblasti řešení problémů obecně, a zejména v oblasti řešení problémů

prostřednictvím počítačů. Poznatky z těchto výzkumů spolu s rozvojem v oblasti počítačových technologií vedly k zásadnímu přepracování konceptuálního rámce pro hodnocení prováděné elektronicky.

Tabulka 11: Konceptuální rámec řešení problémů, PISA 2003

Typy problémů	Rozhodování	Typické problémy, ve kterých musí respondent vybrat z nabízených možností tu, která nejlépe splňuje zadaná kritéria.
	Systémová analýza a design	Typické navrhování a modifikace systémů, ve kterých je třeba uvažovat o systému jako celku. Díky vazbě mezi jednotlivými parametry dojde při změně jednoho parametru ke změnám v celém systému.
	Odstraňování problémů	Typické odstraňování chyb ve fungování přístrojů.
Oborové kontexty	Matematika Přírodověda Literatura Společenské vědy Technika Obchod	
Procesy	Porozumění Charakterizování Reprezentování Řešení Reflektování Komunikování	porozumění podstatě problému charakteristika proměnných a vztahů mezi nimi výběr a modifikace reprezentace problému řešení úlohy reflexe řešení ve vztahu k problému komunikace výsledků
Myšlenkové dovednosti	Analytické uvažování Kvantitativní uvažování Analogické uvažování Kombinatorické uvažování	

Zde vycházeli tvůrci hodnocení z toho, že počítače umožňují žáky lépe motivovat zadáváním dynamických interaktivních problémů a provádět hodnocení efektivněji, neboť umožňují monitorování celého postupu řešení (umožňují například zaznamenávat typ akcí, jejich délku, četnost a návaznost). Počítačová administrace rovněž umožňuje se soustředit na dovednosti, které jsou klíčové pro řešení problémů, a nekontaminovat hodnocení těchto dovedností dovednostmi čtenářskými, matematickými a přírodovědnými, jako tomu bylo při hodnocení prostřednictvím tištěných testů v roce 2003.

V roce 2012, kdy byla hlavní testovanou oblastí opět matematická gramotnost a spolu s ní

bylo opět hodnoceno řešení problémů, byl kladen důraz na autentické komplexní úlohy, které vyžadují přímou interakci řešitele s počítačem k tomu, aby objevoval informace relevantní pro řešení problému. Do definice řešení problému byl též zakomponován afektivní prvek.

Řešení problémových úloh bylo ve výzkumu PISA 2012 definováno jako *schopnost jednotlivce využívat své kognitivní dovednosti k porozumění problémové situaci a k jejímu vyřešení v případě, že způsob řešení není bezprostředně zřejmý. Její součástí je i ochota jednotlivce zabývat se takovými situacemi, aby mohl jako konstruktivní a přemýšlivý občan rozvinout vlastní potenciál.*

Klíčovými prvky, které byly brány v úvahu při vytváření problémových úloh testu PISA 2012, jsou kontext problému, povaha problémové situace a postupy uplatňované při řešení problému. Kontexty byly charakterizovány jako technické a netechnické a osobní a sociální. Problémové situace byly buď statické nebo interaktivní (řešení problémů bylo hodnoceno výlučně elektronicky, což umožňovalo koncipovat interaktivní úlohy). Postupy uplatňované při řešení problému zahrnovaly: zkoumání a porozumění, znázorňování a formulování, plánování a provádění, sledování a posuzování. Zjišťovalo se, do jaké míry je žák schopen identifikovat určitý problém a přejít k jeho řešení, zda rozpoznává a chápe podstatu problému, formuluje problém, rozpozná základní proměnné a jejich vzájemné vztahy, vybere správný postup, přizpůsobí znázornění problémové situaci, dokáže posoudit a překontrolovat svoji práci a prezentovat výsledky. Dále se posuzovala dovednost analytického, kvantitativního, analogického a kombinatorického uvažování.

I když v roce 2015 byla hlavní hodnocenou oblastí přírodovědná gramotnost, bylo řešení problémů zařazeno i v tomto cyklu. Tentokrát však bylo posuzováno řešení problémů ve skupině (*collaborative problem solving – CPS*) (OECD 2013e). Zatímco řešení problémů v PISA 2012 se vztahuje k situaci, kdy jednotlivci řeší samostatně problémy, u kterých není na první pohled zřejmý způsob řešení, při skupinovém řešení problémů spolupracuje na vyřešení problému skupina jednotlivců tak, že spojí své porozumění a úsilí k nalezení společného řešení. Spolupráce má podle autorů hodnocení zjevné výhody oproti individuální práci, neboť:

- umožňuje efektivní dělbu práce;
- umožňuje zapracování informací z rozmanitých zdrojů poznatků, perspektiv a zkušeností;
- posiluje kreativitu a kvalitu řešení stimulací od ostatních členů skupiny.

Spolupráce je pro účely skupinového řešení problémů definována jako *koordinovaná, synchronizovaná aktivita, která je výsledkem kontinuální snahy vytvořit sdílené porozumění problému a postupovat v souladu s tímto sdíleným porozuměním.* Sociální interakce je pro skupinové řešení problémů zásadní, ale nikoli dostatečná, neboť některé sociální interakce nezahrnují společné cíle, zohlednění odlišných perspektiv a organizované úsilí k dosažení cílů.

Vzdělávací systémy kladou v poslední době rostoucí důraz na projektové a badatelské učení. To vyžaduje zaměření výuky na kritické myšlení, řešení problémů, sebekontrolu a spolupráci. Projektová práce často zahrnuje úkoly, při kterých musí skupina žáků spolupracovat k dosažení společného cíle, kterým bývá závěrečná zpráva, analýza nebo prezentace. Skupinové řešení se zpravidla neučí jako samostatná dovednost. Skupinové úkoly bývají často integrovány do rozmanitých předmětů, například do přírodovědných předmětů, matematiky nebo dějepisu. Důraz na komunikaci a spolupráci odpovídá konceptu dovedností pro 21. století, který se čím dál tím větší měrou uplatňuje v kurikulech vyspělých zemí.

Požadavek učit a hodnotit skupinové řešení problémů je veden potřebou připravit žáky pro povolání, která vyžadují schopnost pracovat efektivně ve skupinách a aplikovat schopnosti k řešení problémů v těchto sociálních situacích. Většina úkolů, které vyžadují řešení problé-

mů, je dnes realizována týmy v globalizované a elektronizované ekonomice. Nastal posun od průmyslové výroby k informačním a vědomostním službám. Ale i ve výrobě je práce zřídka vykonávána samostatně pracujícími jednotlivci. S rostoucí dostupností počítačových sítí je od jednotlivců čím dál tím více očekáváno, že budou s využitím technologií pracovat s rozmanitými týmy nacházejícími se v různých lokalitách.

Tabulka 12: Konceptuální rámec řešení problémů, PISA 2012

Kontext	Informační technologie Jiný kontext	úloha simuluje technologické zařízení zpravidla kontext běžného života, jako je rozhodování nebo plánování cesty či pracovního rozvrhu
Problémová situace	Interaktivní Statická	informace se odkrývají postupně při interakci s problémem, problém se může průběžně měnit všechny informace jsou k dispozici na začátku, situace je neměnná; problém může být dobře definovaný (všechny informace jsou k dispozici, je jasný cíl a priority), nebo špatně definovaný (např. nějaké informace nejsou k dispozici, je více cílů, které jsou vzájemně v rozporu apod.)
Procesy	Zkoumání a porozumění Reprezentace a formulování Plánování a vykonávání Monitoring a reflexe	prozkoumání problémové situace, její pozorování a interakce s ní, získávání informací, nalezení překážek, porozumění situaci grafická či verbální reprezentace problému a formulování hypotéz identifikací relevantních faktorů a jejich vzájemných vztahů stanovení celkového cíle a dílčích cílů a stanovení strategie k jejich dosažení, realizace strategie průběžné monitorování postupu včetně ověřování dílčích a finálních výsledků, identifikace nečekaných komplikací a jejich náprava, kritické hodnocení předpokladů a alternativních řešení, získávání dodatečných informací, komunikace výsledků
Myšlenkové dovednosti	Deduktivní uvažování Induktivní uvažování Kvantitativní uvažování Korelační uvažování Analogické uvažování Kombinatorické uvažování Multidimenzionální uvažování	

Dovednost spolupracovat je také potřebná v občanských kontextech, jako je tvorba společenských vazeb, účast v komunitním životě, dobrovolnictví, a také při realizaci administrativních úkonů a využívání veřejných služeb. Od studentů opouštějících vzdělávací systém a vstupujících

na pracovní trh bude vyžadováno, aby uměli řešit problémy ve spolupráci s ostatními a aby tak byli schopni činit s využitím vhodných technologií.

Potřeba kvalitní spolupráce v týmech je zásadní pro úspěch pracovních skupin, rodin, korporací, veřejných institucí i rozmanitých nevládních i vládních organizací. Jeden nekoooperativní člen týmu může negativně ovlivnit úspěch celého týmu, stejně jako dobrý vedoucí je pro týmovou práci pravým požehnáním. Jedním z hlavních faktorů, který přispívá k úspěchu týmové práce, je efektivní komunikace mezi členy týmu. Proto je komunikační zdatnost důležitou součástí konceptuálního rámce skupinového řešení problémů, které je založeno na následujících východiscích:

Komunikovat správné informace a podávat zprávu o uskutečněných krocích správným osobám ve správný čas umožňuje žákům vybudovat sdílené porozumění úkolu. Kompetence zahrnuje schopnost podívat se na problém pohledem ostatních členů skupiny, seznámit se se znalostmi ostatních členů skupiny a společně budovat a monitorovat postup při řešení úkolu.

Za druhé musí být studenti schopni nastavit a udržet efektivní organizaci týmu. To zahrnuje promyšlené rozdělení rolí a průběžnou modifikaci organizace tak, aby byla efektivní pro dosažení cíle. Znamená to poradit si s odlišnými názory, konflikty, překážkami, případně negativními emocemi.

Za třetí musí být studenti schopni porozumět typu kooperace a odpovídajícím pravidlům zapojení. Pravidla se liší v odlišných kontextech ve vztahu k vzájemné pomoci, skupinové práci, budování konsensu, vyjednávání s cílem maximalizovat prospěch pro všechny strany, k diskusím a situacím, kdy všichni zúčastnění nemají všechny informace.

Skupinové řešení problémů bylo v roce 2015 definováno jako *schopnost jedince se efektivně zapojit do procesu, ve kterém se dva a více jednotlivců snaží vyřešit problém prostřednictvím společného porozumění a úsilí potřebného k nalezení řešení a sdílejí znalosti, dovednosti a snahu k dosažení tohoto řešení*.

Hlavní součásti konceptuálního rámce skupinového řešení problémů tak, jak byl aplikován v PISA 2015, jsou uvedeny v tabulce 13.

Koncept měření problémů ve spolupráci vychází z výzkumů, které byly realizovány v uplynulých desetiletích v oblasti komunikace, individuálního a skupinového řešení problémů, počítačem podporované skupinové práce a týmového hodnocení; vyústily v řadu teoretických rámců, modelů a paradigmat. Výzkumy komunikace a skupinového řešení problémů byly zprvu realizovány bez využití počítače a spočívaly v pozorování skupiny v nějaké pečlivě volené situaci nebo při řešení nějakého úkolu. Pro posuzování výkonu skupiny a jednotlivců byly zpravidla používány metody vrstevnického hodnocení a hodnocení nezávislými experty, kteří se zúčastnili práce skupiny v roli pozorovatelů, prostřednictvím rozmanitých škál a dotazníkových šetření.

Přechod na hodnocení prostřednictvím informačních technologií umožnil přivádět respondenty do umně zorganizovaných situací a pozorovat jejich chování a reakce v těchto situacích. Řada prostředí vyvinutých pro měření dovedností k řešení problémů je uzpůsobena pro přirozené rozhodování (*naturalistic decision making*), v němž má každý jedinec vlastní úkoly, identitu a dovednosti, které musí být využity k dosažení cíle, který má dopad na jednotlivce i celou skupinu.

Tabulka 13: Skupinové řešení problémů, PISA 2015

Profil žáka	Nabyté vědomosti	matematika čtení a psaní přírodní vědy a životní prostředí každodenní učení
	Charakteristiky	předpoklady a postoje zkušenosti a znalosti motivace kognitivní schopnosti
Klíčové dovednosti	Dovednosti nezbytné ke spolupráci	zakotvení vysvětlování koordinace plnění rolí přijímání jiné perspektivy naslouchání argumentace vzájemná regulace
	Dovednosti k řešení problému	zkoumání a porozumění (prozkoumání dostupných informací a porozumění problémové situaci) reprezentace a formulování (organizace informací shromážděných v prvním kroku, identifikace relevantních faktorů, formulace hypotéz) plánování a provádění (stanovení konečného cíle a dílčích cílů a cest k jejich dosažení, realizace plánu) monitoring a reflexe (monitorování postupu, kritické hodnocení nalezených řešení a důležitých předpokladů, které je podmiňují)
Kompetence skupinového řešení problému	Vytvoření a udržení sdíleného porozumění	členové skupiny musí stanovit společnou znalost – co každý z nich o problému ví, identifikovat pohledy a perspektivy jednotlivých členů skupiny a domluvit se na společném vidění problému a cest k jeho řešení
	Provedení správných kroků k vyřešení problému	to znamená porozumět omezením, stanovit týmové cíle nutné k řešení, vykonat příslušné aktivity a průběžně monitorovat výsledky ve vztahu ke skupinovým cílům
	Vytvoření a udržení týmové organizace	členové týmu musí být schopni porozumět své roli a rolím ostatních, dodržovat odpovídající pravidla, monitorovat organizaci skupiny a usnadňovat změny nezbytné pro zvládnutí komunikačních obtíží, překážek při řešení a optimalizaci postupu

Kontext	Charakteristika úloh	otevřenost (dobře definovaný nebo špatně definovaný – viz PISA 2012) dostupnost informací (informace mohou být dostupné najednou nebo je respondent musí získávat postupně) vzájemná závislost (míra, do jaké mohou jednotliví členové skupiny vyřešit úlohu samostatně) symetrie cílů (cíle skupinové nebo individuální)
	Médium	semantická rozmanitost vazba na okolní svět prostor řešení problému (respondent je nebo není informován o akcích ostatních členů skupiny)
	Scénář úlohy	typ úkolu (např. skládačka, hledání shody, vyjednávání) kontext úlohy: soukromý/veřejný; technologický/netechnologický; školní (formální)/mimoškolní (neformální) obsahová oblast
	Tým	symetrie rolí (rozsah rolí, které jsou k dispozici pro každého člena týmu) symetrie statusu (všichni členové skupiny jsou na stejné úrovni vs. hierarchická struktura) velikost skupiny

Některá prostředí jsou vytvořena s cílem měřit kognitivní procesy členů skupiny během práce na řešení úkolu, nikoli jejich výsledek. V Evropě je k tomuto účelu široce využíváno prostředí Mycrodyn, jež bylo vyvinuté na univerzitě v Lucemburku. To slouží k pozorování respondentů v komplexních a dynamických situacích (Greiff a kol. 2012). Pro objektivní vyhodnocení výkonu respondentů jsou používány reakce respondentů v konkrétních situacích (které jsou počítačem zaznamenány) a odpovědi na položené otázky.

Ve výzkumu PISA bylo hodnocení realizováno prostřednictvím komunikace respondenta s jedním nebo několika „počítačovými spolupracovníky“. Komunikace se odehrávala prostřednictvím výběru z nabízených sdělení. V průběhu řešení úkolu docházelo k rozmanitým situacím a nedorozuměním, v jejichž rámci měl žák možnost prokázat měřené kompetence. Na závěr měl žák za úkol reflektovat, co by měl příště pro zlepšení spolupráce udělat lépe.

2.2.7 Finanční gramotnost ve výzkumu PISA

Kromě řešení problémů začala být spolu s matematickou gramotností jako hlavní oblastí ověřována i gramotnost finanční, proto je užitečné se seznámit i s jejím konceptem. Zařazení finanční gramotnosti bylo motivováno přesvědčením, že finanční gramotnost hraje v poslední době v životě občanů čím dál tím důležitější úlohu, a proto na ni kladou vyspělé země rostoucí důraz i ve vzdělávání. Finanční gramotnost nabývá podle autorů výzkumu PISA na významu proto, že v posledních desetiletích dochází k velkému přesunu rozmanitých rizik od vlád a zaměstnavatelů k jednotlivcům. Mnoho vlád redukuje státní penze a zdravotní péči, pracující jsou rostoucí měrou zodpovědní za spoření na důchod. Lidé také musí více šetřit, aby pokryli náklady na lékařskou péči pro sebe i pro členy své rodiny. Mnoho výzkumů ukazuje, že si lidé

často nejsou vědomí rizik, která jsou s tím spojena, a nemají znalosti k jejich zvládnutí. Počet finančních rozhodnutí, která jednotlivci musejí učinit, se zvyšuje nejen v důsledku ekonomických změn, ale také v důsledku změn ve společnosti. Například v důsledku vyšší očekávané délky života musí občané více šetřit, aby pokryli delší období strávené v důchodu. Důležité je i plánování a spoření na vzdělávání.

Autoři konceptu vycházejí z toho, že se výrazně zvýšil počet nabízených finančních produktů a zvýšilo se i množství a rozmanitost jejich poskytovatelů. Zároveň se zvýšila složitost produktů; jejich volba tedy vyžaduje porovnání velkého množství faktorů, jako jsou například poplatky, úroky, délka kontraktu nebo rizikovost produktu. Občané tedy musejí být finančně gramotní, aby činili informovaná a zodpovědná rozhodnutí. Ekonomický a technologický rozvoj přinesl globální propojení a masivní změny v komunikaci, finančních transakcích a chování spotřebitelů. Došlo též ke změnám v interakci s poskytovateli finančních produktů. Spotřebitelé potřebují elektronický kontakt, aby mohli provádět a přijímat elektronické platby. Ti, kteří k těmto platbám nemají přístup, často platí za transakce vyšší poplatky. Vzhledem k tomuto vývoji je od jedinců očekáváno, že budou finančně gramotní, aby byli schopni chránit sebe a své blízké a zajistit jejich finanční pohodu.

OECD zahájilo aktivity na poli finanční gramotnosti v roce 2002. V jejich rámci vytvořilo rovněž vodítka pro finanční vzdělávání ve školách. Cílem hodnocení finanční gramotnosti v mezinárodním projektu je pomoci jednotlivým zemím zjistit, zda je jejich přístup k finančnímu vzdělávání efektivní, a pomoci jim identifikovat oblasti, které je třeba ve vzdělávání posílit s cílem vybavit mladé lidi dovednostmi, jež budou potřebovat ke svému finančnímu rozhodování.

Pro účely výzkumu PISA 2012 byla finanční gramotnost definována jako *znalost finančních konceptů a porozumění těmto konceptům a rizikům a dovednosti, motivace a sebedůvěra nezbytné k uplatnění těchto znalostí a porozumění. Cílem je činit efektivní rozhodnutí v rozmanitých finančních kontextech, zlepšit finanční situaci jednotlivců i společnosti a umožnit zapojení do ekonomického života.* Základní charakteristiky konceptuálního rámce jsou uvedeny v tabulce 14.

Finanční gramotnost vykazuje určitý překryv s matematickou gramotností. Tento překryv však není zdaleka tak velký, jak by se mohlo na první pohled zdát. Spadají sem pouze základní aritmetické úlohy, ve kterých mají žáci za úkol aplikovat znalosti v kontextu běžných finančních situací (např. nakupování). V matematice mají kromě toho žáci za úkol řešit řadu úloh, které se vůbec nevztahují k financím, a dále úlohy, ve kterých není na první pohled zřejmá jejich matematická povaha. Do finanční gramotnosti naopak spadají úlohy, ve kterých mají žáci aplikovat pokročilejší finanční znalosti, přičemž potřebují nějaký smysl pro čísla, ale matematika je v těchto úlohách minimální a explicitní. Dále sem spadají úlohy, které matematiku vůbec neobsahují.

Tabulka 14: Konceptuální rámec finanční gramotnosti, PISA 2012

Obsah	Peníze a finanční transakce	funkce peněz, různé typy plateb a finančních převodů, úroky, provádění a sledování finančních transakcí, různé metody plateb, sledování pohybu na účtu prostřednictvím výpisů
	Plánování a hospodaření s financemi	různé typy příjmů a výdajů, hrubý a čistý příjem, rozpočet, výdaje a jejich snižování, stanovení priorit, spoření, půjčky, investice, daně a státní podpora
	Riziko a výnos	pojištění pro různé životní situace, eliminace finančních rizik, fluktuace úrokových sazeb a směnných kurzů, spoření v hotovosti, nákup nemovitostí, zlata apod., lichváři
	Finanční prostředí	práva a povinnosti spotřebitelů, smlouvy a jejich náležitosti a plnění, důvěryhodnost poskytovatelů finančních služeb, finanční poradenství, finanční kriminalita, změny v úrokových sazbách, daních či sociálních dávkách, inflace
Procesy	Identifikace finančních informací	vyhledávání a získávání zdrojů finančních informací a posuzování významu informací pro daný účel
	Analyzování informací ve finančním kontextu	kognitivní činnosti, které probíhají ve finančních souvislostech, jako např. zahrnující vysvětlování, porovnávání a poměřování, syntézu více informací a usuzování z dostupných informací
	Vyhodnocování finančních záležitostí	rozeznání či formulace finančních odůvodnění vyžadující např. vysvětlování, posuzování a zobecňování
	Uplatnění finančních znalostí a porozumění	efektivní jednání v prostředí financí prostřednictvím uplatňování znalostí finančních produktů
Kontexty	Vzdělání a práce	
	Domov a rodina	
	Osobní život	
	Společnost	

2.3 Konceptuální rámce oblastí hodnocených ve výzkumu PIAAC³⁰

Na závěr kapitoly věnované koncepci oblastí hodnocených v mezinárodních výzkumech výsledků vzdělávání je užitečné se ještě krátce podívat na odlišnosti v konceptualizaci testovaných oblastí v dospělé populaci. Zde jsou principiálně hodnoceny stejné oblasti vzdělávání, ale jejich koncepty jsou uzpůsobeny tomu, že jsou aplikovány nikoli v prostředí školy, ale v prostředí dospělého života.

³⁰ Konceptuální rámce oblastí hodnocených ve výzkumu PIAAC jsou spolu s výzkumnými nástroji, datovými soubory a vzorovým testem umístěny na stránkách <http://www.oecd.org/site/piaac/>. České verze konceptuálních rámců a výzkumných nástrojů jsou spolu s českou databází umístěny na www.piaac.cz.

Numerická gramotnost byla ve výzkumu PIAAC definována jako *schopnost získávat, používat, interpretovat a sdělovat matematické informace a představy s cílem zapojovat se do rozmanitých matematických situací života dospělých a zvládat jejich nároky*. Numerická gramotnost zde představuje dovednost řešit úkoly běžného života, které zahrnují nějaké číselné nebo jiné matematické prvky (např. geometrické, prostorové). Podle autorů konceptu numerické gramotnosti se její význam stále zvyšuje, neboť roste množství a různorodost kvantitativních a matematických informací, se kterými se v každodenním životě setkáváme.

Ve výzkumu bylo hodnoceno tzv. numericky gramotné jednání, které zahrnovalo 4 různé aspekty: kontext, typ matematické reakce, matematický obsah a matematickou reprezentaci. Kontext byl klasifikován jako každodenní, pracovní, společenský a jako další vzdělávání. Matematické reakce byly klasifikovány do 5 okruhů: identifikovat, najít či zjistit; zpracovat, použít; uspořádat, spočítat, odhadnout, vypočítat, měřit, modelovat; interpretovat; hodnotit/analyzovat; sdělovat. Matematický obsah byl členěn do 4 matematických oblastí: množství a číslo; rozměr a tvar; zákonitosti, vztahy, změny, data a pravděpodobnost. Tyto obsahy byly reprezentovány pomocí 6 různých reprezentací: objekty a nákresy; čísla a matematické symboly; vzorce, diagramy, mapy, grafy a tabulky; texty; elektronická zobrazení.

Čtenářská gramotnost byla ve výzkumu PIAAC definována jako *schopnost porozumět psaným textům, hodnotit je, používat je a zabývat se čtením, a to s cílem účastnit se života společnosti, dosahovat svých cílů a rozvíjet své vědomosti a potenciál*. Čtenářská gramotnost je pojímána jako dovednost porozumět psanému textu a získávat z něj informace. Předchozí výzkumy ukázaly, že čtenářská gramotnost největší měrou podmiňuje úspěch na pracovním trhu i zapojení do společnosti a je nezbytná pro celoživotní učení.

Úlohy byly konstruovány tak, aby zahrnovaly různé druhy textů, přičemž testy byly klasifikovány podle následujících kritérií: médium (tištěné nebo digitální), formát textu (text souvislý či nesouvislý), typ textu (typ promluvy: popis, výklad, argumentace, instrukce, záznamy), grafická úprava nesouvislého textu (různé typy seznamů, grafické dokumenty, vyhledávací dokumenty, formuláře), vlastnosti specifické pro digitální texty (různé typy hypertextů, interaktivní texty), společenský kontext (práce a zaměstnání, využití pro osobní potřebu, komunita a občanství, vzdělávání a školení). Výzkum ověřoval 3 typy čtenářských postupů: zjistit a rozpoznat informace v textu, integrovat a interpretovat (propojit jednotlivé části textu), hodnotit a reflektovat.

Řešení problémů v technologicky bohatých prostředích zahrnuje *použití digitálních technologií, komunikačních prostředků a sítí k získávání a hodnocení informací, ke komunikaci s ostatními a k provádění praktických úkolů*. Dovednost řešit problémy v prostředí informačních technologií zahrnuje práci s běžnými počítačovými aplikacemi, jako je elektronická pošta, vyhledávače, kalkulátory, programy na zpracování textu. V této oblasti je hodnocena dovednost získávat, vyhodnocovat a zpracovávat informace, přičemž pozornost je věnována jak kognitivním dovednostem, tak zručnosti v práci v počítačovém prostředí.

V této oblasti bylo hodnoceno 5 kognitivních aspektů: stanovení cíle a kontrola postupu, plánování a sebekontrola, získávání a hodnocení informací a použití informací. Zároveň byly hodnoceny 4 technologické aspekty: hardwarová zařízení (zde byl použit pouze počítač, ale tvůrci konceptuálního rámce chtěli zohlednit i možnost jiných zařízení); softwarové aplikace (správce souborů, internetový prohlížeč, e-mail, tabulkový procesor); příkazy, funkce (tlačítka, odkazy, textová pole, kopírovat/ vyjmout – vložit, seřadit, najít); reprezentace (texty, zvuky, čísla, grafika statická i animovaná, videa).

Výzkum PIAAC umožňoval rovněž zjišťování úrovně základních čtenářských dovedností u osob, které mají problémy se čtením jednoduchého textu, s cílem poskytnout informaci

o zastoupení těchto osob v dospělé populaci a specifikovat jejich vzdělávací potřeby. Předchozí výzkumy ukázaly, že ve většině vyspělých zemí je významné zastoupení občanů, kteří mají problémy s porozuměním čtenému textu. Tyto výzkumy však nebyly schopny rozlišit mezi těmi, kteří neumějí číst, a těmi, kteří číst umějí, ale mají nedostatečnou znalost jazyka testu.

Součástí výzkumu byl rovněž doprovodný dotazník, který mapoval demografické charakteristiky respondentů, jejich vzdělávání (počáteční i další), používání čtenářských, matematických a počítačových dovedností v běžném životě, pracovní zkušenosti a rodinné zázemí. Respondenti rovněž vyplňovali dotazník *Požadavky trhu práce*, který podrobně mapoval, do jaké míry využívají ve svém zaměstnání čtenářské, matematické a počítačové dovednosti i další obecné dovednosti, jež podmiňují komunikaci, prezentaci nebo práci v týmu.

3. Metodologie mezinárodních výzkumů

Oddíl 1.6.1 poskytl souhrnnou informaci o realizaci mezinárodních výzkumů výsledků vzdělávání. Tato kapitola se věnuje podrobně dílčím metodologickým aspektům. Zaměřuje se zejména na postupy, jejichž využití má důsledky pro práci s daty. Často se jedná o postupy, které u nás nejsou běžně známé a používané. Postupy jsou zpravidla vyloženy pouze na jednom z pojednávaných výzkumů, neboť jsou pro jednotlivé výzkumy velmi podobné. Kapitola neaspiruje na to podat precizní statistický výklad použitých metod, ale snaží se představit jejich filosofii a poskytnout konkrétní doporučení pro práci s datovými soubory. Datové soubory jsou informačně mimořádně bohaté. Čeští výzkumníci však jejich potenciálu dostatečně nevyužívají. To platí zejména v případě dat z žákovských, učitelských a ředitelských dotazníků. Cílem kapitoly je napomoci širšímu využití těchto zdrojů. Na počátku je popsána konstrukce testů na základě konceptuálních rámců, jež jsou uvedeny v předcházející kapitole, a konstrukce doprovodných dotazníků pro výzkumy IEA a pro výzkum PISA. Následně je podrobněji pojednáno o vážení dat, o metodách zohlednění víceetapovitosti výběru a o metodách škálování. Kapitola vychází z technických zpráv, kterými byly dokumentovány poslední cykly výzkumů (Martin; Mullis 2012, OECD 2014a, OECD 2014b). Poslední část kapitoly popisuje veřejně dostupné nástroje, které je možno s výhodou využít pro zpracování dat. V kapitole jsou na konkrétních příkladech demonstrovány chyby, kterých se snadno dopustíme, nezohledníme-li specifickou povahu dat.

3.1 Konstrukce testů a dotazníků

3.1.1 Konstrukce testů

Již bylo uvedeno, že vývoji testů pro každou vlnu testování předchází revize konceptuálního rámce. Tato revize je prováděna mezinárodními týmy odborníků na jednotlivé testované oblasti, které jsou ustaveny při každém mezinárodním výzkumu. Konceptuální rámce jsou upravovány podle aktuálních vědeckých poznatků s tím, že musí být zachována dostatečná kontinuita, aby výsledný test, který bude na základě konceptuálního rámce vytvořen, umožňoval spolehlivé sledování trendů (např. v TIMSSu odborníci doporučili modifikovat maximálně 5 % hodnocených oblastí). Konceptuální rámce obsahují také informace o testovém designu³¹, neboť testový design určuje dobu testování a typy úloh, které je nutno vyvinout.

V případě výzkumu PIRLS jsou zúčastněné země následně vyzvány, aby zasílaly do mezinárodního centra texty. Volba vhodných textů je velmi obtížný úkol: texty musejí být vhodné pro žáky 4. ročníku z hlediska obsahu, zajímavosti a náročnosti, musejí být dostatečně komplexní, aby umožňovaly rozmanitou škálu otázek, a nesmějí znevýhodňovat žádnou skupinu žáků.

Ve výzkumech IEA jsou nové úlohy vytvářeny na k tomu určeném semináři, v jehož rámci zástupci národních center pod vedením zkušených tvůrců úloh vytvářejí úlohy nové, které jsou podle konceptuálního rámce nezbytné k doplnění souboru testových úloh z předcházejících šetření³². Pro účely vývojového procesu byly vytvořeny speciální pokyny: *TIMSS Item Writing*

³¹ Skladba úloh a jejich rozložení do testových sešitů.

³² Doplnění úloh je nezbytné z toho důvodu, že část úloh je po ukončení každé vlny výzkumů zveřejněna a konceptuální rámce jsou průběžně inovovány, jak bylo uvedeno v předchozí kapitole.

Guidelines a *PIRLS Item Writing Guidelines*. Pro pilotáž je vždy vyvinut zhruba dvojnásobek úloh, než by měl obsahovat finální test, přičemž skladba úloh musí odpovídat struktuře konceptuálního rámce. Úlohy jsou následně revidovány a dopracovávány odborníky na tvorbu testových úloh, v některých případech jsou některé úlohy podrobeny malému pilotnímu šetření, aby se ověřila jejich funkčnost. Na ověřování úloh jsou využívány také kognitivní laboratoře (*cognitive labs*), ve kterých výzkumníci rozmlouvají s žáky nad úlohami a mají možno dobře posoudit srozumitelnost a přiměřenost úlohy pro danou věkovou kategorii. Po dopracování úloh, případných předběžných pilotážích a výběru nejlepších z nich jsou vytvořeny testové bloky a zaslány národním centrům k připomínkám. Národní centra připomínkují zejména aspekty související s převoditelností úloh do jiných jazyků a kultur. Úlohy například nesmějí obsahovat výrazy a rčení, které jsou specifické pro určité kultury a jsou obtížně přeložitelné.

V rámci pilotního šetření dochází k ověřování úloh tak, že každou testovou úlohu musí řešit minimálně 200 respondentů z každé země. Pilotáže se účastní v každé zemi zhruba 30 škol.

Výzkumy TIMSS a PIRLS obsahují zhruba polovinu úloh uzavřených a polovinu úloh otevřených. Úlohy TIMSS jsou rozděleny do 28 bloků: polovina je matematických a polovina přírodovědných. Bloky jsou uspořádány do 14 testových sešitů podle specifického systému, a to tak, že každý sešit obsahuje dva matematické a dva přírodovědné bloky. Každý blok se nachází ve dvou testových sešitech, přičemž umístění bloků se střídá, aby se vykompenzovala chyba spojená s umístěním bloku v testovém sešitu (ukazuje se, že úlohy umístěné na začátku a na konci testového sešitu mají odlišnou úspěšnost). Tedy v jednom případě je žáci řeší na začátku a v druhém případě na konci testování. Pro žáky ve 4. ročníku je celková doba testování 72 minut (18 minut na blok), pro žáky v 8. ročníku 90 minut (22,5 minuty na blok). Při administraci mají žáci mezi první a druhou dvojicí bloků přestávku. Celkový počet úloh pro matematiku i pro přírodovědné předměty je 170 pro 4. ročník a 200 pro 8. ročník.

V rámci výzkumu PIRLS existuje 10 bloků, přičemž každý blok obsahuje text a 12-17 úloh, které se k tomuto textu vztahují. Ty jsou rozděleny do 12 testových sešitů. V rámci výzkumu PIRLS byla vytvořena lehčí varianta testu, tzv. PRE-PIRLS pro země, jejichž žáci nejsou schopni dostatečně zdatně pracovat s běžnými testovými úlohami.

Ve výzkumu PISA jsou postupy velmi podobné jako ve výzkumech IEA s tím, že k aktualizaci konceptuálních rámců dochází zpravidla vždy v oblasti, která je v nadcházející vlně šetření hlavní. Zároveň jsou připravovány konceptuální rámce pro nové oblasti (např. v roce 2012 pro finanční gramotnost). Inovacím v konceptuálních rámcích je věnována velká péče. Například aktualizace matematického konceptuálního rámce v roce 2012 byla založena na studiu matematických standardů vyspělých zemí a konzultaci s matematickými experty. V rámci konzultačního procesu byl vyvinut obsáhlý dotazník, který vyplnilo cca 80 odborníků z celého světa.

I výzkum PISA klade velký důraz na to, aby byly testové úlohy vytvářeny odborníky v různých kulturních prostředích. Z toho důvodu poskytuje mezinárodní centrum k tvorbě úloh národním týmům podporu v podobě pokynů k tvorbě úloh (např. *Item Development for PISA 2012* a *Item Submission Guidelines*) a výcviku pro národní autory úloh. V pokynech je zdůrazněno, že úlohy by měly před zasláním do mezinárodního centra projít ověřením v kognitivní laboratoři s respondenty ve věku odpovídajícím testované populaci. Zúčastněné země možností tvorby úloh využívají. Pro šetření PISA 2012 zaslala národní centra do mezinárodního koordinačního centra zhruba 500 návrhů na testové úlohy. Tyto úlohy byly posouzeny v mezinárodním centru a vybrané z nich byly předány expertním týmům k dalšímu zpracování. Do vývoje úloh bylo zapojeno 10 pracovišť s velkou zkušeností s tvorbou úloh: 2 australská, 1 belgická, 3 německá,

1 japonské, 1 lucemburské, 1 norské a 1 americké. Tato pracoviště dotvořila dodané podněty a vyvinula rovněž množství vlastních originálních úloh. Součástí vývoje byla oponentura v rámci pracoviště i v rámci ostatních expertních týmů, kognitivní laboratoře a menší pilotní ověřování.

Součástí plošného pilotního ověřování bylo v roce 2012 kromě úloh, které byly administrovány již v předcházejících šetřeních, 172 nových matematických úloh. Úlohy, které byly nakonec zařazeny do hlavního šetření, musely splňovat následující kritéria:

- dobré psychometrické vlastnosti;³³
- konsistentní kódování (byly vyřazeny úlohy, které činily problémy při vyhodnocování);
- vysoké hodnocení od národních koordinátorů (úlohy byly hodnoceny z hlediska kvality v jednotlivých národních centrech);
- odpovídající zastoupení úloh dle koncepčního rámce;
- odpovídající zastoupení obtížností úloh.

Ve výzkumu PISA jsou úlohy soustředěny kolem společných textů nebo jiných úvodních materiálů – obrázků, grafů apod. (*stimulus*), podobně jako ve čtenářských testech výzkumu PIRLS. Úvodní texty tvoří spolu s úlohami, které se k nim váží, jednotky, které jsou seskupovány do bloků a promyšleně a systematicky umisťovány do testových sešitů. V PISA 2012 bylo vytvořeno 56 matematických jednotek se 110 úlohami, které představovaly celkem 270 minut testového času. 36 úloh bylo stejných jako v letech 2003, 2006 a 2009, 74 úloh bylo nově vytvořeno a vybráno ze souboru 174 úloh pilotovaných pro PISA 2012. Čtenářský test obsahoval 13 jednotek se 44 úlohami (podmnožina 131 úloh z testu z roku 2009), které reprezentovaly 90 minut testového času. Přírodovědný test byl rovněž 90minutový a obsahoval 18 jednotek s 53 úlohami.

Ve výzkumu PISA jsou úlohy rovněž rozdělovány do bloků, přičemž četnější jsou bloky pro oblast, na kterou je v dané vlně kladen důraz: např. ve výzkumu PISA 2012 bylo celkem 13 bloků, přičemž 7 bloků obsahovalo matematické úlohy, 3 bloky přírodovědné úlohy a 3 bloky čtenářské úlohy. Každý blok reprezentoval 30 minut testovacího času. Bloky byly přiřazeny 13 testovým sešitům, každý sešit obsahoval 4 bloky, tedy odpovídal 2 hodinám testování. Bloky ve vedlejších oblastech zpravidla zůstávají beze změn, v hlavní oblasti jsou opět tři bloky beze změn, aby zprostředkovaly vazbu na předcházející kola šetření, a čtyři bloky přinášejí nový materiál.

Uspořádání bloků do testových sešitů odpovídá vyváženému neúplnému blokovému designu (*balanced incomplete block design*). V tomto designu se každý blok (a tedy každá testová úloha) vyskytuje ve 4 testových sešitech, pokaždé v jedné z možných pozic v sešitě. Každá dvojice bloků se objevuje v právě jednom sešitě. Testový sešit 12 je zcela identický s jedním ze sešitů administrovaných v rámci předchozích kol výzkumu. Každému z respondentů je náhodně přidělen jeden z testových sešitů. Po jedné hodině testovacího času mají žáci cca 5 minutovou přestávku.

Kromě výše uvedených 13 sešitů byl vytvořen ještě tzv. UH sešit (*Une Heure booklet*) pro žáky se speciálními vzdělávacími potřebami. Tento sešit obsahoval zhruba poloviční počet úloh, přičemž v roce 2012 v něm matematika tvořila 50 % úloh a přírodní vědy a čtení po 25 %. Úlohy byly vybrány z běžných sešitů tak, aby byly vhodné pro žáky se speciálními vzdělávacími potřebami.

V posledních letech bylo ve výzkumu PISA zemím s výsledky na spodním konci škály nabídn-

³³ Zejména přiměřená obtížnost a dobrá citlivost úlohy, dále například dobré fungování jednotlivých distraktorů (u úloh s výběrem odpovědi) nebo očekávané vlastnosti chybných odpovědí (u otevřených úloh).

nuto, aby administrovaly méně obtížnou sadu testových sešitů s tím, že jejich výsledky budou prezentovány na stejné škále jako výsledky ostatních zemí. Testové úlohy byly voleny tak, že umožňovaly propojení: např. v roce 2012 byly 2 bloky matematických úloh nahrazeny lehčími úlohami, přičemž zbylé bloky zůstaly stejné. Cílem bylo učinit testování pro žáky méně frustrujícím a získat vyšší rozlišení na spodním konci škály.

V případě, že součástí testování byla nějaká další volitelná oblast, např. v roce 2012 finanční gramotnost, byly pro zúčastněné země připraveny další testové sešity, které kromě původních bloků obsahovaly též bloky s úlohami z této volitelné oblasti. Například pro finanční gramotnost byly vytvořeny další 4 sešity. Každý sešit obsahoval 2 bloky úloh na ověřování finanční gramotnosti, jeden blok matematický a jeden blok čtenářský.

V roce 2012 mohly zúčastněné země administrovat kromě písemných testů ještě elektronické testy. Mohly volit buď pouze test řešení problémů anebo ještě elektronický test matematický a čtenářský. V tom případě byly respondentům přiděleny ještě bloky elektronických úloh.

3.1.2 Konstrukce dotazníků

Důležitou součástí mezinárodních výzkumů výsledků vzdělávání jsou dotazníky pro žáky, učitele a ředitele škol. Dotazníková šetření jsou (stejně jako testy) koncipována prostřednictvím konceptuálních rámců, které přehledně sumarizují typy informací, jež jsou využity pro interpretaci výsledků žáků a jejich vzájemné vztahy. Konceptuální rámce jsou vytvářeny na základě teoretických poznatků o faktorech, které ovlivňují výsledky žáků na úrovni systému, školy, učitele a rodiny.

V tabulce 15 uvádíme konceptuální rámce dotazníkových šetření výzkumu TIMSS 2015. Konceptuální rámec dotazníkových šetření výzkumu PIRLS je analogický konceptuálnímu rámci TIMSS s tím, že se zaměřuje na výuku jazyka a čtenářských dovedností. Do výzkumu PIRLS byl zařazován rovněž rodičovský dotazník, který zjišťuje, jaké aktivity, jež rozvíjejí čtenářství, vykonávali rodiče s dětmi v jejich raném dětství (Mullis; Martin 2013).

Standardní součástí výzkumu PISA jsou pouze dotazníky pro ředitele a žáky. Oblast výuky je ve výzkumu PISA obecně mapována méně podrobně než v šetřeních IEA. Ta si od počátku stanovila jako primární cíl zkoumat souvislosti mezi výukou a výsledky vzdělávání. Z toho důvodu se například uskutečnila v roce 1998 videostudie, v jejímž rámci byly u několika zemí, které dosáhly dobrých výsledků ve výzkumu TIMSS 1995, pořízeny standardizovaným způsobem videonahrávky 100 náhodně vybraných hodin matematiky a přírodovědných předmětů v 8. ročníku. Tyto nahrávky byly následně pečlivě okódovány (podle velmi detailního a sofistikovaného kódovacího systému) a analyzovány, s cílem ukázat rozdíly a společné charakteristiky výuky v jednotlivých zemích (Hiebert a kol. 2003, Roth a kol. 2006). Tohoto výzkumu se zúčastnila i Česká republika.

V rámci pilotáže učitelských dotazníků výzkumu TIMSS 1995 například proběhl také pokus ohodnotit odborné a pedagogické znalosti učitelů. Učitelům bylo předloženo několik pedagogických situací, ve kterých měli z nabízených možností zvolit správný postup a zdůvodnit svoji volbu. Zástupci zúčastněných zemí nakonec tuto část šetření do finálního výzkumu nezařadili, neboť se báli negativní reakce učitelů.

Tabulka 15: Konceptuální rámec doprovodného dotazníkového šetření, TIMSS 2015

Národní kontext	Ekonomické zdroje, demografické trendy, geografické charakteristiky Organizace a struktura vzdělávacího systému Průchod žáků vzdělávacím systémem Vyučovací jazyk Zamýšlené matematické a přírodovědné kurikulum Učitelé a jejich vzdělávání Monitoring implementace kurikula
Kontext rodiny	Domácí výukové zdroje Jazyk komunikace v rodině Vzdělanostní očekávání rodičů a akademická socializace Rané čtenářské, matematické a přírodovědné aktivity
Kontext školy	Sídlo školy Socioekonomické složení žáků školy Omezení výuky nedostatkem matematických a přírodovědných výukových zdrojů Dostupnost učitelů a stabilita sboru Pedagogické vedení Důraz na akademický úspěch Bezpečnost, pořádek a disciplína ve škole
Kontext třídy	Učitelská příprava a pracovní zkušenosti Pokrytí matematických a přírodovědných témat testovaných ve výzkumu TIMSS ve výuce Výukové zdroje a technologie Výukový čas Zapojení žáků do výuky Hodnocení výsledků žáků
Charakteristiky žáků a jejich postoje k učení	Připravenost žáků k učení Motivace žáků Sebepojetí žáků Žákovské charakteristiky

Větší důraz na výuku ve studiích IEA souvisí bezesporu s tím, že organizace byla založena pedagogickými pracovišti a vznikla s cílem realizovat pedagogický výzkum. OECD je koncipována jako organizace sdružující vlády jednotlivých zemí a její důraz spočívá – spíše než na detailním zkoumání pedagogické práce – na zkoumání fungování celých systémů. Výzkumným záměrům je podřízen i výběr respondentů. Zatímco ve výzkumech IEA je testována vždy celá třída a je tedy možno se dotazovat učitelů, kteří ji vyučují, a tím pádem zkoumat rozličné aspekty výuky

z pohledu žáků i učitelů, ve výzkumu PISA jsou vybíráni náhodně žáci z různých tříd a ročníků. Ve výzkumu PISA by tedy bylo obtížné a náročné získat výpovědi od všech učitelů testovaných žáků tak, aby mohly být propojeny s žákovskými daty. Ve výzkumu PISA 2015 došlo k zařazení učitelského dotazníku pro učitele přírodovědných předmětů jako volitelné součásti výzkumu, a to proto, aby bylo možno získat detailnější informaci o přístupu škol k výuce přírodovědných předmětů a ke kooperativním aktivitám. Data učitelů však byla agregována na úroveň školy, nebyla vůbec propojována s daty jednotlivých žáků.

Součástí výzkumů IEA je rovněž národní dotazník, ve kterém zúčastněné země poskytují informace o kurikulu a o vzdělávacím systému a jeho aktuálním vývoji. Tyto údaje jsou následně publikovány v „encyklopediích“ (např. Mullis a kol. 2012a), které poskytují základní komparativní informace o výuce a učení se matematice a přírodovědným předmětům v zúčastněných zemích.

Výzkum PISA je naopak charakteristický velmi podrobným zjišťováním rodinného zázemí žáků (žáci jsou dotazováni nejen na vzdělání, ale rovněž na zaměstnání rodičů a na majetkové a kulturní vybavení domácností). Podoba mapování rodinného zázemí vychází z toho, že jedním z důležitých cílů výzkumu PISA bylo od samého počátku sledování vzdělanostních nerovností v jednotlivých vzdělávacích systémech.

Dotazníkové šetření výzkumu PISA je ovlivněno tím, že se střídají hlavní a vedlejší testované oblasti. Postupem času se ukázalo, že je třeba vymezit jádro, které bude v dotaznících neměnné a zajistí srovnatelnost v čase, a zároveň zajistit určitou flexibilitu, která umožní zaměřit část dotazníku na specifika jednotlivých oblastí. Zároveň bylo třeba vytvořit koncepci dotazníku založenou na poznatcích pedagogických výzkumů a omezit nahodilost, se kterou měli tendenci k tvorbě dotazníku přistupovat tvůrci vzdělávacích politik. K tvorbě konceptuálního rámce, který si kladl tuto ambici, došlo v rámci výzkumu v roce 2012. Dotazníkové šetření ve výzkumu PISA by podle tohoto rámce dlouhodobě měla obsahovat čtyři skupiny proměnných:

1. Obecné proměnné pro všechny kola a cykly: rodinné zázemí, charakteristiky školy týkající se zdrojů a rozhodovacích procesů, charakteristiky výuky, obecné nekognitivní výstupy (záškoláctví, motivace k učení, pocit sounáležitosti se školou apod.).
2. Trendové proměnné související s jednotlivými oblastmi: nekognitivní výstupy spojené s jednotlivými oblastmi (učební strategie a metakognice, přesvědčení týkající se jednotlivých vzdělávacích oblastí, obliba, motivace), příležitosti k učení jednotlivých oblastí, vyučovací metody, kvalita výuky.
3. Tematická rozšíření: volitelné moduly (např. vzdělávací dráha, znalost ICT), kontextové proměnné pro nové oblasti, vysvětlující proměnné pro specifické analýzy (diferenciace vzdělávacích drah, certifikace učitelů apod.).
4. Systémová data, která jsou získávána zpravidla z jiných zdrojů: financování, účast ve vzdělávání apod.

Proměnné, které byly použity ke sledování trendů, byly ve výzkumu PISA samozřejmě používány i v předchozích kolech. Například k měření trendů mezi lety 2003 a 2012 byly použity tyto trendové proměnné: pocit sounáležitosti se školou, vztahy mezi učiteli a žáky, kázeň při výuce matematiky, podpora učitele při výuce matematiky, obava z matematiky, vztah ke škole, instrumentální motivace k učení matematiky, obliba matematiky, matematické sebepojetí a vnímaná zdatnost při učení se matematice.

Cílem konceptuálního rámce z roku 2012 bylo využít zkušeností z předchozích sběrů ke

stanovení jakési „déledobější“ strategie, která zajistí, že budou plánovitě získávány všechny důležité kontextové proměnné a nebude k jejich výběru přistupováno nahodile.

Konceptuální rámec dotazníkového šetření z výzkumu PISA 2015 (OECD 2014c) vycházel z konceptuálního rámce pro rok 2012 a ještě jej rozpracoval do větších detailů. Je založen na východisku, že tvůrci vzdělávací politiky v zúčastněných zemích potřebují být informováni o následujících čtyřech oblastech: 1. výsledky (kognitivní i nekognitivní), 2. rodinné zázemí žáků, 3. procesy vyučování a učení a 4. řízení na úrovni školy a na úrovni systému:

1. Cílem výzkumu PISA je změřit výsledky vzdělávání, kterých žáci dosáhli ve věku 15 let. Tyto výsledky jsou nejen kognitivní, ale též nekognitivní povahy. Úspěch ve škole a v životě nezávisí jen na tom, co umíme, ale také na tom, jak jsme motivováni se učit a spolupracovat s ostatními, jak jsme schopni řídit a monitorovat své učení, čemu věříme, jak jsme ochotni se angažovat a jak jsme schopni respektovat druhé a porozumět jim. Z toho důvodu PISA měří nekognitivní výstupy, jako jsou postoje, přesvědčení, motivace a aspirace a učební strategie. Tyto výstupy jsou zjišťovány zejména prostřednictvím žákovského dotazníku, ale též prostřednictvím dotazníku školního. Mohou být obecné (např. výkonová motivace, předčasné odchody ze vzdělávání) nebo vztažené k jednotlivým oblastem (čtenářské zájmy, obliba matematiky apod.). Ty pak slouží jako vazba mezi konceptuálními rámci jednotlivých oblastí a konceptuálním rámcem dotazníkových šetření. Zde se jako zvlášť silná vazba ukázalo sebepojetí žáků v jednotlivých oblastech, tedy informace o tom, do jaké míry žáci věří tomu, že vyřeší úlohy v dané oblasti.
2. Informace o rodinném zázemí žáka, jako je například socioekonomický status a etnicita, je důležitá pro studium spravedlivosti vzdělávacích systémů. Rozložení vzdělávacích příležitostí a výsledků podle těchto proměnných ukazuje, zda jsou země úspěšné v poskytování stejných příležitostí všem žákům. Ve výzkumu PISA bylo definování a operacionalizaci rodinného zázemí věnováno mnoho úsilí, které nakonec vyústilo v konstrukci ukazatele ekonomického, sociálního a kulturního statusu (*ecsc*). Složky tohoto ukazatele musí být zjišťovány v jednotlivých cyklech stabilním způsobem. Tento ukazatel se používá rovněž v agregaci jako charakteristika školy vypovídající o složení žáků, kteří školu navštěvují. Spolu s tímto ukazatelem je zjišťována i podpora žáků v rodině. S rostoucím důrazem na předškolní vzdělávání je dále mapována celá vzdělávací dráha žáka.
3. Tvůrci vzdělávacích politik potřebují informaci o vyučování, učení a organizaci školní výuky. Tyto charakteristiky jsou zpravidla zjišťovány ve vztahu k hlavní testované oblasti. Sledované koncepty vycházejí z výzkumů efektivity školy, sleduje se kvalifikace učitelů, vyučovací metody a klima třídy a příležitosti k učení. Vyučovací proces je popisován prostřednictvím tří konceptů: struktura a management třídy, podpora žáků učitelem, náročnost učiva (*cognitive challenge*). Vzhledem k tomu, že ve výzkumu PISA není administrován učitelský dotazník, pocházejí informace z žákovského a školního dotazníku.
4. Výzkumy efektivity vzdělávání ukázaly, že na úrovni školy jsou z hlediska kvality učebních výsledků důležité následující faktory: profesní rozvoj, kurikulum, pedagogické vedení a řízení školy, zapojení rodičů, školní klima a využití hodnocení pro zlepšování vzdělávacích výsledků. Tyto koncepty jsou hodnoceny v PISA jako obecné koncepty pomocí školního dotazníku. Zároveň jsou zjišťovány aspekty podpory v hlavní testované oblasti, například dostupnost laboratoří, knihoven, informačních technologií a podobně.

Zároveň musí PISA ovšem zjišťovat aspekty řízení na úrovni systému, jako je například rozdělení

zodpovědností, řízení kvality a podobně. Některé z těchto informací je možno zjistit z jiných zdrojů (např. z ročenky OECD *Education at a Glance*), jiné jsou zjišťovány prostřednictvím školního dotazníku.

Komponenty dotazníkového šetření ve výzkumu PISA 2015 jsou uvedeny v tabulce 16. Dotazníkové šetření je zde rozvrženo do 18 modulů spadajících do 3 hlavních oblastí: rodinné zázemí, procesy (na úrovni školy i systému) a nekognitivní výstupy.

Tabulka 16: Konceptuální rámec dotazníkového šetření PISA 2015

Rodinné zázemí		Procesy			Nekognitivní výstupy	
Rodina	Vzdělávání	Aktéři	Klíčové procesy	Distribuce zdrojů		
		Vyučování a učení				
	5. Mimoškolní zkušenosti s přírodními vědami	1. Kvalifikace učitelů a odborné znalosti	2. Výukové metody v přírodovědných předmětech	12. Časová dotace a kurikulum	4. Výstupy týkající se přírodovědných předmětů: motivace, zájem, přesvědčení	6. Povolání v přírodovědné oblasti
		Školní praxe				
		3. Školní výukové prostředí pro přírodovědné předměty				
7. Socioekonomický status žáka & rodina	9. Vzdělávací dráha v raném dětství	14. Zapojení rodičů	13. Školní klima: vztahy mezi aktéry, důvěra, očekávání	16. Zdroje	10. Obecné chování a postoje	
8. Etnicita & migrace		15. Pedagogické vedení a řízení			11. Předpoklady pro skupinové řešení problémů	
		Řízení				
		17. Rozhodování ve vzdělávacím systému	19. Hodnocení a akontabilita	18. Alokační, selekce a volba		

Zdroj: OECD 2014c

Pro ilustraci toho, jak jsou jednotlivé moduly dále rozpracovávány, uvádíme v tabulce 17 detailní rozpis modulů 4 a 10, které jsou zaměřeny na nekognitivní výstupy. Ve výzkumu PISA 2015

byla hlavní sledovanou oblastí přírodovědná gramotnost; oborově specifické nekognitivní cíle se tedy váží k přírodním vědám.

Tabulka 17: Měření nekognitivních výstupů ve výzkumu PISA 2015 (tučně uvedené položky slouží ke zkoumání trendů)

Oblast	Vztahující se k přírodním vědám (Modul 4)	Obecné (Modul 10)
Já	Sebepojetí (<i>Self-concept</i>) Vnímaná zdatnost (<i>Self-efficacy</i>) Vyrovnávání se s neurčitostí a nejistotou	Akademická vnímaná zdatnost (vnímaná schopnost ovlivnit svůj školní úspěch) Obava z testů Spokojenost: životní spokojenost, spokojenost ve škole (pocit sounáležitosti se školou)
Zájem, postoje a motivace	Zájem o školní předměty Zájem o širší témata Obliba přírodních věd Instrumentální motivace	Postoje ke škole: učební aktivity a výstupy Motivace k dosahování dobrých výsledků
Přesvědčení a preference	Význam přírodních věd: obecný, osobní, vztahující se k trhu práce Prestiž povolání Oceňování vědeckých přístupů k výzkumu Epistemologická přesvědčení Starost o životní prostředí/Optimismus	Otevřenost Vytrvalost „Tah na branku“ Plánování a organizace Prokrastinace
Technologie / ICT	Vnímaná zdatnost (<i>Self-efficacy</i>) Kladný vztah k technologiím Přesvědčení týkající se technologií (přednosti a nedostatky)	Zájem o ICT Vnímaná zdatnost Autonomie v používání ICT ICT v sociálních interakcích
Chování	Četnost využívání technických zařízení	Zdraví Předčasné odchody ze vzdělávání

Zdroj: OECD 2014c

V roce 2012 byly do dotazníkových šetření zavedeny další dvě důležité inovace: korekce postojových škál prostřednictvím kotvících vinět (*vignettes*) a variantnost dotazníků.

Metoda kotvících vinět je v posledních letech ve společenskovědním výzkumu čím dál tím více využívána k očištění postojových škál od počátečního vlivu osobního nebo skupinového nastavení, kulturních odlišností a podobně na výpovědi respondentů (u nás např. Voňková;

Hrabák 2015). V mezinárodních výzkumech jsou například odpovědi na otázky týkající se sebepojetí, vztahu ke škole, k učitelům a podobně, ovlivněny kulturními odlišnostmi (některé kultury jsou optimističtější, více si důvěřují, jsou sebevědomější, jiné jsou zdrženlivější, mají tendenci se více podceňovat apod.). Z toho důvodu byly do výzkumu PISA 2012 zařazeny kotvící viněty. Jedná se o popis hypotetických situací, které žáci ohodnotí na stejné škále, na které potom odpovídají na dotazníkové otázky. Odpovědi na dotazníkové otázky jsou pak korigovány s využitím odpovědí na otázky ve vinětách, které slouží vlastně k jakémusi kalibrování postojové škály. Typicky můžeme například žákům předložit podrobný popis interakcí mezi učiteli a žáky v nějaké hypotetické škole a pak se jich zeptat, jak by na dané škále hodnotili vztahy mezi učiteli a žáky v této škole. Potom žák na stejné škále hodnotí situaci ve své škole³⁴.

V oddíle věnovaném testům bylo uvedeno, že jednotliví žáci řeší pouze část testových úloh. Tohoto neúplného testového designu je využito k tomu, aby bylo otestováno větší množství učiva/vědomostí a dovedností, než je v silách jednoho žáka v běžné testové situaci (2 hodiny testovacího času). Ve výzkumu PISA 2012 bylo poprvé využito tohoto postupu rovněž v případě dotazníků. Byly zkonstruovány 3 verze dotazníků, přičemž část otázek (např. otázky týkající se rodinného zázemí) byla ve všech sešitech identická. Část otázek se v jednotlivých sešitech lišila. Základní otázky byly v souladu s tímto designem zodpovězeny všemi žáky, odlišné moduly byly zodpovězeny dvěma třetinami žáků. Tímto způsobem bylo možno zjistit větší množství kontextových proměnných a zkoumat větší množství závislostí mezi nimi a výsledky žáků.

3.2 Konstrukce výběrů a vážení dat

3.2.1 Konstrukce výběrů

Již bylo uvedeno, že výzkumy IEA a OECD se odlišují definicí cílové populace a s tím souvisejícím způsobem výběru žáků na úrovni školy. Princip výběrů je nicméně v případech obou realizátorů obdobný. Tato kapitola popisuje, jak jsou vybírání respondenti do výzkumu PISA a na základě jakých ukazatelů je posuzována kvalita dosaženého výběru. Následně jsou uvedena specifik výběrů ve výzkumech IEA.

Ve výzkumu PISA je cílová populace definována jako „patnáctiletí žáci, kteří navštěvují minimálně 7. ročník“. To znamená, že zúčastněné země musí zahrnout do cílové populace patnáctileté žáky docházející do všech typů škol na území dané země, které vzdělávají žáky v této věkové kategorii včetně škol mezinárodních.

Přesné vymezení věku v jednotlivých zemích záleží na přesné době testování. Podle mezinárodních pravidel má testování proběhnout ve všech zúčastněných zemích v intervalu 42 dní mezi 1. březnem a 31. srpnem s tím, že se nemá jednat o prvních 6 týdnů školního roku, neboť by v tomto období žáci mohli demonstrovat nižší výkony. Mezinárodní definice cílové populace patnáctiletých byla upravena tak, aby vyhovovala struktuře žáků na severní polokouli při začátku testování v dubnu. Patří do ní žáci mezi 15 lety a 3 měsíci a 16 lety a 2 měsíci. Tedy například při testování v dubnu 2012 to byli v České republice všichni žáci, kteří se narodili v roce 1996 a navštěvovali minimálně 7. ročník školního vzdělávání v ČR. Je povolena tolerance jednoho měsíce, která umožňuje zahájit testování v březnu nebo v květnu a při výběru respondentů stále vycházet z roku narození.

³⁴ O použití vinět ve výzkumu PISA 2012 je možno se detailněji poučit na: <http://www.oecd.org/pisa/keyfindings/PISA-2012-results-Annex%20A6-VolIII-VolIV.pdf>.

V naprosté většině zemí jsou respondenti vybíráni prostřednictvím dvoustupňového stratifikovaného výběru. V prvním kroku jsou vybírány školy, které navštěvují žáci v dané věkové kategorii. Školy jsou vybírány systematicky ze seznamu všech škol navštěvovaných patnáctiletými žáky s tím, že pravděpodobnost výběru školy je úměrná odhadovanému počtu patnáctiletých žáků v dané škole (*systematic probability proportional to size sampling - PPS*).

Před vlastním výběrem jsou školy rozděleny do strat s cílem a) snížit výběrovou chybu, a tedy zvýšit reliabilitu výsledků, b) provést disproporční výběr některých geografických oblastí nebo typů škol, c) zajistit, aby se ve výběru ocitly reprezentativní výběry všech skupin populace. Při konstrukci opory pro výběr jsou používány dva typy stratifikace: implicitní a explicitní. Explicitní strata figurují při výběru samostatně, každé stratum má nezávislou oporu výběru. Typickými explicitními straty jsou regiony. Implicitní strata jsou použita pro seřazení škol v rámci explicitních strat. Zde se typicky setkáváme například s velikostí sídla školy, velikostí školy, etnickým a genderovým složením žáků a podobně. Tato stratifikace zajišťuje striktně pravděpodobnostní zastoupení v rámci všech implicitních strat. Pokud proměnné, které slouží pro implicitní stratifikaci, jsou korelovány s výsledkem testu, slouží tato stratifikace rovněž ke zvýšení přesnosti odhadů výsledků.

Již bylo zmíněno, že výběr škol v rámci jednotlivých strat z opory výběru je realizován systematicky úměrně velikosti. V opoře výběru jsou školy v jednotlivých explicitních stratech seřazeny podle implicitních strat a podle velikosti své cílové populace (*measure of size - MOS*), přičemž u každé školy je uveden akumulovaný součet. Pro každé stratum je stanoven výběrový interval, který je roven podílu celkové cílové populace ve školách obsažených ve stratu (součet *MOS* jednotlivých škol), a počtu škol, které mají být v daném explicitním stratu vybrány. Pak je stanoven náhodný začátek výběru v rámci výběrového intervalu. Škola, do které spadá tento počet žáků, je první vybranou školou. Další školy jsou vybírány analogicky opakovaným přičítáním výběrového intervalu³⁵. Každé škole jsou v rámci výběru škol přiřazeny dvě školy náhradní. Typicky to jsou školy, které jsou v opoře pro výběr zařazeny těsně před vybranou školou a těsně za vybranou školou.

Konstrukce opory pro výběr v ČR je zpravidla relativně komplikovaná, neboť patnáctiletí žáci se nacházejí na všech typech škol úrovně ISCED 2 a ISCED 3. Například v roce 2012 byly jako explicitní strata použity následující typy škol (ZŠ, víceleté gymnázium, čtyřleté gymnázium, školy poskytující střední odborné vzdělávání s maturitou, školy poskytující střední odborné vzdělávání bez maturity, základní školy praktické a praktické školy a učiliště) a regiony pro školy povinného vzdělávání (s cílem získat reprezentativní výběr za jednotlivé regiony tak, aby mohly být mezi sebou porovnávány výsledky žáků v 9. ročnících povinného vzdělávání v jednotlivých regionech). Jako implicitní strata pak sloužily regiony pro školy středního vzdělávání, velikost školy (3 kategorie) a genderové složení žáků školy (toto stratum bylo voleno z toho důvodu, že se od sebe genderové složení žáků v jednotlivých oborech středního vzdělávání hodně liší) (Palečková; Tomášek a kol. 2013).

Ve druhém kroku jsou vybíráni žáci ve školách. K výběru dochází poté, co školy odsouhlasí účast ve výzkumu a dodají seznamy patnáctiletých žáků, kteří školy navštěvují. Pro každou zemi je definována velikost skupiny žáků, kteří by měli být testováni v jednotlivých školách (*target cluster size*). Ta typicky činí 35 žáků. V případě, že počet žáků v dané škole přesáhne 35,

³⁵ S příkladem analogického algoritmu pro výzkum TIMSS je možno se podrobně seznámit na: http://timssandpirls.bc.edu/methods/pdf/Sampling_Schools.pdf.

je ze seznamu prostým náhodným výběrem vybráno 35 žáků³⁶. V případě, že ve škole je méně patnáctiletých žáků, jsou vybráni všichni.

Popsaný způsob výběru směřuje k tomu, aby všichni žáci měli co nejpodobnější pravděpodobnost výběru: větší školy (tedy žáci ve větších školách) mají větší pravděpodobnost, že budou vybrány v prvním kroku, ve druhém kroku však mají větší pravděpodobnost výběru žáci v menších školách (které jsou vybrány v prvním kroku), protože z větší i menší školy vybíráme vždy stejný počet žáků.

Ve výzkumu PISA je kvalita získaného výběru respondentů charakterizována podle míry pokrytí definované cílové populace a podle návratnosti.

Míra pokrytí cílové populace je definována prostřednictvím následujících 5 indexů:

1. Pokrytí populace možných respondentů, tedy žáků, kteří nebyli vyloučeni z testování, respondenty, kteří se účastnili šetření (po převážení – o vážení referuje oddíl 3.2.2).
2. Pokrytí populace patnáctiletých žáků v 7. a vyšším ročníku, kteří nebyli vyloučeni z testování, respondenty, kteří se účastnili šetření (po převážení).
3. Pokrytí národní populace všech patnáctiletých žáků respondenty, kteří se účastnili šetření (po převážení).
4. Pokrytí populace patnáctiletých žáků, kteří jsou ve vzdělávání, respondenty, kteří se účastnili šetření (po převážení).
5. Pokrytí populace patnáctiletých tak, jak figurovala v opoře pro výběr.

V ČR byly hodnoty těchto ukazatelů v PISA 2012 postupně 0,98, 0,98, 0,85, 0,93 a 0,97.

Jak již bylo uvedeno výše, pro posouzení kvality výběru je rozhodující návratnost na úrovni školy a na úrovni žáků. Ve výzkumu PISA jsou počítány tři indikátory:

1. Školní návratnost před zařazením náhradních škol je dána podílem původně vybraných škol, které se zúčastnily a docílily žákovské návratnosti minimálně 50%, ke všem původně vybraným školám (tedy školám uvedeným v čitateli plus vybraným školám, které se nezúčastnily, plus školám, které nedosáhly návratnosti 50 %). ČR v roce 2012 měla váženou školní návratnost před zařazením náhradních škol 98,15 %.
2. Školní návratnost po zařazení náhradních škol se počítá stejně jako index v bodě 1 s tím, že jsou do čitatele zahrnuty i náhradní školy, které dosáhly žákovské návratnosti minimálně 50 %. Vážená školní návratnost po zařazení náhradních škol činila v ČR v PISA 2012 99,61 %.
3. Žákovská návratnost je dána podílem počtu žáků, kteří byli testováni a zařazeni do výsledného souboru, vyjma žáků ze škol, ve kterých byla návratnost mezi 25 % a 50 % (školy s méně než 25% návratností nebyly do výsledného souboru zařazeny) k počtu žáků ve vybraném vzorku bez žáků vyloučených na úrovni školy. Vážená žákovská návratnost ve výzkumu PISA 2012 činila v ČR 90,07 %.

Celková návratnost je pak vypočtena jako součin školní a žákovské návratnosti.

Pro definici cílové populace ve výzkumech TIMSS a PIRLS je základem počet let školní docházky. Vzhledem ke kognitivní náročnosti testů se však organizátoři chtějí vyhnout tomu, aby testovali příliš mladé žáky. Z toho důvodu tudíž ke kritériu počet let školní docházky přidávají

³⁶ Výběr je prováděn v software, který dodává mezinárodní organizátor výzkumu, a tím pádem je i v tomto kroku zajištěn standardizovaný postup.

ještě kritérium věku. V případě, že daná země nesplní věkové kritérium, testuje žáky v následujícím ročníku. Cílové populace pro výzkumy TIMSS a PIRLS jsou tak definovány takto:

- 4. ročník (PIRLS a TIMSS): všichni žáci navštěvující ročník, který odpovídá 4. roku školní docházky od prvního ročníku úrovně ISCED 1 za předpokladu, že je průměrný věk žáků v době testování alespoň 9,5 roku.
- 8. ročník (TIMSS): všichni žáci navštěvující ročník, který odpovídá 8. roku školní docházky od prvního ročníku úrovně ISCED 1 za předpokladu, že je průměrný věk žáků v době testování alespoň 13,5 roku.

Již bylo zmíněno, že výzkumy IEA umožňují vyloučit z testování školy, *jejichž osnovy se výrazně liší od osnov hlavního vzdělávacího proudu*, což České republice dosud umožňovalo vyloučit z testování žáky v základních školách praktických. V ostatních ohledech jsou pravidla pro vylučování žáků a škol v obou výzkumech podobná. I zde platí, že podíl vyloučené populace může činit maximálně 5 %.

Výběr škol probíhá ve výzkumech identicky jako ve výzkumu PISA. Místo výběru žáků však na úrovni školy dochází k výběru celých tříd. Třídy jsou vybírány náhodně ze všech tříd daného ročníku v jednotlivých školách. Protože by příliš mnoho malých tříd mělo za následek zvyšování výběrových chyb, jsou malé třídy pro účely výběru kombinovány do tzv. pseudotříd.

Požadovaný počet škol je stanoven analogicky jako ve výzkumu PISA na základě požadované přesnosti odhadů. Vychází z toho, že by standardní chyba výsledku neměla přesáhnout hodnotu 0,035 směrodatné odchylky. Vzhledem k tomu, že výsledek je stanoven tak, aby měl průměrnou hodnotu 500 a směrodatnou odchylku 100, odpovídá tato standardní chyba pro 95% interval spolehlivosti ± 7 bodům průměrného výsledku a ± 10 bodům pro následující vlny šetření. 95% intervaly spolehlivosti pro dotazníkové proměnné by měly činit $\pm 3,5$ %. Pro většinu zemí je možno těchto hodnot dosáhnout se 150 školami a 4 000 žáky v každém testovaném ročníku. Pro průměrnou velikost třídy 27 žáků stačí k dosažené přesnosti výběr jedné třídy v každé škole ($27 \times 150 = 4\,050$ žáků). Některé země vybírají více tříd proto, aby zvýšily počet respondentů nebo aby mohly zkoumat efekty na úrovni školy.

Mezinárodní centrum může požadovat, aby daná země zvýšila počet vybraných škol v následujících situacích:

- Průměrná velikost třídy je příliš nízká na to, aby se při výběru 150 škol dosáhlo požadovaného počtu žáků.
- Předchozí kola výzkumů ukázala, že daná země není schopna dosáhnout požadované přesnosti, aniž by výběr škol zvýšila.
- Žáci jsou do tříd v rámci škol rozdělováni podle kognitivních schopností, což zvyšuje rozptyl mezi třídami a snižuje přesnost měření.
- Je očekávána nízká úroveň návratnosti.

Výběr škol pro pilotní šetření se řídí ve výzkumech IEA stejným požadavkem jako ve výzkumu PISA. Počet respondentů, kteří budou řešit každou z testových úloh, byl minimálně 200. U výzkumu PIRLS, který má 4 pilotní testové sešity, je požadována pilotáž na 800 respondentech, u výzkumu TIMSS, který má 6 pilotních testových sešitů, je požadována pilotáž na 1 200 respondentech.

Požadavky na kvalitu výsledného vzorku jsou podobné jako ve výzkumu PISA:

- minimální návratnost na úrovni školy je 85 % původně vybraných škol

- minimální návratnost na úrovni třídy je 95 % z původně vybraných tříd
- minimální návratnost na úrovni žáka je 85 % z původně vybraných a náhradních škol

Alternativou je kombinovaná návratnost na úrovni školy, třídy a žáka 75 % z původně vybraných škol.

Třídy s méně než 50% návratností jsou pro účely výpočtu návratnosti považovány za chybějící.

3.2.2 Konstrukce vah

Žáci jsou sice do výzkumů vybíráni náhodně, pravděpodobnost výběru každého z nich se ale liší. Z toho důvodu musí být respondentům přiřazeny váhy, aby bylo zajištěno, že každý respondent reprezentuje odpovídající počet žáků z cílové populace.

Váhy se pro jednotlivé žáky liší z několika důvodů:

1. V rámci výběru mohou být určité podskupiny cílové populace vybírány disproporčně (jsou zastoupeny více nebo naopak méně). Nadvýběr provádíme tehdy, když chceme tyto podskupiny podrobit speciálním analýzám, chceme se o nich dozvědět více. Tak byl například ve výzkumu PISA v ČR prováděn nadvýběr základních škol a víceletých gymnázií, aby byly doplněny na reprezentativní vzorek žáků 9. ročníku povinného vzdělávání (viz oddíl 3.2.1), nebo ve výzkumu PIAAC bylo vybráno více mladých lidí ve věku 16-29 let, než odpovídalo jejich zastoupení v populaci, aby bylo možno získat podrobnější výsledky o absolventech jednotlivých typů škol. Podvýběr je například prováděn z finančních důvodů v případě odlehklých oblastí a těžko dostupných škol.
2. Informace o škole, na jejímž základě je proveden výběr, nemusí být zcela přesná. Například je-li škola vybírána jako velká a nakonec se ukáže, že počet žáků je nízký, a jsou v ní vybíráni všichni žáci, mají tito žáci vyšší pravděpodobnost výběru, než bylo původně předpokládáno.
3. Vybraní žáci nejsou vyloučeni, ale testování se neúčastní z důvodů absence nebo odmítnutí.
4. Úprava vah tak, aby malá skupina žáků neměla příliš vysoké váhy, což může potenciálně vést k velkým standardním chybám.

Finální váha respondenta ve výzkumu PISA je součinem školní váhy, která je dána převrácenou pravděpodobností výběru školy, kterou navštěvuje respondent, a „vnitroškolní“ váhy, která je dána převrácenou pravděpodobností výběru respondenta v rámci dané školy. Váha je dále upravována třemi korekčními faktory: 1. korekce na neúčast škol v daném stratu (podíl počtu zúčastněných původně vybraných škol, 1. a 2. náhradních zúčastněných škol a nezúčastněných vybraných škol a počtu všech zúčastněných škol), 2. korekce situace, kdy byli do testování zařazeni pouze patnáctiletí žáci z určitého ročníku (např. v ČR nadvýběr ZŠ a gymnázií, který byl proveden s cílem získat reprezentativní výsledek za 9. ročník povinného vzdělávání), 3. korekce neúčasti vybraných žáků ze školy.

Na závěr jsou v některých případech poníženy váhy u žáků, kde školní nebo výsledná váha je příliš vysoká.

Školní váha je principiálně konstruována jako podíl výběrového intervalu a počtu žáků z cílové populace v dané škole. Např. měla-li daná škola 100 žáků splňujících definici cílové populace a velikost cílové populace celkem tvořila 150 000 studentů, velikost výběrového

intervalu byla při výběru 150 škol stanovena na 150 000/150, tedy 1 000. Školní váha dané školy tedy činila 1 000/100, tedy 10. To znamená, že tato škola reprezentovala 10 škol. Toto pravidlo bylo modifikováno pro velké a malé školy³⁷.

„Vnitroškolní váha“ je stejná pro všechny žáky, kteří se zúčastnili testování v dané škole, a je rovna podílu cílové populace ve škole (tedy počet patnáctiletých, kteří docházeli do školy k datu testování) a počtu respondentů, kteří se v dané škole zapojili do výzkumu.

Výběrové chyby ve výzkumech TIMSS a PIRLS jsou kombinací složek, které odrážejí pravděpodobnosti výběru a výsledky výběru na třech úrovních: na úrovni školy, třídy a žáka. Na každé úrovni sestává váha ze základní složky, která je rovna převrácené hodnotě pravděpodobnosti výběru na dané úrovni a korekci na chybějící školy, třídy a žáky.

3.2.3 Replikační váhy

V mezinárodních výzkumech výsledků vzdělávání je důležité získat co nejpresnější odhady výsledků žáků v určitých věkových kohortách v jednotlivých zemích co nejúspornějším způsobem. K výběru úsporných a přitom přesných vzorků žáků je používán stratifikovaný více-stupňový skupinkový výběr, kde jsou žáci vybíráni prostřednictvím škol. Tato metoda využívá toho, že vybírat žáky ve školách je jednoduché a přirozené a že tam také snadno mohou být testováni. Tím se stává výzkum finančně únosný. Naše měření je však zatíženo větší výběrovou chybou, než by bylo v případě, že bychom respondenty vybírali striktně náhodně z celé populace žáků, protože žáci z jednotlivých škol mají některé charakteristiky společné (jsou jim dostupné stejné školní zdroje, mají stejné kurikulum, stejné učitele apod.) a v důsledku této skutečnosti panuje ve výběrovém souboru mezi respondenty větší podobnost, než by panovala při prostém náhodném výběru. Prostý náhodný výběr 4 000 studentů zachytí rozmanitost celé populace lépe než výběr 100 náhodně vybraných škol a v každé z nich 40 studentů.

Při práci s daty bychom měli tuto skutečnost zohlednit. Použití standardního statistického software bez patřičných korekcí (například u nás ve společenských vědách hojně používaného statistického balíku SPSS) vede k nesprávným (menším) hodnotám standardních chyb vypočtených statistik. Intervaly spolehlivosti se jeví užší než odpovídá realitě, a tedy častěji nalézáme statisticky průkazné výsledky. Výpočty standardních chyb jsou totiž v běžně dostupných statistických programech založeny na předpokladu prostého náhodného výběru, kdy z množiny všech možných respondentů náhodně vybíráme respondenty, které podrobíme zkoumání.

Koncept standardní chyby je zakotven v teorii výběrů. Teorie říká, že když vybereme několik náhodných výběrů z populace, statistika založená na těchto výběrech bude mít normální rozložení. Střední hodnota této statistiky bude odpovídat hodnotě, kterou bychom získali, kdybychom provedli pozorování v celé populaci. Když pracujeme s výběry, pozorujeme určitou variabilitu v závislosti na daném výběru. Variabilitu průměru v jednotlivých výběrech označujeme jako standardní chybu.

Standardní chyba průměru je v případě prostého náhodného výběru počítána podle vzorce:

$$\sigma_{(\mu)} = \sqrt{\frac{\sigma^2}{n}}$$

³⁷ Pokud byl v nějaké škole počet žáků z cílové populace roven výběrovému intervalu nebo větší, tedy v našem případě 1 000 a více, byla tato škola vybrána s určitostí a měla školní váhu 1.

kde σ je směrodatná odchylka a n počet respondentů.

Čím větší je tedy směrodatná odchylka, tím větší je neurčitost, respektive standardní chyba. Čím vyšší je naopak počet respondentů, tím nižší je standardní chyba. Tento vzorec ovšem předpokládá prostý náhodný výběr.

Standardní chyba průměru pro skupinkový výběr je počítána podle vzorce:

$$\sigma_{(\mu)} = \sqrt{\frac{\sigma^2 \text{ mezi školami}}{n \text{ škol}} + \frac{\sigma^2 \text{ uvnitř škol}}{n \text{ škol} \cdot n \text{ žáků}}}$$

Standardní chyba je tedy v případě prostého náhodného výběru nepřímo úměrná počtu respondentů, v případě skupinkového výběru je přímo úměrná rozptylu mezi skupinkami (školy) a uvnitř skupinek (žáci ve školách) a nepřímo úměrná počtu vybraných škol a počtu vybraných žáků v jednotlivých školách³⁸.

Dopad výběrového designu na standardní chybu je často dokumentován tzv. efektem designu. Efekt designu je roven podílu rozptylu odhadu získaného z komplexního (vícestupňového skupinkového) výběru k odhadu, který by byl získán z prostého náhodného výběru stejné velikosti. Například v ČR v PISA 2012 je efekt designu pro výsledek v matematice roven 4,79. To znamená, že výběrový rozptyl (*sampling variance*) matematického výsledku je ve skutečnosti téměř 5x vyšší, než by byl za předpokladu prostého náhodného výběru se stejným počtem respondentů³⁹. Efekt designu obecně závisí na počtu škol a studentů ve výběru, na rozptylu mezi školami a na stratifikaci výběru.

Tato skutečnost má zásadní dopad na interpretaci výsledků. V jejím důsledku se jeví – například při použití SPSS za předpokladu prostého náhodného výběru – jako statisticky významné i parametry, které ve skutečnosti statisticky významné nejsou. S cílem stanovit správně hodnoty standardních chyb počítaných statistik a statistické významnosti parametrů ověřovaných modelů jsou do datových souborů přidávány tzv. replikační proměnné, které umožňují provést korekci na způsob výběru respondentů, slouží ke správnému výpočtu výběrového rozptylu a následně standardních chyb statistik, a tedy i k posouzení jejich statistické významnosti. Zároveň jsou výběry stratifikovány tak, aby byla výběrová chyba minimalizována⁴⁰.

Replikační proměnné jsou produktem metod, které umožňují určit výběrový rozptyl při zohlednění designu výběru. Tyto metody poskytují řádné odhady výběrových chyb u komplexních výběrových designů. Slouží k výpočtu standardních chyb u jednoduchých statistik a mohou být snadno adaptovány tak, aby poskytovaly správné standardní chyby i u parametrů odhadnutých metodami statistického modelování. Jsou založeny na myšlence rozdělení vzorku na řadu dílčích vzorků a využití variability těchto dílčích vzorků ke stanovení celkového výběrového rozptylu.

³⁸ Tyto vztahy zároveň nejsou lineární, tedy například u prostého náhodného výběru potřebujeme pro 2x menší chybu 4x větší počet respondentů apod.

³⁹ Vzhledem k tomu, že výběry jsou velké, jsou však výběrové chyby relativně malé i při velkém efektu designu.

⁴⁰ Například ve výzkumu PISA 2012 by nabyla standardní chyba matematického výsledku při prostém náhodném výběru hodnoty 1,301, při vícestupňovém nestratifikovaném výběru 4,492 a při stratifikaci a korekci prostřednictvím replikačních vah 2,852.

V případě statistik, které zahrnují rovněž testové výsledky, musíme kromě způsobu výběru zohlednit ještě další zdroj nepřesnosti, kterým je způsob přiřazování testových úloh jednotlivým žákům. Aby testy obsáhly dostatečně široký rozsah učiva, musí být velmi komplexní. To znamená, že není v silách jednotlivých žáků pracovat s celým testem, ale pracují jen s jeho částmi. Pro přidělení testových úloh jednotlivým žákům je využíván tzv. maticový design (*matrix-sampling assessment design*). V tomto designu jsou úlohy rozděleny do bloků a tyto bloky jsou rozděleny do určitého počtu testových sešitů, jak bylo podrobně vysvětleno v oddíle 3.1.1. Každý žák pracuje pouze s jedním sešitem. Výsledky ze všech testových sešitů jsou pak agregovány pomocí teorie odpovědi na položku do celkového výsledku. Pro každého žáka jsou odvozeny odhady testového výsledku, což teorie umožňuje přesto, že každý žák řešil jen část testu (podrobněji se této problematice věnuje následující oddíl). Zpravidla je z odvozeného rozdělení výsledků náhodně generováno 5 plausibilních hodnot (*plausible values* – PV) a jejich rozptyl slouží jako podklad pro stanovení rozptylu odhadu (*imputation variance*) (Mislevy; Beaton; Kaplan; Sheenan 1992). U statistik pracujících s testovými výsledky, tedy založených na *plausible values*, nestačí pro stanovení směrodatné chyby výběrový rozptyl, ale je nutno zahrnout i rozptyl odhadu výsledku. Efekt designu je v případě chyby měření však zpravidla nižší než efekt designu daný vícestupňovým výběrem. V některých mezinárodních výzkumech je zohledňováno i ovlivnění standardních chyb výběrem úloh, které slouží jako kotvící a propojují mezi sebou jednotlivá kola výzkumů.

Jak již bylo uvedeno, díky komplexnímu výběrovému designu používanému v mezinárodních výzkumech výsledků vzdělávání nemohou být standardní chyby počítány podle jednoduchého vzorce založeného na předpokladu prostého náhodného výběru. Ke korekci standardních chyb na skupinkový výběr jsou používány replikační metody, které se sice principiálně shodují, ale liší se v podrobnostech postupu i v proměnných, které jsou pro replikační algoritmy zařazeny do datových souborů. Replikační metody používají rozdíly mezi replikáty, tedy dílčími opakovanými výběry ze stávajícího výběru, a celým výběrem k tomu, aby odhadly rozptyl. Prakticky se to realizuje tak, že jsou systematicky zmenšovány nebo zcela eliminovány příspěvky částí výběru. Pokud systematicky odstraňujeme část výběru, spočítáme dílčí statistiky, srovnáme je se statistikou získanou z celého výběru a zjistíme, že mezi nimi nejsou žádné nebo jsou jen velmi malé rozdíly, znamená to, že výběr je relativně homogenní a že byl vybrán z relativně homogenní populace. Tedy očekáváme, že jiné výběry z té samé populace by byly velmi podobné tomu, který máme. Výběrová chyba je malá. Naopak pokud systematicky eliminujeme část výběru, spočítáme dílčí statistiky, srovnáme je se statistikou získanou z celého výběru a shledáme relativně velké rozdíly, pak předpokládáme, že náš výběr je velmi různorodý a je vybrán z relativně různorodé populace. Očekáváme, že další výběr z té samé populace by mohl vést k odlišným výsledkům. Tedy v tomto případě máme velkou standardní chybu a vysokou neurčitost.

Ve výzkumech IEA je používána replikační technika *jackknife* (*JRR – jackknife repeated replication*) (Johnson; Rust 1992). Tato metoda má dvě alternativy *Jackknife 1* a *Jackknife 2*. *Jackknife 1* je používána pro nestratifikované výběry a spočívá v systematickém odstraňování pozorování z celého výběru a úpravě příspěvků zbylých pozorování. Je například využívána ve výzkumech vědomostí a dovedností dospělých, kde respondenti nejsou vybíráni přes instituce, ale přímo z dospělé populace. Ve výzkumech IEA se používá metoda *Jackknife 2*, která je určena pro stratifikované, respektive vícestupňové výběry. Tato metoda předpokládá, že primární jednotky výběru (*primary sampling units PSU*), v našem případě školy, mohou být spárovány způsobem odpovídajícím výběrovému designu tak, že každý pár představuje pro účely výpočtu standardní chyby jakési pseudostratum. Technika *JRR* pak umožňuje správně

určit příspěvek rozptylu uvnitř těchto jednotek a mezi těmito jednotkami k výběrovému rozptylu. Spočívá v systematickém přiřazení dvojic škol k replikačním zónám (*replication zones*) a tvorbě pseudoreplikací původního vzorku prostřednictvím postupné systematické modifikace vah jednotlivých škol v replikačních zónách. Tvoříme tedy modifikované výběry z původního vzorku, přičemž jejich počet je roven počtu replikačních zón. Statistika, kterou počítáme, se pak vypočte pro původní výběr a pro každý pseudoreplikovaný výběr (tedy s původní sadou vah a se všemi sadami vah vytvořenými replikačním procesem). Rozptyl těchto odhadů pak určuje výběrovou chybu určené statistiky.

Přiřazení škol k výběrovým zónám je v datových souborech TIMSS a PIRLS prováděno v mezinárodním centru. Typicky je první a druhá škola přiřazena k první replikační zóně, třetí a čtvrtá ke druhé replikační zóně apod. Při počtu 150 škol (zamýšlený počet škol ve výzkumech IEA) je vytvořeno 75 replikačních zón. Pokud je škol více a je vytvořeno více párů, jsou sloučeny do 75 zón tak, aby v každé zóně byly 2 „kvaziškoly“. V rámci každé výběrové zóny je náhodně jedné škole přidělen indikátor $u_j = 0$ a druhé škole $u_j = 1$. Při tvorbě pseudoreplikačního výběru je pak váha školy s $u_j = 1$ zdvojnásobena a váha školy s $u_j = 0$ vynulována. Náhodné generování indikátorů je rovněž prováděno v mezinárodním centru a hodnoty indikátorů pro jednotlivé školy jsou uloženy do datových souborů jednotlivých zemí spolu s proměnnou udávající číslo replikační zóny. Tím je zajištěno, že při jednotlivých výpočtech nebude již docházet k náhodnému generování a standardní chyby budou nabývat při opakovaných výpočtech zcela identických hodnot.

Výběrový rozptyl se tedy počítá podle vzorce:

$$Var_{Jrr}(t) = \sum_{h=1}^H [t(J_h) - t(S)]^2$$

kde

H je počet výběrových zón, obvykle 75,

$t(S)$ je stanovovaná statistika pro celý výběr,

$t(J_h)$ je stanovovaná statistika pro jednotlivé pseudoreplikační výběry.

Jak již bylo uvedeno, statistiky jsou stanovovány s využitím výběrových vah, přičemž u replikačních výběrů je vždy pro danou výběrovou zónu u školy s indikátorem $u_j = 1$ váha dané školy zdvojnásobena a druhá škola je z výpočtu eliminována. Pro stanovení výběrového rozptylu je tedy pro každou statistiku nutno provést každý výpočet 76krát: jednou s využitím původních vah a potom 75krát za účelem určení statistiky pro každý pseudoreplikační výběr.

Ve výzkumu PISA jsou aplikovány stejné principy, nicméně použité metody jsou poněkud sofistikovanější, s cílem zvýšit přesnost standardních chyb. Pro výpočet výběrového rozptylu je použita metoda *Balance Replicate Replication* (BRR), respektive její varianta zvaná Fayova metoda. Její podstata je stejná jako metoda Jackknife, ale má oproti ní některé výhody, které jsou dobře zdokumentovány v odborné literatuře (Judkins 1990). Zde se klasty v dílčích replikacích neodstraňují, jako je tomu v metodě Jackknife 2, ale jejich příspěvek je systematicky zmenšován a zvyšován.

Ve výzkumu PISA byla metoda aplikována následovně:

Rozdělením škol do dvojic, respektive do trojic (v případě lichého počtu škol) byly vytvořeny replikační zóny. V replikačních zónách byly školy náhodně očíslovány 1, 2 a (v případě trojice škol) 3. Zón bylo vytvořeno celkem 80. Při tvorbě pseudoreplikačních výběrů byla jedna ze škol vynásobena číslem 1,5 a druhá číslem 0,5, v případě 3 škol jedna škola koeficientem 1,7071 a dvě školy koeficientem 0,6464 anebo dvě školy faktor 0,2929 a jedna škola 1,3536.

Rozptyl byl vypočítán podle vzorce:

$$Var_{BRR}(t) = 0,05 \sum_{h=1}^{80} [t(J_h) - t(S)]^2$$

Váhy pro každý pseudoreplikační výběr jsou přepočítávány tak, aby zohlednily chybějící respondenty (*non-response adjustments*). Proto ve výzkumu PISA na rozdíl od studií IEA nejsou součástí datového souboru replikační zóna a koeficient, ze kterých se váhy postupně počítají, ale přímo výsledně sady replikačních vah. Tedy součástí datového souboru pro každou zúčastněnou zemi je sada 80 proměnných udávajících hodnoty replikačních vah. Tyto váhy díky opravám na chybějící respondenty nemají zpravidla hodnoty 1,5 a 0,5, jak bylo uvedeno v základním algoritmu, ale nabývají mnoha rozmanitých hodnot.

Specializované programy pro práci s daty s vícestupňovým skupinkovým výběrem provádějí při výpočtu vybraných statistik výše popisované algoritmy s využitím replikačních proměnných obsažených v datových souborech a poskytují tak nezkreslené hodnoty standardních chyb a indikátorů statistické významnosti vypočtených statistik. Speciálně pro práci s daty z mezinárodních výzkumů výsledků vzdělávání IEA a OECD byl vyvinut software IDB analyser, který v případě vybraných statistik provádí korekci na výběrovou chybu a využívá k tomu replikační proměnné uvedené v datových souborech (tj. v případě výzkumů IEA informaci o zóně a indikátoru, v případě výzkumu PISA přímo replikační váhy). Práci s tímto softwarem a jeho možnostmi se podrobněji věnuje oddíl 3.4.

K výpočtu výběrových standardních chyb v případě mezinárodních výzkumů výsledků vzdělávání je používán rovněž za tím účelem speciálně vyvinutý software WesVar.

Výše bylo uvedeno, že kromě výběrové chyby se uplatňuje ještě chyba měření. Tou se zabývá následující kapitola. Program IDB analyser umožňuje stanovit kromě chyby výběru i chybu měření a obě chyby kombinovat.

Často potřebujeme rozhodnout, zda jsou dvě vypočtené statistiky statisticky odlišné, tedy potřebujeme vypočítat standardní chybu rozdílu mezi těmito dvěma statistikami. K tomu potřebujeme vydělit rozdíl standardní chybou rozdílu a porovnat s kritickou hodnotou. Při výpočtu standardní chyby rozdílu musíme rozlišovat mezi situací, kdy máme závislé výběry, a situací, kdy pracujeme s výběry nezávislými. V nezávislých výběrech, které jsou vybrány z jiných opor (například jiné země nebo jedna země v různých letech), nejsou chyby korelovány, v závislých výběrech (například muži a ženy) ano. V případě nezávislých výběrů spočteme chybu rozdílu jako odmocninu ze součtu kvadrátů standardních chyb obou výběrů. V případě závislých výběrů musíme vzít v úvahu jejich závislost, tedy nemůžeme jednoduše kombinovat dílčí standardní chyby, ale musíme replikovat rozdíly mezi statistikami a použít rozdíly v těchto replikacích k vypočtení standardních chyb. Situace se dále komplikuje, pokud počítáme s kognitivními skóry vyjádřenými plausibilními hodnotami. Musíme zohlednit i chybu měření (viz níže). V řadě statistických balíčků můžeme významnost rozdílů stanovit prostřednictvím regresní analýzy (viz například Gonzales 2014).

3.3 Postupy škálování a prezentace výsledků

3.3.1 Teorie odpovědi na položku

Při škálování dat není v mezinárodních výzkumech využívána klasická testová teorie, se kterou pracují tvůrci didaktických testů v ČR, ale modernější teorie odpovědi na položku (*Item Response Theory - IRT*). IRT modeluje pravděpodobnost správného zodpovězení testové položky v závislosti na obtížnosti položky a úrovni vědomostí a dovedností respondenta (dále zdatnost), přičemž tato zdatnost zde není vázána na konkrétní test, ale charakterizuje obecnou (latentní) zdatnost respondenta v měřené oblasti. Parametry modelů IRT jsou nezávislé na výběru respondentů i na daném testu, zatímco v klasické testové teorii je skór (*true score*) udávající výsledek respondenta definován v rámci konkrétního testu. IRT poskytuje podstatně větší flexibilitu například v situacích, kdy se testy administrují různým skupinám respondentů, které se od sebe hodně odlišují. IRT tak tvoří dobrý základ pro počítačové adaptivní testování⁴¹.

Již bylo uvedeno, že mezinárodní výzkumy využívají při tvorbě testů maticový design, při kterém různé skupiny respondentů řeší různé testové sešity, aby bylo možno výzkumem obsáhnout širší rozsah učiva. Při zpracování výsledků tedy využíváme toho, že modely IRT umožňují srovnávání výsledků různých testů ověřujících tutéž zdatnost. Velkou výhodou modelů IRT je také skutečnost, že hodnotí obtížnosti položek a zdatnosti respondentů na stejné škále. To umožňuje smysluplné porovnávání výkonů respondentů a obtížnosti položek, tedy ukázat, co umějí osoby, které demonstrovaly určitou úroveň zdatnosti.

Nejjednodušším IRT modelem je jednoparametrický logistický model. Říká se mu také Raschův model, podle dánského matematika George Rasche, který o něm pojednal ve své knize již v roce 1960 (Rasch 1960). Pravděpodobnost správné odpovědi na položku je v Raschově modelu definována jako funkce jedné proměnné – zdatnosti respondenta, a jediného parametru – obtížnosti dané testové položky (b). Obtížnost je definována jako úroveň zdatnosti, při které respondent zodpoví položku správně právě s poloviční pravděpodobností.

Dvouparametrický logistický model přidává k parametru obtížnosti ještě parametr citlivosti položky (a). Ten popisuje sklon charakteristické funkce položky v bodě obtížnosti (b). Odhad parametru citlivosti je blízký nule, pokud položka špatně rozlišuje mezi lepšími a slabšími respondenty. V případě, kdy slabší respondenti odpovídají na položku lépe než respondenti s celkově lepším výsledkem (vyšší zdatností), je citlivost položky záporná. Dvouparametrický logistický model používáme v případě, kdy správné odpovědi na testové položky nejsou snadno uhodnutelné. V případě položek s výběrem jedné správné z nabízených odpovědí, lze ale předpokládat, že i zcela neznalí studenti správnou odpověď alespoň s pravděpodobností $1/n$ (kde n je počet nabízených odpovědí) uhodnou. V takovém případě má opodstatnění tříparametrický logistický model, v němž třetí parametr (c) vyjadřuje pravděpodobnost toho, že i zcela neznalý student odpoví na položku správně.

Pravděpodobnost, že respondent, jehož zdatnost na dané škále je charakterizována latentní proměnnou, zodpoví správně položku i , je tedy:

$$p_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}}$$

⁴¹ Český čtenář se může o metodologii IRT poučit v publikaci Urbánek; Denglerová; Širůček 2011.

kde

θ je zdatnost respondenta na dané škále

a_i je směrnice položky i , která charakterizuje její diskriminační schopnosti/citlivost

b_i je parametr položky i , který udává její obtížnost

c_i je parametr položky i , který udává možnosti respondentů s velmi nízkou zdatností uhádnout správnou odpověď

Ve výzkumech IEA je využíván tříparametrický model pro vyhodnocování položek s výběrem odpovědi a dvouparametrický model pro vyhodnocování položek, ve kterých žáci tvoří své vlastní odpovědi, které jsou vyhodnocovány jako buď správné, nebo nesprávné. Speciální (*partial credit*) model je používán pro položky, které mohou být vyhodnocovány i jako částečně správné. Výzkum PISA užívá modifikovaný jednoparametrický model.

Při zpracování výsledků dochází nejprve ke stanovení parametrů testových položek odděleně pro každou stanovovanou škálu, následně je s využitím těchto parametrů stanovena pravděpodobnostní funkce pro zdatnost jednotlivých respondentů.

Parametry testových položek jsou zpravidla stanovovány na výběru respondentů ze všech zúčastněných zemí (např. ve výzkumu PISA je to 500 náhodně vybraných respondentů z každé země).

V mezinárodních výzkumech je prováděna kalibrace testových položek i na národní úrovni, aby bylo ověřeno, zda položky fungují ve všech zemích řádně. Například je kontrolováno, zda se nějaká položka nechová v nějaké zemi nestandardně (například je proti ostatním položkám významně obtížnější nebo naopak snazší, což může být způsobeno špatným překladem), dále je zkoumána korelace jednotlivých distraktorů s celkovým skórem (*point-biserial*) a její ekvivalent pro otevřené úlohy.

Pro modely IRT platí, že nemají žádné přirozené stupnice, tedy na libovolně zvolené škále získané lineární transformací mohou být stanoveny matematicky ekvivalentní, leč odlišné hodnoty parametrů. Tato neurčitost je řešena tak, že jsou stupnice stanoveny prostřednictvím hodnoty průměru (500) a směrodatné odchylky (100). Takto byla stanovena škála pro první výzkum, od kterého jsou sledovány trendy (v případě výzkumů IEA to byly výzkumy TIMSS 1995 a PIRLS 2001).

Při kalibraci je ještě nutno rozhodnout, jak budou posuzovány chybějící odpovědi v testových sešitech. Obecně rozlišujeme mezi chybějící odpovědí a neřešenou úlohou. Úlohy, pro které platí, že není zodpovězena ani předchozí ani žádná z následujících úloh, jsou pro účely stanovení parametrů testových položek považovány za neřešené, což znamená, že při škálování s nimi zacházíme tak, jakoby nebyly administrovány, tedy jakoby je testový sešit daného žáka neobsahoval. Při stanovování zdatnosti daného žáka jsou však hodnoceny jako nesprávná odpověď.

3.3.2 Vícenásobná imputace

Každé měření vědomostí a dovedností je zatíženo chybou. Chyba může být ve všech případech redukována navýšením počtu testových položek, které respondent řeší. Již bylo opakovaně uvedeno, že ve výzkumech, které zjišťují vědomosti a dovednosti nikoliv u jednotlivců, ale u větších skupin respondentů, je používán maticový design, ve kterém různé skupiny žáků řeší různé testové sešity. Každý žák řeší relativně malé množství úloh, ale agregací přes všechny testované žáky získáme široké zastoupení ověřovaného učiva. Tato výhoda je ovšem u mezinárodních výzkumů vykoupena nemožností určit přesně zdatnost jednotlivých respondentů.

To v tomto případě ovšem nevadí, neboť cílem šetření je zjistit výsledky větších skupin žáků, nikoliv jednotlivců. Aby neurčitost spojená se stanovením zdatnosti jednotlivých respondentů v podobných výzkumech neovlivnila negativně odhad parametrů celé testované populace, byla vyvinuta metoda odhadu tzv. plausibilních hodnot (*plausible values*).

Při aplikaci metody plausibilních hodnot v mezinárodních výzkumech nedochází k tomu, že se nejprve určí zdatnosti jednotlivých respondentů a ty jsou potom agregovány přes celou testovanou populaci. Místo toho jsou využita všechna shromážděná data – odpovědi na testové položky i odpovědi z žákovských dotazníků ke stanovení charakteristik žákovské populace a dílčích subpopulací. Na základě těchto charakteristik jsou z odhadnutých rozdělení zdatnosti vygenerovány sady imputovaných skóre, které jsou označovány jako plausibilní hodnoty. Tyto plausibilní hodnoty jsou pak užívány k prezentaci výsledků a ke standardním statistickým výpočtům.

Tento postup byl vyvinut ve druhé polovině minulého století ke zpracování výsledků pravidelného hodnocení znalostí amerických žáků *National Assessment of Education Progress (NAEP)*. Vychází z Bayesovské statistiky. Je založen na představě, že zdatnost respondentů není díky své latentní povaze známa ani pro studenty, kteří se zúčastnili testování. Tuto zdatnost aproximujeme očekávanou hodnotou odhadnutou na základě informací, které jsou nám známy, tedy z proměnných, které jsme získali v našem šetření. Zdatnost jednotlivých respondentů tak aproximujeme náhodnými hodnotami z jejich rozdělení zdatnosti, které je stanoveno na základě odpovědi respondentů na testové položky a údajů o jejich rodinném zázemí, a z charakteristik testových položek. Tyto náhodné hodnoty označujeme jako imputace (*imputations*) nebo plausibilní hodnoty. Zpravidla používáme náhodných hodnot několik (typicky 5), abychom mohli určit chybu imputace.

Plausibilní hodnoty nemají význam testových skóre jednotlivých žáků, kteří se zúčastnili šetření. Jedná se o imputované skóre žáků s podobnými odpověďmi na testové otázky a s osobnostními charakteristikami, jako mají zúčastnění žáci. Pokud je model správně specifikován, slouží tyto hodnoty ke správnému stanovení hodnot pro cílovou populaci. Plausibilní hodnoty jsou přiděleny všem žákům bez ohledu na konkrétní testové položky, které řešili. Tak například ve výzkumu PISA mají žáci testové skóre i v oblastech, ve kterých vůbec nebyli testováni (například žák, který řešil pouze úlohy z matematiky a přírodovědných předmětů, má stanoveny plausibilní hodnoty i pro oblast čtenářské gramotnosti).

Jak bylo uvedeno výše, při práci s testovými výsledky potřebujeme stanovit nejen chybu výběrovou, ale také chybu měření neboli chybu odhadu výsledku. K tomu potřebujeme spočítat rozptyl odhadu výsledku, a to tak, že příslušnou statistiku počítáme pro jednotlivé plausibilní hodnoty a výsledný rozptyl stanovíme podle vzorce:

$$Var_{imp} = \left(1 + \frac{1}{M}\right) Var(t_i, \dots \dots t_m)$$

kde

M je počet plausibilních hodnot,

t_{1-m} jsou příslušné statistiky spočítané pro jednotlivé PV.

Výsledná standardní chyba je pak vypočtena z celkového rozptylu, který zahrnuje výběrový rozptyl i rozptyl odhadu:

$$Var(t_{pv}) = Var_{jrr}(t_1) Var_{imp}$$

kde výběrový rozptyl $Var_{jrr}(t_1)$ je stanoven pro první plausibilní hodnotu.

Statistické výpočty zahrnující testové skóry by tedy měly být správně provedeny 5krát odděleně pro každou plausibilní hodnotu, výsledky zprůměrovány a stanoven rozptyl dílčích hodnot.

3.3.3 Měření trendů

Jak již bylo opakovaně uvedeno, cílem mezinárodních výzkumů, které probíhají cyklicky, je sledovat vývoj vědomostí a dovedností v jednotlivých oblastech v čase. K tomu je užitečné, aby bylo možno výsledky z jednotlivých let přímo porovnávat. Škálování dat tedy probíhá tak, aby bylo možné výsledky z jednotlivých kol šetření vyjadřovat na stejné škále.

Škálování dat z mezinárodních výzkumů má čtyři kroky:

- a) kalibrace testových položek (odhad parametrů modelu pro každou položku);
- b) zpracování dotazníkových dat pro každého žáka metodou hlavních komponent tak, aby mohly být využity při odhadu rozdělení zdatnosti (ve výzkumech IEA jsou do výpočtu zařazeny komponenty, které vysvětlují 90 % rozptylu⁴²; některé proměnné, např. pohlaví, vstupují do regrese přímo);
- c) generování skórů pro jednotlivé žáky (*plausible values*);
- d) vyjádření těchto skórů na škále z předcházejících měření tak, aby bylo možno sledovat trendy.

Škálování probíhá nejen na celkových škálách – pro matematiku, přírodovědné předměty a čtenářské dovednosti, ale též na dílčích škálách pro jednotlivé obsahové celky a dovednosti.

Škála, na které jsou porovnávány trendy ve výzkumech IEA, byla původně stanovena pro TIMSS 1995 a PIRLS 2001 tak, že přiřadila mezinárodní průměr hodnotě 500 a směrodatnou odchylku hodnotě 100. Možnost prezentovat výsledky z následujících kol výzkumů na stejné škále jako ve výzkumech předcházejících je zajišťována prostřednictvím souběžného škálování dat z každého nového kola výzkumu s daty z kola předcházejícího a posazení výsledků na škálu z předchozích kol prostřednictvím lineární transformace. V souběžném škálování jsou parametry testových položek odhadovány ze současného i minulého hodnocení, přičemž některé položky jsou v obou hodnoceních stejné. Je tedy možné odhadnout rozdělení zdatnosti pro obě hodnocení. Rozdíl mezi těmito dvěma rozděleními reprezentuje změnu ve výsledcích mezi oběma koly.

V dalším kroku je určena lineární transformace, která transformuje rozdělení dat z předchozího kola získané souběžnou kalibrací na původní škálu a nakonec je tato transformace aplikována na data z posledního kola. Tím jsou aktuálně získaná data prezentována na trendové škále.

Ve výzkumu PISA je postup analogický:

1. Jsou stanoveny parametry úloh v aktuálním výzkumu.
2. Je spočtena transformace, která převádí parametry na škálu z předcházejícího kola.
3. Je spočteno rozdělení zdatností s využitím parametrů z aktuálního kola.

⁴² Ve výzkumu PISA jsou do regrese zařazeny položky, které vysvětlují 95 % rozptylu.

4. Zdatnosti jsou převedeny na škálu z předchozího kola pomocí transformace z bodu 2.

Ve výzkumu PISA je přímá porovnatelnost škál od prvního kola v roce 2000 možná pouze pro čtenářskou gramotnost. Matematická a přírodovědná gramotnost byly v tomto roce vedlejšími testovanými doménami, a omezený rozsah ověřovaných vědomostí a dovedností v roce 2000 nedovoloval přímé propojení na pozdější kola, v nichž měly obě oblasti daleko větší záběr. Pro matematiku tak existuje návaznost od roku 2003 a pro přírodovědné předměty od roku 2006.

Je třeba si uvědomit, že měření rozdílů ve výsledcích mezi jednotlivými koly je ovlivněno výběrem úloh, které slouží jako kotvící úlohy, tedy úloh, na základě jejichž parametrů je prováděna transformace. Chyba, která vzniká výběrem kotvících úloh, je v PISE stanovována jako tzv. *linking error*. V PISA je dále prováděna ještě korekce na efekt testového sešitu. Bloky s identickými úlohami jsou v některých testových sešitech umístěny v první polovině testového sešitu a v jiných sešitech ve druhé polovině. Úspěšnost úloh není identická, žáci řeší úlohy odlišně podle toho, jestli se nacházejí na začátku nebo na konci testového sešitu.

3.3.4 Úrovně způsobilosti

Výše uvedené postupy stanovují žákovské skóre na škále 0–1 000 s tím, že většina hodnot se nachází v rozmezí 300 až 700. Platí totiž, že v normálním rozdělení se 95 % hodnot nachází v rozmezí průměr plus minus 2 směrodatné odchylky. Tyto hodnoty ukazují, jakých výsledků dosahují žáci v jednotlivých zemích ve srovnání s žáky v ostatních zemích a jak se mění výsledky žáků v jednotlivých zemích v čase. Neříkají však nic o tom, co umějí žáci, kteří dosáhli určitého výsledku.

Informaci o tom, co umějí žáci, kteří dosáhli určitých skóre, získáváme přiřazením skóre do určitých úrovní způsobilosti, ve kterých jsou popsány vědomosti a dovednosti žáků, kteří těchto úrovní dosáhli. Výzkumy OECD tyto úrovně způsobilosti označují jako *proficiency levels* a výzkumy IEA jako *international benchmarks*. Ve výzkumech TIMSS a PIRLS jsou stanoveny 4 úrovně způsobilosti. Popisy úrovní způsobilosti vznikají tak, že jsou vybrány typické úlohy, které řešili správně žáci, kteří se svými výsledky umístili na těchto úrovních. Následně skupina expertů popisuje vědomosti a dovednosti, které potřebuje žák k tomu, aby správně vyřešil tyto typické úlohy. Hranice pro správné vyřešení je zde stanovena na 65 %. To znamená, že do popisu vědomostí a dovedností odpovídajících dané úrovni jsou zahrnuty úlohy, které žáci, již se umístili na dané úrovni, vyřešili správně ve dvou třetinách případů. Úlohy, které řešili s méně než 50% pravděpodobností, jsou zařazeny na úroveň vyšší.

Ve výzkumu PISA je používán analogický postup s tím, že je vymezeno 6 úrovní způsobilosti. Nejdříve je provedena identifikace možných škál. Experti na jednotlivé hodnocené oblasti se domlouvají, na jakých dílčích škálách budou prezentovány výsledky. Toto rozhodnutí se řídí nejen statistickými parametry (ukazujícími, které testové položky mají k sobě blíže, tedy měří podobné vědomosti a dovednosti), ale zejména expertním rozhodnutím, jaké aspekty jednotlivých oblastí má smysl vypíchnout. Ve výzkumu PISA byly definovány tři škály v každé z hlavních hodnocených oblastí. V oblasti matematické gramotnosti byly vymezeny škály *formulace situací matematicky*, *užívání matematických konceptů*, *fakt*, *postupů a zdůvodňování* a *interpretování*, *aplikování* a *hodnocení matematických výstupů*; ve čtenářské gramotnosti *získávání informací*, *zpracování informací* a *posouzení textu*; v přírodovědné gramotnosti *identifikování přírodovědných problémů*, *vysvětlování jevů vědeckým způsobem*, *používání vědeckých důkazů*.

Při specifikaci úrovní způsobilosti byly testové položky přiřazeny jednotlivým škálám na základě konceptuálního rámce a charakteru úloh a umístěny podle úrovně jejich obtížnosti. V dalším kroku pak experti vypracovali popis dovedností odpovídajících jednotlivým škálám a jejich vývoje při postupu k vyšším úrovním. Škály a jejich popisy tedy vznikaly jako kombinace statistické analýzy a expertní analýzy. První návrhy byly vždy vytvořeny na základě dat z pilotáže. Tyto návrhy byly následně revidovány a zpřesněny na základě dat z hlavního šetření.

Na škálách, které popisují vývoj dovedností, neexistují žádné přirozené dělicí body, vývoj je kontinuální. Pro účely prezentace výsledků je však vhodné škály rozdělit do několika pásem a tato pásma charakterizovat vědomostmi a dovednostmi, jež mají respondenti, kteří se svými výkony v těchto pásmech umístili. Ve výzkumu PISA bylo rozdělení provedeno tak, aby studenti měli nadpoloviční pravděpodobnost, že správně vyřeší všechny úlohy nacházející se na úrovni, do které byli přiřazeni. Student, který se umístil na spodní hranici dané úrovně, měl 52% pravděpodobnost, že úlohy vyřeší správně, student na horní hranici měl tuto pravděpodobnost 62%.

Pro ilustraci jsou v tabulkách 18 a 19 uvedeny popisy úrovní způsobilosti z matematické gramotnosti a řešení problémových úloh z výzkumu PISA 2012.

Tabulka 18: Úrovně způsobilosti pro oblast matematické gramotnosti, PISA 2012 (Palečková a Tomášek 2013)

Úroveň	Rozmezí	Podíl žáků schopných provádět úkoly na této úrovni nebo vyšší (průměr OECD)	Co žáci zvládnou
1	358 až méně než 420 bodů	92 %	Na úrovni 1 žáci umějí odpovědět na otázky ze známého kontextu, pokud otázky obsahují všechny relevantní údaje a jsou jednoznačně definovány. Jsou schopni najít informace a provést rutinní postupy podle přesných instrukcí v explicitních situacích. Umějí realizovat činnosti, které jsou nasnadě a přímo plynou ze zadání.
2	420 až méně než 482 bodů	77 %	Na úrovni 2 žáci umějí interpretovat a poznat situace v kontextech, které nevyžadují víc než přímé úsudky. Umějí vybrat podstatné informace z jednoho zdroje a využívají jednu formu reprezentace. Žáci na této úrovni umějí používat základní algoritmy, vzorce, postupy a konvence. Jsou schopni přímé dedukce a umějí doslovně interpretovat výsledky.
3	482 až méně než 545 bodů	54,5 %	Na úrovni 3 žáci umějí realizovat jasně definované postupy, a to včetně těch, které vyžadují sekvenční rozhodování. Umějí zvolit a aplikovat jednoduché řešitelské strategie. Žáci na této úrovni umějí interpretovat a využívat data pocházející z různých zdrojů informací a vyvozovat z nich závěry. Umějí krátce sdělit své interpretace, výsledky a dedukce.

4	545 až méně než 607 bodů	30,8 %	Na úrovni 4 žáci umějí efektivně pracovat s explicitními modely komplexních konkrétních situací, které mohou obsahovat omezující podmínky nebo vyžadovat vyslovení předpokladů. Umějí zvolit a integrovat různé reprezentace včetně symbolických a umějí je přiřadit k prvkům situací z reálného světa. Žáci na této úrovni využívají rozvinuté dovednosti a umějí v kontextech z reálného světa flexibilně uvažovat, někdy dokonce proniknou hluboko do situace. Umějí zformulovat a sdělovat vysvětlení i argumenty. Vycházejí při tom z vlastních interpretací, argumentace a činnosti.
5	607 až méně než 669 bodů	12,6 %	Na úrovni 5 žáci umějí vytvářet modely komplexních situací a s těmito modely dále pracovat, určit omezující podmínky a formulovat předpoklady. Umějí vybírat, porovnávat a vyhodnotit strategie řešení vhodné pro práci s komplexními úlohami, které z modelů vyplývají. Na této úrovni žáci umějí postupovat strategicky, využívat bohaté a rozvinuté myšlení a uvažování, vhodné navzájem propojené reprezentace a symbolické i formální charakteristiky situací a vzhled do nich. Umějí reflektovat své jednání a formulovat i sdělovat své interpretace a závěry.
6	669 bodů a více	3,3 %	Na úrovni 6 jsou žáci schopni konceptualizovat, zobecnit a použít informace, které získali vlastním zkoumáním a modelováním komplexní problémové situace. Jsou schopni propojit různé zdroje informací i různé reprezentace a jsou schopni flexibilně překládat z jedné formy reprezentace do druhé. Žáci ovládají pokročilé matematické myšlení a uvažování. Žáci jsou schopni využít vzhled a porozumění, ovládají symbolické i formální matematické operace a vztahy. To vše využívají k vytváření nových přístupů a strategií pro řešení nových situací. Žáci jsou na této úrovni schopni formulovat, jak postupují. Umějí reflektovat svá zjištění, výsledky, interpretace, argumenty a posoudit vhodnost těchto výsledků z hlediska původní situace.

Tabulka 19: Úrovně způsobilosti pro oblast řešení problémů, PISA 2012 (ČŠI 2014)

Úroveň	Rozmezí	Podíl žáků schopných provádět úkoly na této nebo vyšší úrovni (průměr OECD)	Co žáci zvládnou
1	358 až méně než 423 bodů	91,8 %	Na úrovni 1 jsou žáci schopni prozkoumat strukturu problému pouze omezeně a snaží se o to jen tehdy, pokud se s velmi podobnými situacemi setkali již dříve. Na základě vlastních pozorování známých situací jsou tito žáci schopni pouze částečně popsat ovládání jednoduchého, denně používaného přístroje. Umějí řešit velmi jednoduché problémy za předpokladu, že má být překonána jedna překážka a k dosažení cíle je zapotřebí provést pouze jeden nebo dva kroky. Nejsou schopni dopředu plánovat nebo stanovovat dílčí cíle.
2	423 až méně než 488 bodů	78,6 %	Na úrovni 2 jsou žáci schopni prozkoumat strukturu neznámého problému a částečně ho pochopit. Snaží se porozumět elektronickým přístrojům s neznámými funkcemi, jako jsou domácí spotřebiče a prodejní automaty, a ovládat je, ovšem pouze s částečným úspěchem. Dokážou otestovat jednoduchou hypotézu a dovedou vyřešit problém, který má jednu konkrétní překážku. Dovedou naplánovat a provést pouze jeden krok naráz k dosažení dílčího cíle, avšak mají jistou schopnost sledovat celkový postup řešení.
3	488 až méně než 553 bodů	56,6 %	Na úrovni 3 jsou žáci schopni zacházet s informacemi, které jsou jim předloženy v několika různých formátech, prozkoumat strukturu problému a rozpoznat jednoduché vztahy mezi jeho součástmi. Dovedou ovládat jednoduchá elektronická zařízení, ale složitější přístroje jim činí potíže. Dobře si poradí s jednou překážkou a umí navrhnout několik řešení a ověřit, jestli danou překážku opravdu překonali. Pokud je překážek víc nebo pokud jsou některé funkce vzájemně propojené, dovedou používáním jedné proměnné zjistit účinek na ostatní proměnné nebo funkce. Umí navrhnout a provést zkoušky, které potvrdí nebo vyvrátí danou hypotézu. Chápou, že musí plánovat dopředu a sledovat pokrok a jsou schopni v případě potřeby vyzkoušet i jiné možnosti.

4	553 až méně než 618 bodů	31,0 %	Na úrovni 4 jsou žáci schopni pozorně prozkoumat středně složitý problém. Pochopí vztahy mezi součástmi podstatnými pro řešení problému. Dovedou ovládat středně složitě elektronické přístroje, jako jsou neznámé prodejní automaty nebo domácí spotřebiče, ale nedělají to vždy naprosto rutinně. Umí plánovat několik kroků dopředu a sledovat pokrok svého řešení. Obvykle jsou schopni na základě zpětné vazby tyto plány přizpůsobit nebo přeformulovat dílčí cíle. Dovedou systematicky zkoušet různé možnosti a kontrolovat, jestli bylo překonáno více překážek najednou. Umí zformulovat hypotézu, proč systém nefunguje správně, a popsat, jak to otestovat.
5	618 až méně než 683 bodů	11,4 %	Na úrovni 5 jsou žáci schopni systematicky prozkoumat složitý problém, aby pochopili strukturu důležitých vztahů a informací. Když se setkají s neznámými, středně složitými přístroji, jako jsou prodejní automaty nebo domácí spotřebiče, jsou schopni se je rychle naučit ovládat. Při hledání nejlepších strategií, jak dosáhnout cíle, umí přemýšlet dopředu a mají na paměti všechna daná omezení. Když se setkají s nečekanými obtížemi nebo když udělají chybu ve správném postupu, dovedou okamžitě přizpůsobit své plány nebo zpětně vysledovat chybu.
6	683 bodů a více	2,5 %	Na úrovni 6 jsou žáci schopni vytvořit úplný, ucelený a srozumitelný model struktury jakéhokoli problému, což jim je umožňuje efektivně řešit. Dovedou prozkoumat strukturu problému velmi efektivně a pochopit všechny jeho podstatné vztahy. Informace jim mohou být předkládány v různých formátech, dokonce i v těch, které vyžadují rozdílnou interpretaci a začlenění do příslušných částí. Když se setkají s velmi složitými přístroji, jako jsou domácí spotřebiče, které fungují neobvyklým nebo nečekaným způsobem, rychle se je optimální cestou naučí ovládat. Umí formulovat obecné hypotézy o systému a dokážou je řádně otestovat. Umí dovést předpoklad k logickému závěru a poznají, když k vyvození závěru nemají dostatek informací. Aby došli k řešení, dovedou vytvářet komplexní, pružné, vícefázové plány, které v průběhu řešení neustále sledují. V případě potřeby změní strategii, přičemž zohlední veškerá omezení, a to jak zjevná, tak skrytá.

3.4 Práce s daty

3.4.1 Datové soubory

České datové soubory ve formátu *.sav (soubor pro statistický balík SPSS) jsou k dispozici na stránkách České školní inspekce v sekci věnované mezinárodním výzkumům. Jsou zde spolu s českými verzemi dotazníků. Nalezneme tu žákovské soubory, které obsahují proměnné ze žákovských dotazníků, a plausibilní hodnoty, které určují výsledky ve všech testovaných oblastech (včetně dílčích škál). Ve výzkumu PISA je k dispozici ještě školní datový soubor, ve výzkumech IEA nadto ještě učitelský a v některých případech i rodičovský soubor. Na mezinárodních stránkách výzkumů jsou umístěny rovněž *Codebooks*, které specifikují hodnoty jednotlivých proměnných obsažených v mezinárodních datových souborech.

České datové soubory obsahují zpravidla kromě mezinárodních proměnných ještě několik proměnných národních, které byly získány pouze v ČR. Například v žákovském souboru se typicky jedná o proměnné označující typ školy (ve výzkumu PISA všechny typy škol navštěvovaných patnáctiletými žáky – viz tabulka 20, ve výzkumu TIMSS školy/třídy s rozšířenou výukou nějakého předmětu vs. běžné základní školy). Soubory dále obsahují i známky z hlavních předmětů, jež měli žáci na vysvědčení.

Chceme-li se poučit o tom, jak řešili žáci jednotlivé úlohy, musíme vyhledat na mezinárodních stránkách výzkumů datový soubor, který obsahuje odpovědi žáků na jednotlivé testové úlohy. Na mezinárodních stránkách nalezneme rovněž výpisy četností jednotlivých odpovědí pro všechny testové a dotazníkové položky (ve výzkumu PISA jsou označovány jako *Compendia*, ve výzkumech IEA jako *Almanacs*). Jednoduché porovnávání úspěšnosti v jednotlivých úlohách tedy můžeme provádět s využitím těchto výpisů, aniž bychom pracovali s datovými soubory.

Na následující straně je uvedena ukázka z PIRLS 2011 z dokumentu P11_LIT_ItemAlmanac⁴³, který obsahuje data o kognitivních úlohách pro všechny zúčastněné země. Zde se jedná o úlohu s výběrem správné odpovědi ze čtyř nabízených možností. V prvním sloupci je informace o počtu žáků, kteří v jednotlivých zemích úlohu řešili, ve druhém sloupci je informace o obtížnosti úlohy. Následují procentuální podíly žáků, kteří volili jednotlivé nabízené možnosti, dále podíl žáků, kteří úlohu neřešili, a nakonec procentuální úspěšnost dívek a chlapců. Zde vidíme, že se jednalo o úlohu s vysokou úspěšností správného řešení.

Pokud chceme provádět srovnání s jinými zeměmi, musíme si z mezinárodních stránek stáhnout soubory vybraných zemí. Na stránkách výzkumu PISA jsou datové soubory obsahující data ze všech zúčastněných zemí ve formátu *.txt, a syntaxe pro načtení dat z těchto souborů do statistických balíků SPSS a SAS. Na stránkách výzkumů TIMSS a PIRLS je možno si stáhnout datové soubory jednotlivých zemí ve formátu SPSS nebo SAS.

Chceme-li analyzovat dohromady školní a žákovské proměnné, musíme soubory spojit. Spojení provádíme přes identifikátor školy. Žákovské a rodičovské soubory je možno spojit přes identifikátor žáka.

⁴³ Viz <http://timssandpirls.bc.edu/pirls2011/international-database.html>.

COUNTRY	N	DIFF %	A %	B %	C %	D %	OMITTED %	1.GIRL % RIGHT	2.B.OY % RIGHT
Australia	1201	91.4	91.4	1.2	1.6	5.5	0.3	94.4	88.3
Austria	921	95.8	95.8	0.3	0.3	3.7	0.0	95.8	95.8
Azerbaijan	957	91.8	91.8	1.7	0.5	5.9	0.1	93.7	90.2
Belgium (French)	733	87.2	87.2	1.4	0.6	10.5	0.2	87.3	87.0
Bulgaria	1047	93.9	93.9	0.9	0.7	4.3	0.1	92.1	95.7
Canada	4568	93.6	93.6	1.2	0.7	4.4	0.1	94.5	92.8
Chinese Taipei	856	95.1	95.1	0.4	0.2	4.2	0.1	96.8	92.6
Colombia	753	91.5	91.5	1.3	2.0	14.9	0.3	91.1	91.9
Croatia	923	97.8	97.8	0.1	0.1	1.9	0.1	99.6	95.8
Czech Republic	894	94.9	94.9	0.1	0.5	4.4	0.0	94.9	95.0
Denmark	908	94.5	94.5	0.3	0.3	4.9	0.0	95.3	93.6
England	793	91.4	91.4	1.0	1.5	6.1	0.0	93.7	89.3
Finland	931	96.0	96.0	0.3	0.6	2.9	0.2	97.5	94.7
France	898	92.8	92.8	0.6	1.5	4.7	0.4	92.8	92.8
Georgia	944	92.6	92.6	1.3	1.0	4.9	0.2	93.8	91.6
Germany	759	94.9	94.9	0.4	0.0	4.2	0.4	95.0	94.3
Hong Kong SAR	779	96.7	96.7	0.7	0.0	2.3	0.2	96.5	97.0
Hungary	1021	91.4	91.4	0.5	1.4	6.3	0.3	91.2	91.6
Indonesia	937	82.1	82.1	1.0	1.1	15.3	0.5	87.4	76.6
Iran, Islamic Rep. of	1143	84.8	84.8	2.2	2.6	9.9	0.5	85.9	83.7
Ireland	894	93.2	93.2	0.5	0.3	6.0	0.0	95.1	91.1
Israel	830	94.9	94.9	0.9	0.6	3.5	0.1	95.1	94.7
Italy	815	96.1	96.1	0.1	0.1	3.6	0.1	94.8	97.2
Lithuania	923	93.4	93.4	0.9	0.0	5.6	0.1	94.4	92.4
Malta	702	89.9	89.9	2.3	2.2	11.5	0.1	86.1	81.7
Morocco	1512	52.1	52.1	12.0	8.2	26.5	1.2	56.9	48.0
Netherlands	794	94.5	94.5	0.8	0.9	3.8	0.0	94.3	94.6
New Zealand	1117	90.5	90.5	1.1	1.5	6.7	0.1	93.2	88.1
Northern Ireland	708	95.8	95.8	1.0	0.3	3.0	0.0	97.6	94.0
Norway	617	90.3	90.3	0.1	1.1	8.3	0.1	92.7	87.4
Oman	2064	72.2	72.2	5.5	5.0	16.6	0.6	79.7	64.7
Poland	955	90.3	90.3	0.4	0.8	8.3	0.2	92.1	87.7
Portugal	811	93.2	93.2	0.8	1.0	4.9	0.1	96.0	90.5
Qatar	803	70.8	70.8	5.7	3.4	15.8	0.3	72.8	69.0
Romania	930	88.0	88.0	1.2	0.6	8.9	1.3	91.7	84.5
Russian Federation	883	98.6	98.6	0.0	0.5	1.0	0.0	99.1	98.0
Saudi Arabia	870	73.0	73.0	2.1	3.4	21.1	0.5	80.6	64.9
Singapore	1252	92.2	92.2	1.6	1.8	4.2	0.1	93.9	90.7
Slovak Republic	1110	90.5	90.5	0.9	0.7	7.9	0.1	92.4	88.7
Slovenia	900	94.0	94.0	0.4	0.4	5.2	0.0	95.2	92.9
Spain	1716	86.3	86.3	1.3	1.4	10.3	0.1	85.6	82.9
Sweden	916	93.6	93.6	0.5	0.6	4.2	1.1	94.1	93.3
Trinidad and Tobago	771	81.3	81.3	2.2	3.1	13.4	0.0	85.1	77.8
United Arab Emirates	2895	73.5	73.5	5.2	4.3	16.7	0.3	77.7	69.2
United States	2537	89.7	89.7	1.4	1.3	7.4	0.2	90.5	88.9
International Avg. (45)	50401	89.2	89.2	1.5	1.4	7.8	0.2	90.9	87.5
Botswana (6)	839	57.3	57.3	4.5	7.2	30.7	0.3	59.2	55.4
Honduras (6)	773	81.1	81.1	1.8	2.1	14.6	0.5	79.0	83.4
Kuwait (6)	623	63.9	63.9	9.1	5.6	20.8	0.6	68.3	58.2
Morocco (6)	1399	74.6	74.6	3.5	2.2	18.8	1.0	79.3	70.1
Alberta, Canada	732	92.4	92.4	1.5	1.1	5.0	0.0	92.8	92.0
Ontario, Canada	857	93.7	93.7	0.7	0.8	4.7	0.1	95.0	92.5
Quebec, Canada	850	92.4	92.4	2.3	0.6	4.5	0.2	94.2	90.6
Maltese - Malta	709	82.8	82.8	2.7	1.4	11.8	0.3	87.0	80.5
Eng/Špa. (5) - RSA	711	65.4	65.4	6.6	6.5	20.0	1.5	69.2	62.6
Andalusia, Spain	864	87.1	87.1	0.5	0.6	11.4	0.3	90.3	84.0
Abu Dhabi, UAE	818	70.8	70.8	5.4	5.2	18.4	0.1	76.6	64.8
Dubai, UAE	1194	81.1	81.1	3.1	2.6	12.3	0.9	81.9	80.4
Florida, US	524	91.3	91.3	1.2	1.3	6.2	0.0	93.4	89.3

3.4.2 Použití žákovských vah

V předcházejících kapitolách byly zmíněny vybrané metodologické aspekty mezinárodních výzkumů výsledků vzdělávání, které je třeba zohlednit při práci s žákovskými daty. Prvním aspektem, který je třeba zohlednit, je složení výběru respondentů. K tomu používáme žákovské váhy, kterými se podrobně zabýval oddíl 3.2. Žákovské váhy například zohledňují skutečnost, že někteří respondenti jsou ve výběru záměrně zastoupeni méně a jiní více, než odpovídá jejich výskytu v populaci, a že někteří respondenti, kteří byli zařazeni do výběru, se šetření nezúčastnili. Váhy jsou koncipovány tak, že jejich součet představuje velikost testované populace patnáctiletých žáků, tedy každý žák zastupuje určitý podíl žáků cílové populace.

Ve výzkumu PISA jsou cílovou populací patnáctiletí žáci. Jak již bylo uvedeno, výběr je relativně komplikovaný a je tudíž skutečně důležité data správně vážit. V tabulce 20 je podíl žáků ve výběru PISA 2012 nevážený a převážený studentskou váhou W_FSTUWT.

Z tabulky je zřejmé, že v případě, že data nebudou zvážena, budou ve výběru reprezentováni žáci víceletých gymnázií výrazně větší měrou, než odpovídá jejich zastoupení v populaci patnáctiletých. To se samozřejmě odrazí v provedených výpočtech. Například průměr první plausibilní hodnoty v matematice pro nevážený soubor činí 519,77 a pro vážený soubor pouze

498,79. To ve výsledcích odpovídá rozdílu mezi Finskem a Českou republikou. Z této ilustrace je zřejmé, že chyba, které bychom se dopustili, kdybychom váhy nepoužili, je značná.

Ve výzkumu TIMSS 2011, který probíhal ve 4. ročníku ZŠ, byl rozdíl mezi váženým (žakovská váha TOTWGT) a neváženým průměrem pro první plausibilní hodnotu v matematice nižší: 514,80 pro nevážený soubor a 510,66 pro soubor vážený.

Tabulka 20: Podíl žáků z jednotlivých typů škol – PISA 2012

Typ školy	neváženo (%)	váženo (%)
Základní škola	40,4	47,1
Gymnázium víceleté	27,4	12,6
Gymnázium čtyřleté	6,4	6,7
Střední odborné studium s maturitou	17,5	22,4
Střední odborné studium bez maturity	6,6	8,4
Škola praktická, speciální	1,7	2,8

Jak již bylo uvedeno výše, žakovská váha váží data na cílovou populaci. V datových souborech IEA je obsažena také váha, která váží data na počet respondentů (*house weight* HOUWGT).

Když budeme realizovat mezinárodní analýzy s žakovskou váhou, budou výsledky více ovlivněny zeměmi, které mají větší počet studentů. Tak například analýzy evropských zemí by byly nejvíce ovlivněny výsledky německých, francouzských a britských žáků. Kdybychom počítali s *house weight*, byly by výsledky ovlivněny nestejným počtem respondentů v jednotlivých zemích. Pokud chceme, aby v mezinárodních analýzách byly všechny země zastoupeny stejně, volíme místo žakovské váhy váhu *senátní* (*senat weight* - v PISE proměnná *senwgt_STU*, v TIMSS proměnná *SENWGT*), která zachovává poměry mezi jednotlivými žakovskými skupinami v rámci každé země a je zároveň koncipována tak, aby populace každé ze zemí představovala stejný počet žáků (v PISA 1000 žáků, v TIMSS 500 žáků), tedy aby všechny země přispívaly k výsledkům stejnou měrou.

3.4.3 Práce s plausibilními hodnotami a replikačními váhami

Již bylo zmíněno, že při práci s daty z mezinárodních souborů hrozí nebezpečí špatného odhadu standardních chyb a z toho plynoucího chybného stanovení statistické významnosti výsledků. Chyby se dopouštíme tím, že a) nepracujeme správně s plausibilními hodnotami a b) že předpokládáme prostý náhodný výběr v situaci, kdy data byla získána prostřednictvím vícestupňového výběru, kde je mezi respondenty větší podobnost.

Již bylo řečeno, že provádíme-li analýzy s plausibilními hodnotami, měli bychom je provést pro každou plausibilní hodnotu zvlášť, výsledky zprůměrovat a standardní chybu stanovit z rozptylu jednotlivých hodnot. Zároveň bychom měli všechny výpočty realizovat s využitím replikačních vah. Ty jsou ve výzkumech OECD přímo obsaženy v datových souborech, ve výzkumech IEA jsou v datových souborech koeficienty, jejichž prostřednictvím je možno replikační váhy vygenerovat.

Pro usnadnění práce s plausibilními hodnotami a replikačními váhami je uživateli k dispozici program IDB analyser, který je možno si zdarma stáhnout na stránkách mezinárodních

výzkumů⁴⁴. Tento program se používá v kombinaci se statistickým balíkem SPSS k provádění jednoduchých analýz datových souborů ze všech mezinárodních výzkumů výsledků vzdělávání v žákovské i dospělé populaci (včetně datových souborů z výzkumů občanské výchovy CIVED a ICCS a výzkumů informačních technologií SITES a ICILS). Zároveň umožňuje analyzovat i data z výzkumů učitelů TEDS a TALIS. IDB analyser umožňuje počítat průměry, procentuální podíly, korelace, zastoupení respondentů na jednotlivých úrovních způsobilosti a percentilové hodnoty a umožňuje provádět rovněž lineární a nově i logistickou regresi. Po navolení příslušného datového souboru, typu analýzy a proměnných vygeneruje syntaxi pro SPSS. Program IDB analyser obsahuje kromě analytického modulu, který provádí výše zmiňované analýzy, rovněž modul pro spojování souborů. Ten umožňuje například konstrukci mezinárodních souborů s vybranými zeměmi, ve kterých budou obsaženy jen proměnné vybrané pro určitý typ analýzy.

Pro provádění složitějších analýz již musí uživatelé vyvinout vlastní syntaxi nebo použít jiný statistický software, například SAS, STATA, R, MPlus a podobně.

Následující tabulka obsahuje průměrné výsledky žen a mužů v jednotlivých oblastech, zjišťovaných ve výzkumu vědomostí a dovedností dospělých PIAAC, a standardní chyby těchto průměrných výsledků. Pro ilustraci jsou uvedeny také hodnoty získané v software SPSS při prostém zprůměrování plausibilních hodnot a ignorování skupinkové povahy dat.

Z tabulky 21 je zřejmé, že průměry jsou stejné, ale standardní chyby jsou v případě, že nezohledníme povahu dat, 6-7x nižší. To vede k chybnému stanovení statistické významnosti v případě rozdílů v oblasti řešení problémů. Při řádném výpočtu se jeví rozdíly mezi muži a ženami vzhledem k relativně velké chybě jako statisticky nevýznamné; nicméně výpočet v SPSS, pokud nezohledňuje povahu dat, indikuje statistickou významnost.

Tabulka 21: Průměrné výsledky mužů a žen a standardní chyby při různých způsobech výpočtu, PIAAC 2012

oblast		SPSS		IDB	
		průměr	s.e.	průměr	s.e.
čtenářská gramotnost	žena	272,32	0,02	272,32	1,30
	muž	275,68	0,02	275,68	1,26
numerická gramotnost	žena	271,19	0,02	271,19	1,30
	muž	280,20	0,02	280,20	1,36
řešení problémů	žena	280,91	0,03	280,91	1,49
	muž	284,87	0,03	284,87	1,72

Při zpracování dat z mezinárodních výzkumů výsledků vzdělávání je často využíván software HLM (*hierarchical linear modeling*), který byl vyvinut pro práci s daty, jež mají hierarchickou povahu. Data pořizovaná v mezinárodních výzkumech tuto hierarchickou povahu mají. Pracujeme zde s výsledky žáků, kteří jsou umístěni ve třídách. Tyto třídy se nacházejí ve školách v různých vzdělávacích systémech. Proměnné, které se vztahují k práci učitele, jsou společné všem žákům v jedné třídě, proměnné charakterizující školu a její organizaci jsou společné všem žákům v jedné škole. Stejně tak proměnné charakterizující vzdělávací systém jsou společné všem

⁴⁴ <http://www.iea.nl/eula.html>

žákům tohoto systému. Hierarchické lineární (nebo též víceúrovňové) modelování umožňuje testovat modely, které mají více úrovní a do kterých vstupují proměnné na té úrovni, která jim odpovídá (více např. Soukup 2006). Tyto modely nám například umožňují testovat, do jaké míry jsou výsledky žáka ovlivněny jeho socioekonomickým statutem (na úrovni žáka), a do jaké míry jsou ovlivněny složením žáků třídy nebo školy (socioekonomický status agregovaný na úroveň třídy nebo školy), postoji, výukovými metodami učitelů (které jsou stejné pro celou třídu) a podobně. Víceúrovňové modelování lze samozřejmě realizovat v řadě dalších prostředích (SPSS, SAS, R, Stata apod.). Software HLM umožňuje při výpočtech data řádně navázat a řádně pracovat s plausibilními hodnotami jako závislými proměnnými (tedy vysvětlovat výsledky žáků). Zároveň logicky zohledňuje způsob výběru. Jednoduché příklady víceúrovňového modelování uvádíme v kapitole věnované výsledkům mezinárodních výzkumů v ČR.

V posledních letech získává mezi uživateli na oblibě software Mplus, který umožňuje v uživatelsky přívětivém prostředí řádně provádět složitější analýzy, jako je například víceúrovňové strukturní modelování nebo logistická regrese. Některé jeho přednosti uvádíme v kapitole 4 v oddíle zabývajícím se vývojem analytických metod.

Logistickou regresi využíváme v případě, že chceme vysvětlit závislou proměnnou, která nabývá diskrétních hodnot (více např. Řeháková 2000). Nejčastěji se setkáváme s binární logistickou regresí, ve které nabývá závislá proměnná dvou hodnot. Logistickou regresi například využíváme při studiu vzdělanostních přechodů, kdy zkoumáme, jak je skutečnost, že žák aspiruje na nějaký typ vzdělání, nebo že byl přijat ke vzdělávání určitého typu, ovlivněna například jeho rodinným zázemím, školními výsledky nebo výsledky testu. I zde se setkáváme s tím, že při řádné aplikaci replikačních vah v programu MPlus se jeví nevýznamné některé nezávislé proměnné, které při jednoduchém výpočtu v SPSS, bez zohlednění povahy dat, statistickou významnost vykazovaly, případně se jevíly významné interakce mezi nimi.

Příklad takové situace ukazují modely, jejichž parametry jsou uvedeny v následující tabulce 22. Modely odhadují šance přechodu z 5. třídy do víceletého gymnázia v závislosti na pohlaví dítěte, průměru známek v 5. ročníku, nejvyšším dosaženém vzdělání obou rodičů a výsledku v testu z matematiky, který dítě řešilo v závěru 4. ročníku. Kromě výše uvedených proměnných vstupují do modelu ještě interakce mezi vzděláním rodičů a výsledkem v matematickém testu, a mezi pohlavím a školními známkami. Tyto interakce ověřují hypotézy, že šance na přechod do víceletého gymnázia se posilují v případě chytrých dětí vzdělaných rodičů a v případě pilných dívek. Pro výpočty byla využita data z longitudinálního výzkumu CLoSE (*Czech Longitudinal Study of Education*), který sledoval děti, jež se v roce 2011 zúčastnily testování v rámci mezinárodního výzkumu TIMSS & PIRLS 2011, jež bylo realizováno na přechodu na druhý stupeň základní školy, respektive do osmiletých gymnázií.

První model (levý sloupec tabulky 22) nezohledňuje způsob výběru respondentů a provádí výpočet tak, jako by se jednalo o prostý náhodný výběr. Z tabulky je zřejmé, že všechny proměnné i interakce jsou statisticky významné na hladině 0,05. To by znamenalo, že děti s lepšími známkami mají statisticky vyšší šance na úspěšné přijetí do víceletého gymnázia a že ostatní proměnné se pozitivně projevují jen ve vzájemné interakci. Tedy že pilná děvčata mají větší šance na přijetí do víceletého gymnázia (stejně jako chytré děti vzdělaných rodičů), ale že samostatně mají dívky nižší šance – stejně jako děti s lepším matematickým výsledkem a děti vzdělaných rodičů.

Druhý model (průměrný sloupec) zohledňuje výběr respondentů. Zohlednění má za následek nárůst standardních chyb a zcela odlišný pohled na statistické významnosti, kde se jeví nevýznamné nejenom obě interakce, ale také pohlaví a výsledek v matematickém testu. Tento

výpočet napovídá, že výše uvedená interpretace není správná. Při odebrání interakcí se stane statisticky významným výsledek v matematickém testu, pohlaví však nikoli. V pravém sloupci tabulky je uveden výsledný model, který říká, že šance na přijetí ke studiu víceletého gymnázia se zvyšují s prospěchem na základní škole, matematickými dovednostmi (zde slouží jako aproximace studijních předpokladů) a vzděláním rodičů. Výsledný model tedy ukazuje, že i při zohlednění studijních předpokladů a školního prospěchu žáků mají děti ze vzdělaných rodin větší šanci na studium ve víceletém gymnáziu.

Uvedený příklad dobře ilustruje chybu, jaké se můžeme dopustit, nezohledníme-li při výpočtu skutečnost, že respondenti nebyli vybráni pomocí prostého náhodného výběru. Vzhledem k tomu, že se jedná o velké soubory (tisíce respondentů), jeví se řada koeficientů jako statisticky významná, i když ve skutečnosti významná není, což může vést ke špatným interpretacím výsledků.

Tabulka 22: Šance na studium na víceletém gymnáziu – koeficienty modelů bez použití a s použitím replikačních vah

	výpočet bez replikačních vah s interakcemi			výpočet s replikačními váhami a interakcemi			finalní výpočet s replikačními váhami		
	B	s.e.	sig.	B	s.e.	sig.	B	s.e.	sig.
pohlaví (dívka)	-0,973	0,069	0,000	-1,362	0,969	0,160			
průměr známek	0,562	0,069	0,000	0,732	0,152	0,000	0,466	0,066	0,000
vzdělání rodičů	-0,345	0,056	0,000	-0,616	0,312	0,049	0,262	0,045	0,000
matematický skór	-0,002	0,001	0,005	-0,004	0,004	0,273	0,222	0,041	0,000
vzdělání rodičů x matematický skór	0,001	0,000	0,000	0,002	0,001	0,101			
pohlaví (dívka) x průměr známek	0,229	0,009	0,000	0,325	0,20	0,104			
konstanta	3,304	0,126	0,000	3,809	1,789	0,033	6,578	0,929	0,000

3.4.4 Další nástroje umožňující jednoduchou práci s daty

Pro jednoduchý pohled na data z výzkumu PISA 2012 byly na stránkách výzkumu PISA k dispozici interaktivní nástroje: *Interactive data selection* a *Multi-dimensional Data Request*. První z nich umožňoval interaktivně vytvářet dvourozměrné tabulky, které ukazovaly četnosti odpovědí na vybrané otázky a u každé odpovědi ukázaly průměrný výsledek v hlavních testovaných oblastech (čtenářská, matematická a přírodovědná gramotnost). Informace byly poskytnuty pro země vybrané uživatelem a zároveň obsahovaly stejné informace o průměru zemí OECD. Výsledky bylo možno si stáhnout do excelového souboru. Níže je ukázán print screen obrazovky s výstupem, kde zadání tvořila následující otázka týkající se učebních strategií. Výstup byl požadován pro ČR a pro Německo.

Vyber jedno tvrzení, které nejlépe vystihuje tvůj způsob práce v matematice:

- Když se učím na prověrku z matematiky, snažím se přijít na to, kterým pojmem jsem ještě dobře neporozuměla.
- Když se učím na prověrku z matematiky, snažím se přijít na nové možnosti řešení.
- Když se učím na prověrku z matematiky, ověřuji si, zda si pamatuji to, co jsem se už naučila.

Pod odpověďmi obou vybraných zemí jsou ještě odpovědi zemí OECD. V položce *OECD average* jsou průměry zemí OECD, v položce *OECD total* jsou průměrné odpovědi za předpokladu, že každá země přispívá k průměru úměrně svému počtu patnáctiletých žáků. Tabulka například ukazuje, že pamětné učení se matematice je mezi patnáctiletými žáky ČR rozšířenější než mezi německými žáky i mezi průměrnými žáky zemí OECD.

Downloadable table in CSV / Excel format: 243731.CSV

Variable:	ST53Q02	Learning Strategies- Improve Understanding vs. New Ways vs. Memory
	Full question:	For each group of three items, please choose the item that best describes your approach to mathematics
Category:	1	Improve understanding
	2	new ways
	3	check memory

					Reading		Mathematics		Science	
Country	Variable	Category	%	(%SE)	Mean	(SE)	Mean	(SE)	Mean	(SE)
Czech Republic	ST53Q02	1	39.77	(0.94)	506	(3.72)	509	(3.77)	518	(3.39)
Czech Republic	ST53Q02	2	15.49	(0.79)	491	(5.44)	516	(5.48)	516	(5.20)
Czech Republic	ST53Q02	3	42.53	(0.92)	497	(3.29)	500	(3.51)	509	(3.17)
Czech Republic	ST53Q02	m	2.21	(0.34)	447	(16.92)	469	(15.09)	472	(14.38)
Germany	ST53Q02	1	64.92	(1.01)	523	(3.04)	526	(3.15)	539	(3.33)
Germany	ST53Q02	2	11.82	(0.58)	487	(6.44)	505	(6.99)	513	(6.46)
Germany	ST53Q02	3	20.36	(0.80)	530	(5.01)	528	(5.17)	543	(4.94)
Germany	ST53Q02	m	2.90	(0.39)	447	(14.90)	473	(14.58)	479	(14.68)
OECD Total	ST53Q02	1	57.04	(0.30)	507	(1.32)	496	(1.39)	507	(1.44)
OECD Total	ST53Q02	2	15.12	(0.22)	470	(1.81)	472	(1.74)	477	(1.64)
OECD Total	ST53Q02	3	25.71	(0.26)	494	(1.85)	484	(1.62)	495	(1.72)
OECD Total	ST53Q02	m	2.12	(0.08)	410	(4.55)	413	(3.82)	419	(4.28)
OECD Average	ST53Q02	1	54.61	(0.16)	509	(0.57)	504	(0.57)	512	(0.56)
OECD Average	ST53Q02	2	14.41	(0.11)	474	(1.03)	484	(1.02)	486	(1.01)
OECD Average	ST53Q02	3	28.47	(0.14)	497	(0.70)	490	(0.67)	499	(0.68)
OECD Average	ST53Q02	m	2.51	(0.05)	406	(2.16)	420	(1.89)	422	(2.03)

a - The category does not apply in the country concerned. Data therefore missing

c - There are too few observations to provide reliable estimates

m - Missing or invalid response category (this category is excluded from computations of group percentages in the international report)

OECD Average - the average of the valid percentages and mean performance of OECD countries.

OECD Total - (OECD as single entity) - each country contributes in proportion to the number of 15-year-olds enrolled in its schools.

Ve druhém nástroji *Multi-dimensional Data Request* mohl uživatel zvolit až čtyři otázky ze školního nebo žákovského dotazníku a zobrazit vztahy mezi těmito proměnnými a výsledky žáků ve vícerozměrné tabulce. Požadavek byl po odeslání zařazen do fronty, kde čekal na zpracování. Výsledky obdržel uživatel na emailovou adresu, kterou zadal do systému.

Výzkum vědomostí a dovedností dospělých OECD PIAAC nabídl aplikaci *International Data Explorer*,⁴⁵ která umožňuje tvorbu sofistikovaných vícerozměrných tabulek interaktivně.

Aplikace *International Data Explorer* je v současné době k dispozici na stránkách americké organizace NCES (*National Center for Education Statistics*) nejen pro výzkum PIAAC, ale rovněž pro výzkumy PISA, TIMSS a PIRLS⁴⁶. Kromě kognitivních výsledků obsahuje systém dotazníkové proměnné týkající se rodinného zázemí, postojů k učení a vzdělávání a školního vzdělávání (PISA, PIRLS, and TIMSS), proměnné z učitelských dotazníků týkající se vyučovacích metod, pedagogických přesvědčení, výukových zdrojů, praxe a počátečního a dalšího vzdělávání (PIRLS, TIMSS) a proměnné ze školních dotazníků týkající se školy, uplatňovaných postupů a školních zdrojů (PISA, PIRLS, and TIMSS).

Na stránkách OECD je dále zúčastněným zemím k dispozici celá sada indikátorů, které byly získány z mezinárodních výzkumů PISA, PIAAC a TALIS, včetně indikátorů, které vstupovaly do ročenek *Education et a Glance*. Tyto indikátory si lze stáhnout v excelových souborech⁴⁷.

Tyto pomůcky reagují na situaci, kdy běžně statisticky poučený uživatel nemůže s daty z mezinárodních výzkumů řádně pracovat bez dalšího poučení. Organizátoři výzkumů se tedy snaží poskytnout nástroje, které umožní hledat odpovědi na běžně kladené otázky bez vlastní práce s daty. Řada uživatelů by nicméně potřebovala informace, které by jim poskytly patřičnou nadstavbu nad jejich statistické znalosti. V ČR je možno se pro poučení o záležitostech dat a správné práci s nimi obrátit na publikace Petra Soukupa (např. Soukup 2016).

⁴⁵ <http://piaacdataexplorer.oecd.org/ide/idepiaac/>

⁴⁶ <https://nces.ed.gov/surveys/international/ide/>

⁴⁷ <http://gpseducation.oecd.org/IndicatorExplorer>

4. Vývoj pohledů na cíle, možnosti a využití výzkumů

Tato kapitola reflektuje, jak se za dobu realizace změnily aspekty mezinárodních výzkumů, jak se změnilo využití získaných dat a jaký je potenciál práce s daty do budoucna. První část kapitoly popisuje, jak se vyvíjely metody konstrukce testu a škálování, jak ovlivnil možnosti a realitu výzkumů technologický rozvoj a jak se měnily pohledy na měřené cíle. Druhá část kapitoly se věnuje využití informací, které jsou v rámci mezinárodních výzkumů shromážděny. Popisuje využití dat a hlavní typy získaných poznatků a specificky se zaměřuje na možnosti využití dat ve výzkumech efektivity vzdělávání. Podává přehled typických analytických postupů a jejich zdokonalování. V závěru podává přehled využití dat českými výzkumníky.

4.1 Ohlédnutí za metodologickým vývojem

4.1.1 Vývoj v oblasti konstrukce testu a dotazníku

Historie mezinárodních výzkumů výsledků vzdělávání velmi dobře ilustruje vývoj oboru označovaného jako *large scale assessments (LSAs)*, pro který je typickou úlohou charakterizovat nějakou populaci respondentů z hlediska jejich vědomostí a dovedností. Tyto výzkumy se tedy zaměřují na zjišťování skupinových výsledků. Jak již bylo uvedeno dříve, je hodně důležité je odlišovat od plošných testů (*large-scale testing programmes*), které jsou naopak zaměřeny na hodnocení jednotlivců. Vzhledem k tomu, že oba typy měření mají zcela jiné cíle a jejich výsledky jsou užívány ke zcela jiným účelům, liší se významně používanými metodologickými přístupy. Některé postupy se však překrývají, a proto vývoj v jedné oblasti je často využíván i v oblasti druhé.

Na vývoji mezinárodních výzkumů IEA a OECD je možno dobře pozorovat metodologický vývoj, který proběhl v oblasti měření vědomostí a dovedností v posledních 50 letech. V mezinárodních výzkumech jsou uplatňovány nejnovější metody a zároveň je věnována velká pozornost kvalitě realizace v jednotlivých zemích. V návaznosti na moderní poznatky se postupně standardizovaly metody výběru respondentů, výpočtů vah a výpočtů výběrových chyb tak, jak jsou podrobně popsány v kapitole 3.

Velký vývoj nastal v oblasti konstrukce testů. V průběhu druhé poloviny 20. století začaly být do testů kromě úloh s výběrem odpovědi (*multiple-choice items*) zařazovány rostoucí měrou rovněž úlohy, ve kterých respondenti tvoří vlastní odpovědi (*open-ended items*). Byly vyvinuty precizní metody hodnocení otevřených úloh a školení hodnotitelů, jakož i metody zajišťování mezinárodní srovnatelnosti. Do konstrukce testu a vymezení měřených oblastí silně zasahuje rovněž vývoj informačních technologií a změny náhledů na cíle vzdělávání. Těmito aspekty se budeme zabývat v dalších částech tohoto oddílu.

Velký vývoj nastal v oblasti škálování. Škálování testů vychází z metod, které byly vyvinuty v rámci amerického programu *National Assessment of Educational Progress (NAEP)*, který byl prvně realizován ve Spojených státech v roce 1969 a od té doby ve Spojených státech pravidelně hodnotí vývoj výsledků vzdělávání prostřednictvím výběrových šetření. NAEP byl od počátku inovativní tím, že jeho cílem nebylo měřit rozdíly mezi jednotlivci, tak jak bylo v té době (prostřednictvím klasické testové teorie) zvykem, ale zjišťovat důležité vědomosti a dovednosti u skupin respondentů. Tyto skupiny byly vymezeny prostřednictvím proměnných, které byly

relevantní z pohledu vzdělávání, jako je například pohlaví, přistěhovalecký status a etnicita. Cílem programu bylo s pomocí expertních skupin definovat důležité cíle v jednotlivých oblastech vzdělávání, zjišťovat, do jaké míry je jednotlivé skupiny dosahují a jak se jejich vědomosti a dovednosti mění v čase.

NAEP začal být velmi záhy kritizován za to, že jeho interpretace jsou založeny na velmi omezeném počtu úloh, které se nacházely v jednotlivých testech. V roce 1980 realizoval *Education Testing Service (ETS)* NAEP na základě metodologie Samuela Messicka, Alberta Beatona a Frederica Lorda, která byla popsána v publikaci *National Assessment of Educational Progress reconsidered: a new design for a new era* (1983). Představili možnosti využití teorie odpovědi na položku (IRT), která má proti klasické testové teorii výhodu v tom, že přímo podporuje tvorbu škál pro různé soubory testových úloh. Umožnila tedy aplikaci neúplného testového designu, při kterém různí respondenti řeší různé sady testů. Tato metoda umožňuje hodnotit testované oblasti v plné jejich šíři, neboť se výzkumníci nemusí spolehnout pouze na jedinou sadu úloh, kterou vyřeší v omezeném testovacím čase jeden respondent. Použité metody umožnily pokrýt testovanou oblast šířeji a zobecnit interpretace z jednotlivých úloh na celou testovanou oblast, kterou jednotlivé úlohy reprezentují.

Kromě využití metodologie IRT vedla práce týmu NAEP k vývoji nových metod, včetně postupů marginálních odhadů (*marginal estimation*), které umožňují optimalizovat prezentaci výsledků pro velmi komplexní šetření (von Davier a kol. 2006).

Vyvíjely se i pohledy na vhodnost modelů používaných pro zpracování výsledků. Jednparametrický (Raschův) model (který předpokládá stejné parametry úloh s výjimkou obtížnosti) byl postupně nahrazován modelem dvouparametrickým (který zohledňuje kromě různé obtížnosti ještě různou rozlišovací schopnost úloh) a tříparametrickým (který zohledňuje navíc ještě „hádání odpovědi“ v případě úloh s výběrem odpovědi). Postupně začalo docházet k tomu, že žáci jsou testováni jen v některých oblastech a jejich výsledky v dalších oblastech jsou odhadovány na základě zjištěných výsledků a dotazníkových proměnných. Výsledky v jednotlivých oblastech nejsou vyjadřovány jedním skórem, ale několika náhodnými hodnotami z rozdělení naměřené zdatnosti (*plausible values*). Této problematice se podrobněji věnovala kapitola 3.

Vývoj v oblasti škálování rozhodně nemůžeme považovat za ukončený. Výše uvedené modely jsou pro účely konkrétních mezinárodních výzkumů modifikovány, zároveň nepanuje shoda na tom, jaký model je pro zpracování dat z jednotlivých výzkumů nejvhodnější. Například výzkum PISA byl kritizován za používání jednparametrického Raschova modelu. I když se zdá, že tato volba modelu neovlivnila negativně prezentované výsledky, neustálá polemika bezesporu bude nutit výzkumníky k hledání stále nových dokonalejších řešení (Gustafsson; Rosén 2014). Negativním důsledkem tohoto vývoje ovšem může být posilování současné tendence, kdy metodologie mezinárodních výzkumů a jejich výsledky budou čím dál tím méně intuitivně pochopitelné a stanou se srozumitelnými pouze pro hrstku poučených odborníků, kteří dokáží kontinuálně držet krok s metodologickým vývojem. To může paradoxně vést ke ztrátě důvěry v tyto výzkumy a jejich postupy.

Charakteristickým aspektem vývoje mezinárodních výzkumů je rostoucí tendence obohacovat informace získané z testování pomocí obsažných dotazníků. Současné výzkumy v žákovské i v dospělé populaci obsahují rozsáhlé dotazníky pro různé respondenty. Například žákovské dotazníky se dotazují na obecné postoje ke vzdělávání, škole a školním aktérům a testovaným předmětům, na každodenní vzdělávací a volnočasové aktivity. Ve výzkumech dospělých se vyskytují otázky týkající se požadavků pracovního trhu, čtenářských a matematických akti-

vit na pracovišti i v běžném životě a společenských výstupů, jako je občanská angažovanost, zapojení do charitativních aktivit a důvěra ve společnost, ve které žijeme.

Aplikace metodologie odpovědi na položku i na dotazníkové proměnné umožňuje tvorbu škál charakterizujících zájmy a názory respondentů. Stejně tak jsou vytvářeny škály charakterizující míru čtenářských a matematických aktivit a využívání informačních technologií v práci i v běžném (tedy mimoškolním či mimopracovním) životě. Použití metodologie IRT umožňuje studovat rozdíly mezi jednotlivými zeměmi v charakteristikách respondentů stejným způsobem, tedy na stejných škálách a ve stejné míře podrobnosti jako rozdíly ve vědomostech a dovednostech. Data z dotazníků jsou v souvislosti s kognitivními skóry používána k zodpovězení čím dál tím složitějších otázek vzdělávací politiky, které souvisejí se vztahem mezi vzděláváním, vědomostmi a dovednostmi a dalšími typy výstupů.

Precizovány byly rovněž metody, které umožňují měřit trendy. Jak již bylo zmiňováno, jedním z důvodů popularity mezinárodních výzkumů je skutečnost, že umožňují sledovat vývoj vědomostí a dovedností v čase. To neznamena jen možnost kontinuálně sledovat, jak se mění postavení žáků z dané země ve srovnání s ostatními zeměmi, ale přímo porovnávat výsledky v jednom šetření s výsledky v šetřeních minulých. To vyžaduje vyjádřit obtížnosti úloh a výsledky žáků ve stejných jednotkách a na stejné škále. Metody přepočítávání z jednoho výzkumu na druhý nejsou triviální, jak jsme ukázali v kapitole 3.

Metodologická komplexnost mezinárodních výzkumů ovšem vedla k situaci, kdy jen menšina uživatelů dat je schopna pracovat s daty správně, tj. například správně používat odhady žákovských výsledků (*plausible values*) a replikační váhy sloužící k řádnému odhadu výběrové chyby, neboť standardní statistické programy to jednoduše neumožňují. Mezinárodní výzkumy jsou přitom důležitým zdrojem dat pro akademické analýzy i analýzy pro účely vzdělávací politiky. IEA i další organizace, které se podílejí na přípravě a realizaci mezinárodních výzkumů⁴⁸, proto začaly vyvíjet pro uživatele aplikace, které jim umožní správné výpočty standardních chyb a správnou práci se skóry stanovenými metodologií IRT, a nabízí uživatelům pravidelná školení. Aplikace jsou dostupné na stránkách těchto institucí spolu s datovými soubory a dokumentací. Tyto aplikace však stále nejsou použitelné pro všechny typy statistických analýz. Rozvoji metodologických přístupů k analýze dat se budeme podrobněji věnovat v oddíle 4.2.3.

4.1.2 Vývoj v technologické oblasti

Zcela novou etapu v mezinárodních výzkumech výsledků vzdělávání zahájila počítačová administrace. Ta byla zahájena v roce 2006 ve výzkumu PISA v rámci přírodovědné gramotnosti. Toto měření však vykazovalo řadu metodologických nedostatků a zúčastnilo se jej velmi málo zemí. V roce 2009 byla zjišťována dovednost interaktivní práce s textem v elektronickém prostředí. Toto měření proběhlo úspěšně. Výkony žáků v elektronickém testování však nebyly srovnatelné s výkony při řešení úloh v tištěných sešitech, tedy v konečném vyhodnocení nepřispívaly k hodnocení čtenářské gramotnosti, ale tvořily specifickou škálu čtení v elektronických médiích. Odpovědi žáků zde byly i nadále vyhodnocovány školenými hodnotiteli. Tento model tedy neumožňoval okamžité vyhodnocení výsledku.

Jak již bylo uvedeno výše, elektronickou administraci mezinárodních výzkumů výrazně

⁴⁸ Nejvýznamnějšími vědeckými institucemi, které se podílejí na realizaci mezinárodních výzkumů, jsou *Education Testing Service (USA)* a *Australian Council for Educational Research (Austrálie)*.

posunul kupředu výzkum v dospělé populaci PIAAC, který úspěšně administroval elektronicky nejen úlohy čtenářské, ale i úlohy matematické a úlohy na řešení problémů, a to v takové podobě, že byly vyhodnocovány automaticky. Výzkum PIAAC zároveň zavedl do mezinárodních výzkumů adaptivní testování, které umožňuje cílenější a tím pádem přesnější měření, a ukázal, že elektronickou platformu je možno vyvinout pro rozmanité jazyky - testování probíhá nejen v asijských jazycích, ale také v hebrejštině, která je psána nikoliv zleva doprava, ale zprava doleva. Na zkušenosti výzkumu PIAAC navazuje výzkum PISA, který dosud nabízel elektronické moduly jako volitelné, od roku 2015 však již je standardní způsob administrace elektronický.

S tím ovšem souvisí důležitá otázka, zda bude možno zachovat měření trendů, tj. zda je prostřednictvím počítačů možno měřit stejné vědomosti a dovednosti jako ty, které byly zjišťovány prostřednictvím tištěných testů. Mezi tvůrci vzdělávacích politik se nicméně vede také debata o tom, zda je to vůbec potřeba. Zda tedy například vědomosti a dovednosti, které byly definovány jako klíčové v roce 2000, kdy výzkum PISA začínal, jsou stejně relevantní i v roce 2015. Někteří odborníci se domnívají, že trendy je třeba v každém případě zachovat, jiní se domnívají, že je naopak zapotřebí vymezit testované oblasti nově a ověřovat to, co je aktuálně považováno za klíčové, tedy začít sledovat trendy v poněkud jiných oblastech.

Informační technologie umožňují výrazně zefektivnit výzkumy a testovat nové typy vědomostí a dovedností. Umožňují administrovat úlohy, ve kterých mají respondenti za úkol tvořit například myšlenkové mapy nebo jinak organizovat své vědomosti nebo se zapojit do diskusí a projektů s jinými respondenty (Chudowsky; Pellegrino 2003). Umožňují rovněž administrovat dynamické interaktivní úlohy založené na rozmanitých scénářích nebo simulacích, a tím také hodnotit komplexní přírodovědné vědomosti a dovednosti a dovednosti zkoumat a řešit nestandardní problémy, což klasické písemné testy neumožňovaly. Umožňují také shromažďovat další užitečné informace, jako jsou například údaje o době řešení jednotlivých úloh a jejich součástí, a údaje o tom, jak respondent při řešení úlohy postupoval. Kromě správnosti výsledné odpovědi tedy umožňují posoudit i efektivitu řešení.

Technologie umožňují provádět testování efektivněji a učinit je pro žáky zábavnější. Zároveň umožňují zpřístupnit testy hendikepovaným žákům. V současné době jsme často svědky snahy adaptovat stávající testy tak, aby byly dostupné pro žáky s rozmanitými znevýhodněními. Někteří odborníci se domnívají, že v budoucnosti bychom měli postupovat opačně a od začátku navrhovat testy tak, aby byly přístupné žákům s rozmanitým postižením a znevýhodněním a s omezenou znalostí jazyka testu. Měly by mít i pro tyto žáky vysokou validitu. Tím automaticky umožníme větší zapojení všech žáků (např. Johnstone a kol. 2009; Thompson a kol. 2002).

Počítačové systémy, které jsou vyvíjeny a zdokonalovány v současné době, propojují kvalitní výuku, učení a hodnocení. Využívají k tomu simulací a animací, videoklipů, virtuálních laboratoří, textu a úloh na vyhledávání informací. Vznikají hodnocení dovednosti vytvářet psané texty, které zohledňují způsob, jak žáci v dnešní době píšou, tedy umožňují používat textové procesory s běžně dostupnými pomůckami. Technologie umožňují rovněž poskytování hodnotné zpětné vazby jednotlivým respondentům. Moderní počítačové systémy se tak snaží vytvářet synergie mezi testováním za účelem sumativního hodnocení žáků a škol (hodnocení učení - *assessment of learning*), pomáhají plánovat a přizpůsobovat učení jednotlivým žákům (hodnocení pro učení - *assessment for learning*) a vytvářet hodnocení, které je učiteli i žáky považováno rovněž za hodnotnou výukovou a učební zkušenost (hodnocení jako učení - *assessment as learning*). Počítačové systémy se tak snaží sjednocovat testování, formativní hodnocení a poskytování podpory. Pokud mají takto fungovat, musí být však všichni aktéři, tedy tvůrci systémů i jejich

uživatelé, ve shodě, pokud jde o vzdělávací cíle a výstupní standardy. Možnostmi využití těchto moderních systémů hodnocení se zabývá kapitola 4.1.4.

Jak již bylo uvedeno, v mezinárodních výzkumech jsou počítačové systémy používány ke třem rozličným účelům. Zaprvé jsou používány k lepšímu hodnocení tradičních oblastí: matematiky, jazyka a přírodovědných předmětů. Zde je využíván jejich potenciál vytvořit bohatší a interaktivnější prostředí a umožňují hodnotit ty aspekty daných oblastí, které bylo doposud obtížné hodnotit.

Druhou oblastí, kde je možné informační technologie s výhodou využít, je hodnocení obecnějších přenositelných dovedností, jako je samotná dovednost pracovat s informačními technologiemi či dovednost porozumět různě prezentovaným informacím, organizovat je, používat a sdělovat.

Třetí velmi důležitou oblastí je používání informačních technologií k hodnocení komplexnějších konstruktů, které charakterizují přemýšlení o dovednostech pro 21. století. Sem patří například tvořivost nebo řešení problémů ve spolupráci.

V roce 2009 uveřejnil Kozma přehled výhod a nevýhod zavádění informačních technologií do mezinárodních výzkumů.

Jako přednosti uvedl například (volně citováno dle Bellar 2013):

- Snížení nákladů na pořizování, sběr, agregaci, verifikaci a analýzu dat.
- Možnost adaptovat obtížnost testů pro jednotlivé respondenty tak, že měření bude více odpovídat jejich úrovni a bude přesněji vypovídat o jejich schopnostech.
- Možnost efektivního sběru a kódování odpovědí včetně automatického nebo poloautomatického vyhodnocování otevřených odpovědí.
- Možnost sledovat postup při řešení úloh, nejen finální odpověď.
- Možnost využít elektronických nástrojů, které jsou běžně používány, jako jsou nástroje pro organizaci myšlenek a psaní textů, nástroje pro analýzu dat ze společenských věd a vizualizace a modelování v přírodních vědách.
- Možnost poskytnout tvůrcům kurikula, výzkumníkům, učitelům i žákům detailní informace, které mohou být využity ke zlepšení dalšího učení.

Mezi technologickými výzvami uvedl například:

- Vysoké počáteční náklady (hardware, software, síťování, vývoj software včetně lokalizace, personál pro technickou podporu a údržbu).
- Nutnost volby mezi existujícími aplikacemi, jež znemožňují standardizaci, ale umožňují žákům pracovat ve známém prostředí; zároveň mohou znevýhodnit některé žáky, kteří jsou zvyklí pracovat v prostředí jiném, a umělými aplikacemi, které jsou standardizované a znevýhodňují všechny stejně, neboť s nimi nikdo běžně nepracuje.
- Nutnost integrovat aplikace a systémy tak, aby mohla být shromažďována a agregována standardizovaná informace.
- Nutnost volit mezi nezávislou a internetovou aplikací, přičemž nezávislá klade velké nároky na standardizaci a jsou s ní spojeny náklady na agregaci dat; internetová zase klade nároky na kvalitu připojení a kapacitu, jakož i na zajištění bezpečnosti.
- Potřeba pracovat s širokou škálou jazyků, písma a systémem symbolů v úlohách i při vyhodnocování odpovědí.

Lze očekávat, že spolu s inovacemi v administrativních platformách a s tím souvisejícím zařazo-

váním nových úloh a nových testových oblastí se bude rozšiřovat a měnit i obsah doprovodných dotazníků. Inovace budou tvořit cyklus, který velmi pravděpodobně povede k novým metodám hodnocení, novým interpretačním modelům a novým oblastem hodnocení.

4.1.3 Ověřované vzdělávací cíle

První kapitola této publikace ukázala, že mezinárodní výzkumy se tradičně zaměřovaly na žákovské populace a testovaly zejména typické akademické oblasti, jakými jsou matematika, čtenářské dovednosti a přírodní vědy. Později byly i s příchodem měření v dospělé populaci zařazeny čtenářská a matematická gramotnost a řešení problémů. S rostoucím významem informačních technologií se objevila šetření dovedností v práci s informačními technologiemi, digitální čtení a řešení problémů v prostředí informačních technologií. V pravidelných intervalech jsou testovány i vědomosti a dovednosti v oblasti výchovy k občanství.

Organizátoři mezinárodních výzkumů výsledků vzdělávání si začínají rostoucí měrou uvědomovat, že obsah mezinárodních výzkumů a způsob, jakým jsou prezentovány výsledky, zásadním způsobem ovlivňují obsahy i metody vzdělávání na celém světě. Tvůrci vzdělávacích politik se (ve snaze zlepšit výsledky svých žáků v mezinárodních výzkumech) snaží zařazovat do vzdělávacích programů vědomosti a dovednosti, které jsou v mezinárodních výzkumech zjišťovány. Kritici mezinárodních výzkumů, ale i jejich tvůrci se obávají, že tak mnohdy činí na úkor jiných vědomostí a dovedností, které měřeny nejsou.

V poslední době proto sílí požadavek co nejvíce rozšířit rozsah ověřovaných vědomostí a dovedností, aby nedocházelo ke zužování kurikula a aby zároveň byla informace, kterou komparativní studie přinášejí, co nejužitečnější (např. Loony 2011). V této souvislosti vznikla řada iniciativ, které se snaží definovat takzvané „kompetence pro 21. století“ v celé jejich šíři a naučit se tyto kompetence rovněž hodnotit. Jako příklad takové iniciativy je možno uvést projekt *Assessment and Teaching of 21st Century Skills*, který byl sponzorovaný společnostmi Microsoft, Intel a Cisco v letech 2009-2012. Tvůrci projektu, jimiž jsou přední světoví odborníci v oblasti vzdělávání, vyjadřují přesvědčení, že současný obsah školního vzdělávání nepřipravuje žáky dostatečně na život a práci ve 21. století. Zaměstnavatelé se podle jejich mínění potýkají s tím, že čerství absolventi postrádají dovednosti, které jim umožňují vymýšlet, budovat a udržovat v chodu informačně bohaté podniky. Za základní kameny vzdělávání sice stále považují čtení, psaní, matematiku a přírodovědné předměty, ale domnívají se, že vzdělávání musí být obohaceno o další oblasti. Jako hlavní nedostatky uvádějí nedostatečný výcvik v efektivním využívání informačních technologií a v týmové spolupráci. Ambicí tvůrců je vyvinout počítačové úlohy na skupinové řešení problémových úkolů (*collaborative problem solving*), které obrátí pozornost tvůrců vzdělávacích politik na celém světě k těmto důležitým dovednostem. Kompetence pro 21. století jsou v uvedeném projektu pracovníě definovány takto⁴⁹:

Způsoby myšlení

- Tvořivost a inovace
- Kritické myšlení, řešení problémů, rozhodování
- Kompetence k učení, metakognice

⁴⁹ Více informací na: <http://atc21s.org/>.

Způsoby práce

- Komunikace
- Týmová práce

Pracovní nástroje

- Informační gramotnost
- ICT gramotnost

Život na světě

- Občanství – lokální a globální
- Život a pracovní kariéra
- Osobní a společenská zodpovědnost a kompetence – včetně kulturního uvědomění

Iniciativ, které si kladou za cíl vymezit „kompetence pro 21. století“, je mnoho. I když se jejich konceptuální rámce liší v podrobnostech a struktuře, mají zpravidla velmi podobné složky. Všem iniciativám zacíleným na definování dovedností pro 21. století je především společná snaha dobře definovat nekognitivní dovednosti, které jsou zpravidla členěny na intrapersonální dovednosti (např. schopnost stanovovat si cíle a reflektovat postupy k jejich dosahování, cílevědomost, angažovanost) a interpersonální dovednosti (zejména dovednost spolupracovat a domlouvat se) a potřeba naučit se je rozvíjet a hodnotit. Ve výše uvedené struktuře projektu ATC21 je nacházíme ve 2. a 4. bodě.

Společnost *Education Testing Service* pravidelně pořádá konference pro odborníky a tvůrce vzdělávacích politik, které jsou věnované mezinárodním a národním výzkumům výsledků vzdělávání. Jejich cílem je hovořit o dopadech těchto šetření na vzdělávací systémy a hledat cesty, jak zvýšit jejich relevantnost a užitečnost pro zlepšování vyučování a učení a formování vzdělávacích politik. Pravidelným tématem je vymezování a měření nekognitivních vzdělávacích cílů.

Zajímavý příspěvek do debaty o měření nekognitivních dovedností přinesla pracovní skupina *Committee on the Assessment of 21st Century Skills*, která se scházela v průběhu let 2010 a 2011 ve Spojených státech a měla za úkol vymezit dovednosti pro 21. století a posoudit možnosti jejich hodnocení. V rámci tohoto projektu vznikla užitečná inventura zkušeností s měřením kognitivních výstupů v oblasti řešení nerutinních problémů, systémového myšlení a kritického myšlení a nekognitivních výstupů v oblasti intrapersonálních a interpersonálních dovedností v rozmanitých oblastech lidského konání (Koenig a kol. 2011)⁵⁰.

V posledních letech je zdůrazňován význam sociálně-emočních kompetencí. Elias (2013) uvádí, že cílem všech vzdělávacích systémů je vzdělat mladé lidi tak, aby:

- byli plně gramotní, aby byli schopni profitovat z bohatství psané a mluvené řeči v rozmanitých formách;
- rozuměli matematice a přírodovědným předmětům na takové úrovni, která je připraví pro svět budoucnosti a posílí jejich schopnost myslet kriticky, obezřetně a kreativně;
- uměli řešit problémy;
- přijímali zodpovědnost za své zdraví a pohodu;

⁵⁰ Interpersonální dovednosti zde byly vymezeny jako: komplexní komunikace, sociální dovednosti, týmová práce, kulturní citlivost a přijímání odlišností. Intrapersonální dovednosti jako sebeřízení, hospodaření s časem, vlastní rozvoj, sebekontrola, adaptabilita, schopnost jednat.

- vytvořili si efektivní sociální vazby a naučili se spolupracovat s lidmi z odlišných kultur a prostředí, naučili se jim rozumět a vytvářet si k nim přátelské vztahy;
- byli ohleduplnými jedinci, kteří budou mít zájem o druhé a budou je respektovat;
- rozuměli tomu, jak funguje společnost, a byli připraveni přijímat zodpovědnost za její vývoj;
- měli dobrý charakter a činili rozumná morální rozhodnutí.

Elias (2003) uvádí, že sociálně-emoční vzdělávání je někdy označováno jako „chybějící článek“, protože reprezentuje mnohdy opomíjenou důležitou součást vzdělávání, která propojuje akademické vědomosti se specifickými dovednostmi nezbytnými pro úspěch ve škole, rodině, společnosti, na pracovním trhu a v životě jako takovém. Pro společnost je velmi nebezpečné, když vybavuje děti vědomostmi bez sociálně-emočních dovedností a silného morálního nasměrování.

Všechny iniciativy se shodují v tom, že význam nekognitivních dovedností pro uplatnění na trhu práce a v životě ve 21. století je minimálně stejně tak velký jako význam dovedností kognitivních. Zároveň je zřejmé, že skutečnost, že je neumíme hodnotit (a to nejen na mezinárodní, ale ani na národních úrovních), vede k tomu, že jsou ve výuce mnohdy opomíjeny. Snaha napravit tento nedostatek vychází rovněž z předpokladu, že dobrý koncept hodnocení má potenciál napomoci porozumění těmto kompetencím v odborné i široké veřejnosti a pomůže hledat cesty k jejich cílenému rozvíjení.

Jak již bylo uvedeno, rozvoji nekognitivních dovedností ve školách velmi škodí, že pozornost je zaměřena výlučně na dovednosti kognitivní. Důraz na kognitivní dovednosti má samozřejmě své opodstatnění. Testované dovednosti jsou důležitými prediktory vzdělávacích výsledků a životního úspěchu, metodologie výzkumů je velmi kvalitní a ve srovnání například s měřením autentických výkonů žáků nebo jejich sociální dovedností relativně nenákladná. Zdaleka však neplatí, že kognitivní dovednosti jsou to jediné, na čem záleží v procesu vývoje zdravé a produktivní dospělé osobnosti. Přestože jsou důležitými prediktory produktivity a příjmu na individuální i společenské úrovni, empirické výzkumy ukazují, že jejich dopad je daleko menší, než se obecně předpokládá. Zaměstnavatelé, kteří si stěžují na špatnou výbavu zaměstnanců, daleko častěji zmiňují absenci žádoucích postojů, jako jsou zodpovědnost, svědomitost a sociální chování, než nízkou úroveň vědomostí a dovedností.

Podle Levina (2013) platí, že akademické výsledky jsou korelovány s nekognitivními atributy a slouží jako zástupný indikátor při předpovídání ekonomických výstupů, což vede k jejich přeceňování v situaci, kdy nekognitivní indikátory nejsou k dispozici. Dále platí, že nekognitivní atributy nejsou pouze korelovány s kognitivními, ale přispívají k nim. Výlučný důraz na kognitivní aspekty může mít velmi škodlivé dopady, neboť může negativně ovlivnit chování učitelů, kteří budou ignorovat nekognitivní dovednosti a nebudou doceňovat úlohu učitelů a škol v rozvoji těchto dovedností.

Levin (2013) uvádí, že nejkomplexnější výzkum programů na rozvoj nekognitivních dovedností u mladých lidí (Durlak a kol. 2011) ukázal 5 důležitých faktorů:

- sociálně emoční dovednosti – identifikace emocí ze sociálních signálů, stanovování cílů, zaujímání stanoviska, řešení mezilidských problémů, řešení konfliktů a rozhodování;
- postoje k sobě a druhým – pozitivní vztah k vlastní osobě, ke škole a k sociálním tématům;
- pozitivní sociální chování – schopnost dobře vycházet s druhými (identifikovaná vnějším pozorovatelem);

- problémové chování – vyrušování, neposlušnost, agresivita, šikana, delikvence;
- emoční stres – deprese, úzkost, stres, stranění se společností (identifikované vnějším pozorovatelem).

Levin (2013) uvádí, že pro zkoumání vztahu mezi nekognitivními dovednostmi, akademickými výsledky a dosaženým vzděláním je nejčastěji používáno následujících pět faktorů (*big five*):

- otevřenost – invenční a zvědavý jako opak k neměnný a opatrný
- svědomitost – efektivní a výkonný jako opak k lehkovážný a nedbalý
- extraverte – vstřícný a energický jako opak k osamělý a rezervovaný
- ochota – přátelský a soucitný jako opak k chladný a nevládný
- neurotismus – citlivý a nervózní jako opak ke klidný a sebejistý

Měření nekognitivních dovedností je na volitelné bázi realizováno i v rámci mezinárodních výzkumů. Např. Německo administrovalo v rámci výzkumu TIMSS a PIRLS 2011 v rámci žákovského dotazníku pro žáky 4. ročníku baterii na zjišťování sociálních kompetencí. Zjišťovány byly následující dimenze: kognitivní vcítění, emocionální vcítění, schopnost ovlivňovat pocity druhých, prosociální chování, vznětlivost, schopnost se prosadit. Sociální kompetence byly posuzovány na třech úrovních a byly dávány do souvislosti s rodinným zázemím, mediálními aktivitami (televize, počítačové hry) a vnímaným postavením ve třídě (pocit sounáležitosti se školou, přátelství, pocit odcizení) (Fay a kol. 2015).

Nekognitivní výstupy se v mezinárodních výzkumech dosud snažily alespoň částečně podchytit výzkumy občanské výchovy. Analýza dat z těchto výzkumů například ukázala, že třídní klima otevřené diskusi pozitivně ovlivňuje znalosti a angažovanost. Výzkumy občanské výchovy zahrnují část dovedností označovaných jako dovednosti pro 21. století, například dovednost porozumět mediálním sdělením a ekonomickým a globálním problémům.

Existují doklady o tom, že měření občanských dovedností pomáhá vzdělavatelům zacílit výuku a profesní rozvoj učitelů (Torney-Purta; Amadeo 2013). V této oblasti se zřetelně ukazuje, že nemůžeme očekávat, že mladí lidé budou ovládat efektivní strategie v občanském i pracovním životě, pokud je nebudeme těmto strategiím učit. V případě výzkumů občanství navíc zjišťování postojů dospívajících k menšinám a přistěhovalcům umožní porozumět tomu, jak se tyto postoje ve společnosti utvářejí.

Přemýšlením o závažnosti nekognitivních dovedností a potřebou rozšířit rozsah hodnocených dovedností byl motivován i vývoj oblasti řešení problémů ve spolupráci a její zařazení do výzkumu PISA. Z konceptuálního rámce této oblasti, který byl prezentován v kapitole 2, je zřejmé, že je zacílen jak na cíle v oblasti komunikace a spolupráce (interpersonální dovednosti), tak na cíle v oblasti kognitivní a technologické. Ve snaze zařadit do výzkumů důležité aktuální vědomosti a dovednosti je pro výzkum PISA 2018 vyvíjen koncept tzv. globálních kompetencí. Kontrakt na vývoj konceptuálního rámce pro globální kompetence vyhrála nadnárodní vzdělávací organizace Pearson.

Nástroj pro měření globálních kompetencí by měl měřit povědomí žáků o propojenosti světa, ve kterém žijeme, a jejich schopnost efektivně se vypořádat s výzvami, které tato propojenost přináší. OECD považuje rozvoj globálních kompetencí za novou odnož výchovy k občanství, která pomáhá žákům k tomu, aby se stali autonomními individualitami, jež mohou vyjednávat o globálních otázkách a účastnit se globálních aktivit, a k tomu aby byli schopni se vyrovnat s dilematy a kontroverzemi, které vznikají v důsledku globalizace a nemají žádná jednoduchá řešení. K nalezení řešení je však důležité vědomí kulturní rozmanitosti, které pomáhá zodpo-

vědět důležité otázky související s tím, jak v multikulturních společnostech vyvinout sdílené hodnoty (Reimers 2013).

Podle americké organizace *National Educational Association (NAE)* sestávají globální kompetence ze čtyř hlavních složek:

1. mezinárodní povědomí (znalost světové historie a ekonomických a politických systémů),
2. oceňování kulturní diversity (schopnost poznat lidi z jiných kultur, porozumět jim a oceňovat je),
3. znalost cizích jazyků,
4. kompetitivní dovednosti (schopnost obstát v mezinárodní konkurenci).

4.1.4 Moderní systémy kombinující formativní a sumativní hodnocení

Jak již bylo uvedeno, rozvoj mezinárodních výzkumů výsledků vzdělávání probíhal paralelně s rozvojem národních testovacích programů. Tyto postupně začaly velmi významně ovlivňovat životy žáků a učitelů zejména ve Velké Británii a Spojených státech, kde jsou podrobovány čím dál ostřejší kritice. Jedná se o programy, ve kterých jsou v určitých fázích vzdělávací dráhy hodnoceni všichni žáci a výsledky testů mají dopad na rozhodování o jejich dalším vzdělávání, případně mají dopad na hodnocení a osudy učitelů a škol (*high stakes testing*). Závěry studií dopadů testovacích programů na výsledky žáků poskytují značně kontroverzní výsledky. Mnozí odborníci dospěli k závěru, že standardizované testování samo o sobě, bez systematické podpory vyučování a učení, nemá pozitivní dopady na výsledky žáků a tím méně na profesní rozvoj učitelů (Loony 2011). Mnozí tvůrci vzdělávacích politik naopak argumentují, že testování a s ním spojená akontabilita⁵¹ mají příznivý dopad na výsledky vzdělávání, což je vidět na tom, že se výsledky zlepšují, a ohrazují se proti jeho zpochybňování (např. Ritzen 2014). Koretz (2005) však ukázal, že testování vede ke strategickému chování škol a učitelů, a cílené přípravy na testování vedou k umělému zvyšování testových skóre, aniž by se zvýšily vědomosti a dovednosti žáků. To se projeví okamžitě v situaci, kdy se žáků dotážeme na trochu jiné učivo nebo jim zadáme jiný typ úlohy, než jsou zahrnuty v testech, i když patří do stejného okruhu vzdělávacích cílů. Ukazuje to, že se nezlepšuje celková zdatnost žáků v dané oblasti, ale pouze ty její aspekty, které jsou součástí testu, přesně v té podobě, v jaké v testu figuruje. Blíže se této problematice věnuje kapitola 5.

I když panují různé pohledy na to, zda pravidelné testování zlepšuje výsledky vzdělávání, málokdo pochybuje o tom, že má řadu nežádoucích vedlejších efektů. Nejčastěji je diskutováno oklešťování kurikula, snaha škol zbavovat se žáků, kteří jim kazí výsledky v testech, tendence k podvádění (zejména na straně škol a učitelů), a negativní dopad na motivaci učitelů (Morris 2011). Výzkumy ukazují, že při posuzování přínosů a negativních dopadů testování je vždy třeba věnovat velkou pozornost strategickému chování učitelů. Masters a Foster (2000) například poukazují na strategické chování učitelů v případě testů, které ověřují dosažení nějaké minimální úrovně (tedy testů ověřujících nějaký minimální standard). V této situaci se učitelé zaměřují na práci s žáky, kteří se nacházejí v blízkosti této úrovně, aby zajistili, že

⁵¹ Akontabilita (z anglického *accountability*) představuje povinnost škol zodpovídat se z výsledků žáků. Mají-li žáci či školy špatné výsledky, jsou školy poháněny k zodpovědnosti a musí čelit nepříznivým důsledkům. Veselý (2012) definuje akontabilitu obecně jako „vztah mezi aktérem a fórem, ve kterém aktér má povinnost vysvětlovat a zdůvodňovat své jednání, přičemž fórum může klást otázky a posuzovat aktéra, který také může čelit důsledkům“.

v testu uspějí a zlepší celkový výsledek školy. Zcela však zanedbávají žáky, kteří jsou vysoko nad nebo hluboko pod touto úrovní.

Morris (2011) dává následujících šest doporučení, jak eliminovat negativní dopady testování:

- Jasně vymezit účel testování a tomuto účelu podřídít testový design, implementaci i využití výsledků.
Tedy neřítit se primárně tím, jaký test umíme vytvořit, ale tím, co potřebujeme zjišťovat.
- Standardy ověřované testy pečlivě provázat s kurikulem a stanovenými cíli vzdělávání.
To znamená klást v kurikulu důraz na to, co je důležité, a to také ověřovat. Tedy zároveň také ověřovat to, co se žáci učí.
- S velkou opatrností aplikovat testy, které slouží více účelům.
To například znamená snažit se vyvarovat situací, kdy jeden test slouží hodnocení výsledků vzdělávání a zároveň i jako nástroj výběru pro další studium.
- Pečovat o to, aby ředitelé, učitelé a zřizovatelé byli poučeni v problematice hodnocení a testování a dobře rozuměli jejich limitům a rizikům.
Ukazuje se, že pedagogové, ale i zřizovatelé škol dostatečně nerozumějí testování a neumějí výsledky testů interpretovat a těžít z nich poučení pro další práci a rozhodování. Často se hovoří o tom, že učitelé a zřizovatelé nedisponují dostatečnou datovou gramotností (*data literacy* nebo *assessment literacy*).
- Redukovat podvádění a strategické chování zapojením učitelů do procesu tvorby a vyhodnocování výsledků a získáváním jejich podpory pro testování od samého počátku.
Je důležité, aby učitelé získali pocit, že testy ověřují stejné cíle, které sledují ve své výuce, a mohou jim poskytnout užitečnou informaci o vědomostech a dovednostech jejich žáků, a že zároveň nebudou za výsledky žáků popotahováni. Pak je větší pravděpodobnost, že cílům dobře porozumějí a budou je schopni ve výuce rozvíjet a nebudou podvádět při přípravě na test a jeho realizaci.
- V systémech, kde na základě testů probíhá rozhodování o osudech žáků a učitelů, aplikovat více měření, tedy nezakládat rozhodování pouze na jednom hodnocení.

Masters a Forster (2000) také zdůrazňují závažnost toho, aby do hodnocení byly zařazeny i úlohy, které nejsou běžně počítačově vyhodnotitelné, a aby do hodnocení těchto úloh byli zapojeni učitelé. Uvádějí příklady úkolů zaměřených na hodnocení písemného nebo ústního projevu žáků. Upozorňují na to, že pokud se budeme v hodnocení řídit pouze tím, co je jednoduše vyhodnotitelné, vědomě budeme redukovat školní vzdělávání na dosahování pouze těchto cílů a budeme rezignovat na řadu jeho důležitých aspektů. Povinnost vyhodnotit výkony žáků na základě hodnotícího schématu připraveného profesionály a možnost hovořit o žákovských výkonech a jejich charakteristikách s jinými učiteli u příležitosti zácvičku, má velmi pozitivní dopad na profesní rozvoj učitelů. Prostřednictvím této zkušenosti se učí lépe porozumět vzdělávacím cílům a tomu, jak se v práci žáků odráží míra jejich dosahování.

Pokusy o odstranění výše uvedených nedostatků u národních testovacích programů vedly k tvorbě systémů, které monitorují průběžný pokrok žáků. Slouží učitelům (ale i žákům a rodičům) jako podklad pro formativní hodnocení a pro rozhodování o dalším vzdělávacím postupu. To znamená, že jsou navázány na kurikulum a zjišťují cíle vzdělávání, které jsou v kurikulu definovány. Vzhledem k tomu, že jejich účel je primárně formativní, mohou zahrnovat i úlohy, které nejsou úplně striktně objektivně vyhodnotitelné. To znamená, že mohou poskytnout

informace o široké škále výsledků vzdělávání a zahrnovat úlohy, které ověřují obtížnější hodnotitelné vzdělávací cíle. Zároveň mohou aplikovat rozmanité hodnotící postupy.

Pro sledování pokroku žáků jsou vytvářeny popisy rozvoje sledovaných vědomostí a dovedností – vývojová kontinua (u nás používáme často termín „mapy učebního pokroku“). Tato kontinua obsahují popisy vědomostí a dovedností na jednotlivých stupních vývoje, tedy ukazují, jaké nové dovednosti si žák postupně osvojuje, když se v dané oblasti zlepšuje. Jejich součástí jsou rovněž úlohy, které tyto vědomosti a dovednosti ověřují, tedy ty, jež by měli zdárně vyřešit žáci, kteří se nacházejí v určité fázi vývoje. Vzhledem k tomu, že vědomosti a dovednosti žáků v jedné třídě, respektive v jednom ročníku mají široký rozptyl, je velmi výhodné, když mohou žáci dostávat úlohy různé obtížnosti odpovídající jejich aktuální úrovni. To znamená, že každý žák pracuje na té úrovni, která je pro něj v danou chvíli nejvhodnější, kde je schopen se nejefektivněji učit. Přiřazování úloh vhodné obtížnosti je nejlépe realizováno prostřednictvím adaptivního testování, kdy je obtížnost volena na základě předchozího výkonu žáka. Zakotvení výkonu žáka na vývojové kontinuum má tu výhodu, že nese bohatou informaci o žákových vědomostech a dovednostech. Přesně sděluje žákovi, co si již osvojil, co už umí, a zároveň ukazuje cestu vpřed, tedy na jaké vědomosti a dovednosti je třeba se zaměřit, chceme-li postoupit dále. Tedy poskytuje informaci, která při běžném známkování chybí.

Systém, který průběžně monitoruje pokrok jednotlivých žáků, poskytuje zároveň „sumativní“ informaci o aktuální úrovni žákových vědomostí a dovedností, tedy umožňuje například monitorovat podíl žáků, kteří v daném ročníku dosáhli minimálního vzdělávacího standardu, je-li takový standard definován. Průběžné sledování žáků a sledování jejich výsledků ve vhodně volených úlohách, které se nacházejí v blízkosti tohoto standardu, umožňuje také lépe porozumět obtížím, projevujícím se u žáků, kteří mají s dosažením standardu problémy. Při prezentaci výsledků je možno také vymezit nejen minimální standard, ale také standardy pro různé úrovně obtížnosti, a lze prezentovat podíl žáků na těchto úrovních. Výhoda tohoto přístupu spočívá v tom, že se obrací pozornost pedagogické veřejnosti nejen k minimální úrovni, ale také k vyšším úrovním, které popisují, jak vypadají excelentní výkony.

Pokud je v rámci vzdělávacích systémů, které používají výše uvedené systémy hodnocení, shoda stran cílů, mohou tyto moderní systémy hodnocení poskytnout žákům, učitelům a tvůrcům vzdělávacích politik nástroje, které jim umožní rozpoznat, zda se žáci od prvního stupně základní školy učí tak, aby byli úspěšní v dalším vzdělávání a v životě. Umožní též včas identifikovat jejich potřeby a cíleně a včas odstranit jejich nedostatky. S takovými systémy nemusí školy přerušovat výuku, aby administrovaly žákům externí testy, a nemusejí ztrácet čas přípravou na testování. Systémy by tedy měly posloužit zároveň pro sumativní i formativní účely.

Všechny výše uvedené negativní dopady testování, které se snaží překonat moderní systémy hodnocení, jsou vytýkány rovněž mezinárodním výzkumům, přes to, že se nejedná o testy, které mají přímý dopad na osudy žáků nebo jejich učitelů (tedy nejsou „*high stakes*“). Některé kritiky budí dojem, že někteří aktéři mezi oběma typy testování dobře nerozlišují. Neuvědomují si, že mezinárodní šetření se týkají v naprosté většině systémů pouze malé části žáků a učitelů a že výsledky nemají žádný dopad na osud jednotlivých žáků, jednotlivých učitelů ani škol. To může negativně ovlivnit motivaci žáků snažit se při testování, ale zároveň to eliminuje většinu negativních aspektů testování. Kritici ovšem pravděpodobně vycházejí z předpokladu, že mezinárodní aplikací získává testování jakousi větší oporu, což se negativně projeví na domácích testovacích aktivitách. Podrobněji se kritikou mezinárodních výzkumů zabýváme v oddíle 5.1.

Jako cesta k překonání negativních dopadů je i v případě mezinárodních výzkumů zvažován přechod na model komplexních systémů s výše uvedenými charakteristikami. I když vývoj výše popsaných systémů pro integraci sumativního a formativního hodnocení již intenzivně probíhá a v řadě vzdělávacích systémů již podobné systémy v nějaké podobě fungují⁵², není úplně zřejmé, zda je možné a dokonce žádoucí dosáhnout něčeho takového na mezinárodní úrovni. K tomu je potřebné sjednocení vzdělávacích cílů, obsahů a výstupních standardů s interními a externími hodnotícími úlohami. I když mnozí argumentují, že mezinárodní výzkumy významně sbližují kurikula zúčastněných zemí, ve vymezení cílů vzdělávání a výstupních požadavků na žáky existuje stále velká diverzita, která plošnou aplikaci podobných systémů na mezinárodní úrovni vylučuje. Aplikace takových systémů se tedy minimálně v blízké budoucnosti nejeví reálnou, zejména proto, že nejsou sjednocena kurikula a jejich sjednocování se ani nejeví žádoucím. Zcela zásadním předpokladem úspěchu takového programu je právě jeho těsné navázání na národní kurikula a národní cíle vzdělávání. Zároveň existují velké rozdíly mezi jednotlivými vzdělávacími systémy v dostupnosti potřebné infrastruktury. Metody používané v mezinárodních výzkumech budou však zcela jistě vývojem těchto komplexních platform významně ovlivněny.

4.2 Využití dat z mezinárodních výzkumů – historie a současnost

4.2.1 Hledání faktorů ovlivňujících výsledky vzdělávání

Aspekty ovlivňující využití dat

Jak již bylo uvedeno v úvodu, původním cílem mezinárodních výzkumů bylo odpovědět na otázku, jaké faktory na úrovni systému, školy, učitele a žáka ovlivňují efektivitu vzdělávání. Zpočátku byly také koncipovány tak, aby na tuto otázku principiálně mohly odpovídat, neboť obsahovaly longitudinální složku. Například výzkum SIMS (*Second International Mathematics Study*) obsahoval pretest a posttest. Jako longitudinální byly koncipovány i výzkumy *Classroom Environment Study* a *Preprimary Project* realizované v 80. letech Asociací pro hodnocení výsledků vzdělávání. Opuštění longitudinální povahy výzkumů z 80. let v jejich další historii silně redukovalo jejich potenciál přispět k vysvětlení determinant kvalitní výuky. Výjimkou v tomto ohledu byly videostudie realizované v letech 1995 a 1999 jako součást výzkumu TIMSS. Podrobné analýzy nahraných vyučovacích hodin pomohly porozumět tomu, jak učební postupy přispívají ke vzdělávacím výsledkům. Jedním z aspektů, které byly ve výzkumu sledovány, byla i míra, v jaké učitelé při výuce chybují. Zde se ukázalo, že čeští učitelé dělají nejméně faktických chyb ve srovnání s kolegy z ostatních zúčastněných zemí. Výsledky týkající se věcného chybování učitelů však nikdy nebyly zveřejněny z obavy z reakce učitelů.

⁵² V budování systému hodnocení výsledků vzdělávání dosáhli například pozoruhodných výsledků maďarští kolegové. Univerzita v Szegedu vyvinula elektronický systém, jehož prostřednictvím mohou učitelé diagnostikovat učební obtíže žáků a sledovat pokrok jednotlivých žáků v hlavních oblastech vzdělávání v období celého povinného vzdělávání (viz např. Csapó 2012, Csapó 2014, Csapó a kol. 2014). Maďarští tvůrci vzdělávacích politik uvádějí, že tento systém je příčinou zlepšení výsledků maďarských žáků ve čtenářské gramotnosti, neboť umožňuje ranou pedagogickou diagnostiku a s ní spojenou včasnou identifikaci obtíží u jednotlivých žáků a umožňuje průběžně sledovat i efektivitu nápravných opatření.

Velkou předností studií IEA ve vztahu k faktorům ovlivňujícím vyučování a učení bylo a je, že se věnují specificky různým předmětům a studují celé školní třídy, což podporuje důraz na oborové didaktiky. S tím souvisí i přínos studie *TED-M*, která se zabývala vzděláváním učitelů. Studie OECD mají v tomto ohledu daleko menší potenciál.

Dosavadní úspěšnost mezinárodních výzkumů při stanovování faktorů, které na úrovni systému a školy ovlivňují výsledky vzdělávání, je celkově relativně nízká. Mezinárodní výzkumy jako celek nicméně pomohly porozumět tomu, jak se výsledky žáků v jedné zemi liší od výsledků v jiných zemích a pomáhají generovat hypotézy o tom, jak různé faktory na straně žáka, třídy, učitele, školy, kurikula a systému přispívají k učení žáků. Umožňují rozložit rozptyl ve výsledcích žáků na individuální, školní a systémovou úroveň a zároveň poskytují data o rozmanitých faktorech, které tyto tři úrovně zahrnují a které by (podle jiných výzkumů či teoretických předpokladů) měly ovlivňovat výsledky žáků. Statistické modely umožňují s využitím víceúrovňových dat pomoci porozumět složitým vazbám mezi vstupy, procesy a jejich interakcemi v ovlivnění výstupů. Pokud jsou k dispozici informace o zdrojích a nákladech, mohou mezinárodní výzkumy pomoci porozumět ekonomické efektivitě systémů.

Klieme (2013) zdůrazňuje, že mezinárodní výzkumy byly původně koncipovány tak, aby testovaly model zamýšleného, realizovaného a dosaženého kurikula. Současné nároky na ně kladené jsou však vyšší. Při všech analýzách je třeba zohledňovat hierarchickou strukturu systému, který má v principu čtyři úrovně: žáci, třídy, školy, systém. Tedy pro porozumění komplexním mechanismům ve vzdělávacím systému je třeba testovat přímé, zprostředkované a moderující efekty na různých úrovních. Opatření vzdělávací politiky sice působí na vyšších úrovních, ale jsou také manifestována na úrovních nižších. Tento aspekt je v naší publikaci opakovaně zmiňován.

Jak již bylo uvedeno výše, hlavní očekávání mezinárodních výzkumů výsledků vzdělávání spočívalo od počátků jejich realizace v tom, že ukáží, jaké postupy v jednotlivých vzdělávacích systémech vedou k dobrým vzdělávacím výsledkům, a že tyto postupy bude možno implementovat do systémů prokazujících nižší úspěšnost. Data jsou analyzována výzkumníky z celého světa, kteří s jejich využitím publikují odborné články i zprávy pro tvůrce vzdělávacích politik na národní i mezinárodní úrovni. Opakovaně byly činěny pokusy určit faktory, které zapříčiňují, že některé vzdělávací systémy jsou úspěšnější než jiné, tedy které jsou zodpovědné za to, že v některých vzdělávacích systémech dosahují žáci lepších výsledků než v jiných. Jak již bylo uvedeno, tyto analýzy jsou motivovány vírou, že je možno aspekty vzdělávacích systémů, které podmiňují kvalitu výsledků, replikovat i v jiných systémech s podobným výsledkem.

Velkým omezením těchto analýz je však skutečnost, že se jedná o výzkumy průřezové, které neobsahují žádnou longitudinální složku, a tedy v principu vylučují vyvozování kauzálních závěrů. Vyslovují se pouze o korelacích mezi výsledky a kontextuálními proměnnými, ale neumožňují vyvozovat závěry o tom, jaký jev podmiňuje jev jiný. Proto nemohou samy o sobě sloužit pro testování hypotéz o příčinách a následcích, ale mohou ukazovat na souvislosti mezi různými faktory a generovat hypotézy.

Produkty mezinárodních výzkumů

Podle Kliemeho (2013) nabízejí mezinárodní výzkumy primárně tři typy produktů: a) indikátory, které monitorují fungování, produktivitu a spravedlivost vzdělávacích systémů, b) znalosti faktorů, které se vztahují k efektivitě vzdělávání, c) velmi kvalitní databáze, které mohou odborníci na celém světě používat ke studiu vědeckých otázek i otázek vzdělávací politiky.

Tvůrci vzdělávacích politik se zejména zajímají o systém indikátorů, jehož užitečnost je dána a) způsobem definování a operacionalizace kognitivních a nekognitivních výstupových indikátorů, které mohou sloužit tvůrcům vzdělávací politiky v jednotlivých zemích jako podklad pro výběr a prioritizaci vzdělávacích cílů, b) zkoumáním a prezentací faktorů, které mohou být ovlivněny vzdělávací politikou a profesí, c) tvorbou mezinárodních měřítek, které ukazují tvůrcům vzdělávacích politik, kde se mohou poučit z jiných zemí.

Výběr indikátorů je určen požadavky vzdělávací politiky, která se musí zabývat fungováním systémů, jejich produktivitou a spravedlivostí. Je podmíněn sdílenými prioritami, které jsou společné všem vyspělým zemím. Dvořák (2012) zmiňuje čtyři hlavní postuláty globální vzdělávací kultury:

1. Ideál sociální spravedlnosti/rovnosti šancí založený na přesvědčení, že nejefektivnější cestou k dosahování spravedlivosti je společné vzdělávání.
2. Víra v přínosnost rozvoje osobnosti každého člena společnosti pro pokrok a rozvoj celku.
3. Důraz na kognitivní/akademické aspekty vzdělávání, který se projevuje globální konvergencí kurikula.
4. Význam vzdělávacích výsledků jedince a dosaženého vzdělání pro úspěšnost jedince.

Indikátory získané v mezinárodních výzkumech poskytují možnost porovnávat situaci jednotlivých zemí ve vztahu k těmto aspektům, což zvyšuje význam výzkumů.

Indikátory z mezinárodních výzkumů jsou využívány i pro stanovování cílů vzdělávacích politik na evropské úrovni. Například Strategie *Education & Training 2020*⁵³ vytyčuje jako jeden z pěti cílů pro země Evropské unie do roku 2020 snížit podíl žáků s nedostatečnou úrovní matematické, čtenářské a přírodovědné gramotnosti pod 15 %. Podílem žáků s nedostatečnou úrovní v jednotlivých oblastech se rozumí podíl žáků, kteří se svými výkony umístili ve výzkumu PISA pod druhou z šesti úrovní způsobilosti.

Ukázku indikátorů z mezinárodních výzkumů můžeme nalézt například v publikaci *Education at a Glance*.

Hlavní poznatky mezinárodních výzkumů výsledků vzdělávání můžeme podle Kliemeho (2013) shrnout do následujících tří bodů:

- velké rozdíly mezi zeměmi
- velké změny ve výsledcích jednotlivých zemí
- velké rozdíly ve spravedlivosti (socioekonomický status, pohlaví, etnicita)

Výzkum PISA systematicky obrací pozornost odborné veřejnosti k problematice vzdělanostních nerovností na úrovni povinného vzdělávání. Podrobné mapování rodinného zázemí umožňuje analytikům pravidelně informovat o míře rozdílů mezi žáky a školami v jednotlivých systémech a o síle vazby mezi výsledky vzdělávání a rodinném zázemí. Velmi podrobně jsou analyzovány výsledky přistěhovalců první a dalších generací. Důraz, který OECD klade na tento typ indikátorů, významně přispěl k tomu, že se otázkou nerovností na úrovni povinného vzdělávání začaly zabývat i systémy, které jí dosud nevěnovaly pozornost, a že se otázka spravedlivosti (*equity*) ocitla na prvním místě v žebříčku deklarovaných politických priorit v mnoha vyspělých zemích.

Přesto někteří odborníci argumentují, že otázka spravedlivosti stále stojí ve stínu průměrných výsledků a že by při vyhodnocování poznatků z mezinárodních výzkumů měla být prioritní.

⁵³ http://ec.europa.eu/education/policy/strategic-framework/index_en.htm

Strietholt (2014) upozorňuje na skutečnost, že ve studiích zabývajících se spravedlivostí autoři nedostatečně popisují svá hodnotová východiska, tedy málo vysvětlují, jaké nerovnosti jsou pro ně ještě přijatelné a jaké nikoli. Tato skutečnost vede podle autora k potvrzování současných praktik. Ve studiích spravedlivosti se bez hodnotových východisek nemůžeme obejít. Blíže se diskutím o dopadech výzkumu na vzdělávací politiku věnuje kapitola 5.

Pokusy o identifikaci úspěšných systémů

V rámci analytických aktivit OECD bylo učiněno několik pokusů identifikovat charakteristiky systémů, které dosahují dlouhodobě dobrých výsledků. Nejdůkladnější analýza (OECD 2004b) se zaměřila na reformní kroky v zemích, které dosáhly dobrých výsledků ve výzkumu PISA 2000. I když identifikovala některé společné trendy ve všech systémech (např. zvyšování autonomie, snaha o kontrolu výstupů apod.), nepřinesla očekávaný seznam kroků, které vedou k efektivitě vzdělávacích systémů.

V letech 2007 a 2010 získaly velkou pozornost analýzy konzultantské společnosti McKinsey, která ve spolupráci s OECD a odborníky z dalších renomovaných institucí učinila pokus určit faktory, které ovlivňují úspěšnost vzdělávacích systémů v mezinárodních komparativních výzkumech. V analýze z roku 2007 (McKinsey 2007) byly analyzovány společné charakteristiky systémů, které trvale dosahovaly nadprůměrných výsledků ve výzkumu PISA. Analýza ukázala, že charakteristická pro všechny tyto systémy je snaha 1.) získat pro povolání učitele ty nejlepší studenty, 2.) vybavit je potřebnými vědomostmi a dovednostmi v rámci počátečního studia i systematického dalšího vzdělávání a 3.) zajistit, aby systém poskytoval co nejlepší vzdělání každému dítěti.

V roce 2010 byla publikována druhá analýza společnosti McKinsey, která využívala dat z výzkumu PISA, ale její záběr byl daleko širší. Výzkumníci použili škálu vytvořenou agregací škál z šetření PISA a na tuto škálu přepočítali i výsledky dalších šetření v rozmanitých vzdělávacích systémech z celého světa. Pro detailnější analýzy pak zvolili 17 systémů, které dosáhly v uplynulém období výrazného zlepšení (bez ohledu na to, zda aktuálně dosahované výsledky byly v mezinárodním srovnání nadprůměrné či podprůměrné). Výsledky analyzovaných systémů se pohybovaly v 5 pásmech: neuspokojivé (např. Ghana, Brazílie) uspokojivé (např. Chile, Argentina, Jordánsko), dobré (např. Litva, Polsko, Německo, Slovinsko), výborné (např. Hongkong, Singapur), vynikající (Finsko). Analyzované systémy zároveň zahrnovaly systémy s nízkými i vysokými investicemi do vzdělávání. Výzkumníci navštívili zlepšující se systémy a dotazovali se tvůrců vzdělávacích politik na příčiny zlepšení. Celkem navštívili 200 tvůrců vzdělávacích politik a získali seznam 575 opatření. Následně formulovali doporučení pro vzdělávací systémy v různých stadiích rozvoje (McKinsey 2010):

1. stadium – osvojení základů matematiky a mateřského jazyka: podpora nedostatečně vzdělaných učitelů, zajištění základních potřeb žáků, zajištění minimální úrovně pro všechny školy.
2. stadium – konsolidace základů systémů: získávání kvalitních dat, vytvoření systému kontroly škol, koncepce systému financování, vytvoření pevné organizační struktury, hledání konsensu o pedagogickém modelu.
3. stadium – péče o profesi učitele a ředitele: tvorba systému dalšího vzdělávání a profesního postupu.
4. stadium – přesun péče o kvalitu z centra do škol: důraz na poskytování kolegiální zpětné vazby, podpora inovací a experimentu.

Analýza ukázala, že ve stadiu 4 měly země, ve kterých měly školy vyšší autonomii, lepší vzdělávací výsledky než školy s nižší autonomií. Ve stadiu 3 tomu bylo však naopak. Tento výsledek napovídal, že pokud se přistoupí ke zvyšování autonomie dříve, než učitelé a ředitelé dosáhnou určité profesní úrovně, může to paradoxně kvalitu vzdělávání zhoršit.

Výsledky studií McKinsey je možno srovnat s obecnými doporučeními, která na základě identifikace úspěšných vzdělávacích systémů a jejich charakteristik formuloval F. Whelan (Whelan 2009):

- Mít méně (lepších) učitelů.
- Získat pro profesi učitele ty správné lidi.
- Zajistit, aby každá škola měla efektivní vedení.
- Stanovit vysoké standardy a zjišťovat, zda jsou dosahovány.
- Vytvořit takové struktury, které podporují iniciativu, zodpovědnost a spolupráci.
- Investovat do rozvoje profesních znalostí a dovedností.
- Potírat nerovnosti ve vzdělávání.

Analýzy McKinsey jsou nicméně kritizovány mnohými členy akademické obce pro příliš zjednodušující přístup. Podrobněji se kritikou analýz a interpretací dat z výzkumu PISA zabývá kapitola 5.

4.2.2 Mezinárodní výzkumy a výzkumy efektivity školního vzdělávání

Historie výzkumů efektivity školního vzdělávání

V 80. a 90. letech minulého století byly mezinárodní srovnávací výzkumy považovány za významnou součást výzkumů efektivity školního vzdělávání (*school effectiveness research*)⁵⁴. Diskuse o mezinárodních výzkumech výsledků vzdělávání probíhala značnou měrou v Británii a šla ruku v ruce s diskusí o výzkumech efektivity školního vzdělávání a o metodách a dopadech národního testování, které již v té době bylo velmi masivní (např. Goldstein 2004). Později se mezinárodní výzkumy výsledků vzdělávání vydělily z výzkumů efektivity jako samostatný „pragmatický“ proud, přičemž výzkumy efektivity školního vzdělávání si zachovaly svůj akademický přístup a od pragmatičnosti mezinárodních výzkumů se „distancovaly“. Významná část výzkumníků, věnujících se efektivitě vzdělávání, se snaží využívat dat, která poskytují mezinárodní výzkumy výsledků vzdělávání nebo s nimi spojené národní výzkumy, k hledání odpovědí na své výzkumné otázky. Snaha poučit se o efektivitě vzdělávání zároveň stála na počátku vzniku mezinárodních výzkumů. Část kritiky, které čelí mezinárodní výzkumy, je společná i výzkumům efektivity vzdělávání. Pro porozumění úloze mezinárodních výzkumů výsledků vzdělávání v rozvoji zkoumání efektivity vzdělávání a pro porozumění kritice mezinárodních výzkumů, již se zabývá kapitola 5, je užitečné se poučit o vývoji obou proudů a o jejich vzájemné komunikaci.

Výzkumy efektivity školního vzdělávání a mezinárodní výzkumy vědomostí a dovedností se nevyvíjely zcela nezávisle – byly mezi nimi rozmanité překryvy a interakce, a přestože probíhaly paralelně, vzájemně se obohacovaly. Důležité bylo, že mezinárodní výzkumy mobilizovaly zdroje pro pedagogický výzkum. Tvůrcům vzdělávacích politik pomocí jednoduchých a srozumitelných výstupů ukázaly, že vzdělávání je důležité a že má velký smysl o něm získávat relevantní infor-

⁵⁴ Nyní hovoříme o *education effectiveness research* a *school improvement research*, o rozdílech mezi těmito dvěma proudy pojednáváme dále.

mace, na jejichž základě bude možno zvyšovat jeho efektivitu. Tvůrci vzdělávacích politik tedy začali být ochotni pedagogický výzkum financovat a v některých případech jej též iniciovali. Odvrácenou stranou politického a mediálního zájmu o výsledky výzkumů však byla častá zjednodušení a misinterpretace spočívající v publikaci prostých srovnání prostřednictvím žebříčků a v nabídkách prosté aplikace jednoduchých opatření bez zohlednění kulturních a společenských kontextů vzdělávacích systémů. Přes rozličná pochybení vedly však získané poznatky k tomu, že se tvůrci vzdělávacích politik začali zamýšlet nad tím, co by mohli dělat lépe. Možnost seznámit se s charakteristikami a výsledky jiných systémů podnítila tvůrce vzdělávacích politik k tomu, aby investovali do výzkumu efektivit vzdělávání.

Z hlediska analytického potenciálu mají mezinárodní databáze velkou výhodu v tom, že mezi hodnotami proměnných je daleko větší rozptyl než při národních měřeních a výsledky mají tudíž větší statistickou sílu. Při práci s databázemi tak bylo možno zkoumat nejen statistickou významnost závislostí, ale i jejich povahu. Sekundární analýzy získaných dat umožnily oddělit obecné tendence od charakteristik národně specifických.

Výzkumy efektivit školního vzdělávání naopak prospívaly mezinárodním výzkumům teoretickými konstrukty. V posledních cyklech výzkumů je věnována rostoucí péče konstrukci doprovodných dotazníků v souladu s teoretickými modely tak, že umožňují zkoumat například dopady klimatu a chování učitelů na učení žáků. Zároveň byl rozpoznán význam zachování konzistentnosti faktorů přes jednotlivé cykly, aby bylo možno zkoumat dlouhodobé trendy. Výzkumy efektivit vzdělávání také upozornily na limity používání průřezových dat pro kauzální závěry, což vedlo k metodologickým inovacím, kterými se zabýváme dále. Ovlivňování mezi oběma disciplínami tedy bylo bezesporu užitečné a je třeba, aby i nadále pokračovalo.

Výzkumy efektivit školního vzdělávání se začaly rozvíjet v 70. letech minulého století poté, co Coleman a kol. (1966) ukázali, že pouze malý podíl rozptylu ve výsledcích žáků je vysvětlitelný školou a že tedy zcela zásadní podíl na výsledcích žáků má jejich rodinné zázemí. V návaznosti na to se výzkumníci snažili rozkrýt, jak vliv školy na výsledky žáků zvýšit. S tím souvisela zásadní otázka, co podmiňuje kvalitní učení a vyučování (Kyriakides a Charalambos 2014).

Výzkum v 80. letech se tedy zaměřil na potvrzení skutečnosti, že školy mohou být důležité a že některé školy jsou efektivnější než jiné. Postupně si však kladl čím dál vyšší cíle. V dalších letech se pozornost zaměřila na pokusy rozdíly mezi školami kvantifikovat. V 90. letech následovala snaha zjistit, co přispívá k učení žáků, tedy identifikovat faktory, které vysvětlí rozdíly ve výsledcích žáků jednotlivých škol. Vzniklé modely ukázaly potřeby lepšího teoretického ukotvení. Přelom století se zaměřil na konstrukce a testování modelů, jejichž cílem bylo vysvětlit, jak faktory na jednotlivých úrovních – žák, třída, škola, systém přispívají k výsledkům žáků. V této fázi se jasně vyprofilovaly tři různé perspektivy (Kyriakides a Charalambos 2014).

První perspektiva je ekonomická. Ekonomové vztahují výstupy ke vstupům, přičemž provádějí kontrolu pro různé proměnné (*education production function*). Ekonomické práce jsou založeny na předpokladu, že lepší vstupy povedou k přírůstku ve znalostech.

Druhá perspektiva je sociologická a zaměřuje se na studium socioekonomického statusu a jeho dopadu na výsledky žáků a okolností, které školám umožňují vliv socioekonomického statusu kompenzovat. Vycházejí z předpokladu, že zmírnění znevýhodnění pomůže zlepšit výsledky nejen znevýhodněných žáků, ale celého systému.

Třetí perspektiva pochází z dílny psychologie vzdělávání a zabývá se tím, jak motivace žáků a jejich rozmanité schopnosti přispívají k učení a také jak může být efektivita učení zvýšena přístupem a chováním učitelů. Psychologická zkoumání postupně vedla k přesvědčení aktuálně

sdílenému mnoha odborníky, že faktory na straně učitele jsou nejdůležitější. Pozornost se postupně přesunula od učitelova chování k jeho vědomí a myšlení.

Změna zacílení výzkumů s sebou přirozeně přináší i vývoj metodologický. V první fázi se zkoumání omezilo na srovnávání více a méně efektivních škol, ve druhé a třetí fázi byly předmětem výzkumu studie kohort a nyní začíná být kladen čím dál tím větší důraz na výzkumy longitudinální. V posledních dvou fázích byly také studie přímých efektů obohaceny o studie efektů nepřímých. Začaly být dále vytvářeny modely přidané hodnoty.

Dynamický model efektivity vzdělávání

Příznačnou charakteristikou aktuální etapy je důraz na komplexnost sledovaných jevů. Výzkumníci vycházejí z poznání, že efektivita je dynamická a v čase se mění, mění se s žáky, učiteli a předměty. To znamená, že je třeba se zaměřit ne na statické jevy, ale na sledování růstu a změny. Tedy nezajímá nás pouze, jaké faktory podmiňují, že některé školy jsou lepší než jiné, ale zejména to, jaké faktory způsobují, že některé školy jsou schopny se dále vyvíjet a zlepšovat svoji práci.

Na volání po teoretickém ukotvení výzkumů efektivity vzdělávání a jejich současného zaměření reagovali například Creemers a Kyriakides (2010) modelem, který provázal výzkum efektivity vzdělávání (*education effectiveness research*) s výzkumem zlepšování školy (*school improvement research*). Výzkumy efektivity vzdělávání a zlepšování školy byly původně dva odlišné proudy, které se postupně sbližují, autoři však uvádějí, že jejich vztah je stále problematický. Výzkumy efektivity vzdělávání byly vždy zaměřeny na sledování výsledků vzdělávání a faktorů na úrovni třídy, školy a systému, které tyto výsledky ovlivňují. Zkoumání zlepšování školy se naopak zaměřovalo na procesy změny ve třídách a ve školách bez ohledu na výsledky vzdělávání. Výzkumy se věnovaly například studiu změny klimatu, přístupu učitelů a podobně, ale nemodelovaly ani nesledovaly dopad těchto jevů na výsledky žáků.

Autoři uvádějí, že výzkum efektivity vzdělávání může být nahlížen jako balík výzkumů v různých oblastech: výzkum chování učitelů, kurikula, diferenciace, organizace školy a vzdělávací politiky. Hlavní výzkumnou otázkou zde je, které faktory ve výuce, kurikulu a učebním prostředí na různých úrovních (třída, škola, komunita, systém) mohou přímo či nepřímo vysvětlit rozdíly ve výsledcích žáků, vezmeme-li v úvahu charakteristiky žáků, jako jsou jejich vrozené dispozice, vstupní výsledky a rodinné zázemí. Výzkumy efektivity vzdělávání se v posledních desetiletích značně zlepšily díky kritice výzkumného designu, výběrových postupů a statistických technik. Co se týče teoretického aspektu, zlepšilo se vymezení konceptů a vztahů mezi nimi, ale dlouho neexistoval teoretický model, který by propojoval jednotlivé prvky. Tento model je potřebný k tomu, aby zasadil uskutečněné výzkumy do jednotného rámce a zabránil tomu, aby se určité výzkumy realizovaly stále dokola. Dále ukazuje, kam je třeba zaměřit pozornost v budoucnu. Měl by také pomoci praktikům ve školách výsledky výzkumu lépe uchopit. Z výše uvedeného vyplynula potřeba tvorby modelu, který vezme v úvahu dynamickou povahu efektivity vzdělávání a poskytne oporu nejen výzkumníkům, ale zejména tvůrcům vzdělávací politiky a praktikům.

Autoři vycházejí z toho, že výzkumníci nemohou převzít od učitelů zodpovědnost za školní vzdělávání, ale že je jejich povinností přispět k efektivitě a zlepšování vzdělávání poskytnutím poznatkové základny pro praxi a pomocí školám navrhnout vhodné intervence. Vztah mezi vědou a praxí nebyl ve výzkumu efektivity vzdělávání (i jinde) vždy úspěšný.

Významným rysem vytvořeného modelu (Creemers; Kyriakides 2010) je výše zmiňovaná skutečnost, že je zaměřen na dynamickou povahu jevů, tedy nezkoumá aktuální stav, ale dopad intervencí. Jeho hlavním smyslem je usnadnit modelování skutečných intervencí (například sadu opatření na eliminaci šikany, která se uplatňuje na různých úrovních vzdělávacího systému). S tím souvisí souběžné působení více faktorů a také nelinearita uplatňování rozmanitých faktorů (v realitě neplatí, že čím více, tím lépe – například v případě oborově předmětových znalostí učitelů je nezbytná nějaká základní úroveň, s dalším růstem znalostí se již výsledky žáků nezlepšují). Navržený model je samozřejmě užitečný i pro analýzy dat z mezinárodních výzkumů.

4.2.3 Rozvoj metodologických postupů

Analytické přístupy

Snaha o maximální využití bohatých dat získávaných v rámci mezinárodních výzkumů a důraz na to, aby výsledky byly prezentovány a interpretovány korektním a nezávadějícím způsobem, vedly k rozvoji analytických postupů. Začaly se zdokonalovat analytické metody, které umožňují data analyzovat stále sofistikovanějším způsobem.

Pro analýzu dat jsou typicky využívány metody víceúrovňové analýzy, které umožňují pracovat s hierarchickou strukturou dat, tedy zohlednit skutečnost, že některé vlivy působí na úrovni žáka, jiné na úrovni třídy či školy a další na úrovni systému. O metodologickém vývoji, který vedl k využití víceúrovňového modelování při modelování dopadů faktorů na úrovni třídy a školy na výsledky vzdělávání a jejich přednosti, podrobně referuje např. Goldstein (1987) nebo Bryk a Raudenbush (2002), u nás Soukup (2006).

Kromě víceúrovňového lineárního modelování, které umožňuje vysvětlovat závislou proměnnou (typicky výsledky vzdělávání měřené testy) pomocí soustavy lineárních rovnic modelujících vlivy proměnných na úrovni žáka, školy, případně vzdělávacího systému a jejich interakcí, je čím dál tím častěji používáno víceúrovňové strukturní modelování. Předností víceúrovňového strukturního modelování je – kromě možnosti modelovat souběžné vlivy na jednotlivých úrovních – rovněž konstrukce latentních proměnných pomocí konfirmativní faktorové analýzy. Latentní proměnné zohledňují chybu měření a jsou proto spolehlivějšími indikátory než manifestní (přímo měřené) proměnné. Víceúrovňové strukturní modelování tedy kombinuje přednosti strukturního modelu, který sestává z části, která modeluje latentní proměnné na základě měřených proměnných (*measurement part*), a z části, která modeluje vztahy mezi proměnnými (*structural part*), a přednosti víceúrovňové analýzy, která umožňuje zohlednit hierarchickou povahu reality. Víceúrovňové strukturní modelování umožňuje tvořit komplexnější modely než víceúrovňové lineární modelování, neboť umožňuje modelovat i složité vztahy mezi proměnnými na jednotlivých úrovních⁵⁵.

Význam víceúrovňových metod spočívá v tom, že do jisté míry umožňují vyvozovat příčinné závěry z dat získaných jednorázovým pozorováním (*observational data*), neboť při hledání vztahu mezi závislou a nezávislou proměnnou umožňují zohlednit další vlivy na úrovni žáka, třídy a školy, tedy očistit vztah mezi závislou a nezávislou proměnnou o tyto vlivy. V rámci mezinárodních výzkumů jsou výzkumníci odkázáni na data získaná jednorázovým pozorováním

⁵⁵ Tato inovace je přirozeně provázena i určitými nevýhodami, jako je rostoucí složitost výpočtů a nutnost volby specializovaných programů.

a dostávají se do situací, ve kterých je obtížné vyvozovat kauzální závěry, tedy říci, že jedna proměnná podmiňuje druhou. Realita je často taková, že proměnná, o které předpokládáme, že je závislá, naopak částečně ovlivňuje proměnnou, u které předpokládáme nezávislost. Také se můžeme dostat do situace, že obě proměnné jsou ovlivňovány třetí proměnnou, kterou jsme v modelu opomínali (problémy endogenity nebo reverzní kauzality). Jak jsme se zmiňovali v předchozí kapitole, další nebezpečí spočívají v podhodnocení standardních chyb.

Nejčastějším problémem průřezových studií je však skutečnost, že zpravidla nepozorujeme všechny vlivy, které by měly být zohledněny (problém nepozorované heterogenity). Vztah mezi proměnnými, očistěný od měřených vlivů, tedy stále nemusí umožňovat činit správné závěry o příčinách a následcích. Je zřejmé, že faktorů, které by měly být zkoumány ve vzdělávací politice, je opravdu mnoho, a pravděpodobnost, že nějaký neměříme, je velmi vysoká. Typicky třeba vrozené schopnosti studentů, učitelské schopnosti učitelů, preference rodičů nebo hodnoty spolužáků.

Tento oddíl popisuje vybrané metody, které nějakým způsobem řeší problémy endogenity a nepozorované heterogenity. Tyto metody jsou používány v pedagogickém výzkumu, ale mají svůj původ v ekonomii (Schlotter a kol. 2014).

Poslední dobou se uplatňuje metoda, jak obejít výše uvedené obtíže pomocí hledání identických respondentů, kteří byli vystaveni rozdílnému působení zkoumané proměnné (*matching*). Tato metoda odstraňuje problém s endogenitou, nicméně neodstraňuje problém opomínutých proměnných.

Všechny zde uvedené problémy řeší nejlépe technika randomizovaného experimentu. V experimentu přidělujeme respondenty náhodně do skupiny, která je vystavena nějaké intervenci, a do skupiny kontrolní. Pokud jsou skupiny dost velké a přiřazení náhodné, jsou vlastnosti obou skupin identické. Změny ve výsledku jsou pak způsobeny intervencí a můžeme je přímo pozorovat. Například ve známém projektu podpory předškolních dětí ze znevýhodněného prostředí *Perry Preschool Programme* byly znevýhodněné děti náhodně přiřazeny do dvou skupin, z nichž v jedné se dětem dostalo podpory a ve druhé nikoli. Děti byly dlouhodobě sledovány a bylo vyhodnocováno, jak se liší úspěšnost ve škole a v životě u dětí, kterým se dostalo podpory, a u těch, jimž se této podpory nedostalo.

Experimenty však často nejsou implementovány rigorózně, v průběhu experimentu například dochází k přemísťování zkoumaných osob mezi experimentální a kontrolní skupinou nebo se stává, že chování osob je ovlivněno vědomím, že jsou zařazeny do experimentu. V pedagogice je dále často obtížné kontrolovat podmínky experimentu. Zde se často setkáváme s tím, že zajímavé výzkumné problémy nelze z etických, praktických i ekonomických důvodů studovat pomocí randomizovaného experimentu.

Někdy však můžeme využít situace, která takové experimenty vytváří přirozeně. Například v situacích, kdy je nějaké přiřazení realizováno na základě loterie, mělo by být rozdělení respondentů náhodné. Můžeme tedy zkoumat ty, kteří v loterii uspěli, a ty, kteří nikoli, ale musíme brát v úvahu, že výsledky můžeme zobecnit pouze na populaci, která se do loterie zapojila, protože může mít jiné společné vlastnosti, než ti, kteří se nezúčastnili.

Skutečnost, že nemůžeme postupovat prostřednictvím randomizovaného experimentu, obcházíme několika dalšími technikami, které umožňují výzkumně využít situace, které nastávají zcela přirozeně. Jednou z nich je technika instrumentálních proměnných. Tato technika vychází z předpokladu, že rozptyl výsledné proměnné, která nás zajímá, je složen z několika částí. Část rozptylu je způsobena proměnnými, které nebyly identifikovány. Část nicméně může být použita ke stanovení kauzálních závěrů. V této situaci se snažíme oddělit rozptyl,

jenž je způsoben touto (instrumentální) proměnnou, která není korelována s výsledkem ani s nezjišťovanými proměnnými. Musíme tedy najít vhodnou instrumentální proměnnou, která není korelována s výsledkem jinak, než prostřednictvím sledovaného efektu. Schlotter a kol. (2014) jako příklad analýzy dopadu uvádějí prodloužení povinné školní docházky na výši příjmu. V případě délky povinného vzdělávání můžeme předpokládat, že není korelována s výši příjmu jinak než prostřednictvím získaných vědomostí a dovedností. Je to tedy vhodná instrumentální proměnná. Podrobnější rozbor situace však dobře ukazuje, že i zde musíme být obezřetní, protože délka povinného vzdělávání je rozhodující pouze u těch občanů, kteří odcházejí ze školy po ukončení povinného vzdělávání a nepokračují dále ve studiu.

Dalším příkladem využití přirozených situací je regresní diskontinuita. Týká se specifických situací, kdy účast, respektive neúčast v nějakém programu záleží na tom, zda překonáme nějakou ostrou hranici. Sledování výzkumných objektů, které se nacházejí buď těsně pod touto hranicí nebo těsně nad ní, simuluje randomizovaný experiment. Vycházíme z toho, že mezi objekty, které jsou v těsné blízkosti hranice, by neměl být žádný systematický rozdíl. Jediný rozdíl spočívá v tom, zda se jim podařilo hranici překonat a tudíž se jim dostalo nějaké specifické intervence (například nějaké podpory). Dlouhodobým sledováním skupin těsně pod hranicí a těsně nad hranicí můžeme dopad této intervence vyhodnotit. Zde opět vycházíme z toho, že situace probíhá podle pravidel. Pokud by mohlo nastat, že některé objekty překonají hranici podvodem, budou vznikat systematické rozdíly mezi těmi, kteří jsou schopni/ochotni podvádět, a těmi, kteří nikoli. Příkladem takové situace může být například podpora škol, které mají určitý podíl znevýhodněných žáků a sledování dopadu této podpory na výsledky žáků či jiné aspekty práce školy.

Na panelových datech můžeme často aplikovat metodu DiD (*difference in difference*). V tomto přístupu sledujeme dvě skupiny, z nichž jedna je vystavena sledovanému vlivu a druhá ne. Na rozdíl od randomizovaného experimentu v této situaci nepředpokládáme, že zmíněné dvě skupiny jsou identické. Předpokládáme však, že všechny změny ve výsledku jsou způsobeny sledovanou intervencí. Tedy předpokládáme, že kdyby nedošlo k námi sledovanému působení, rozdíly ve výsledku obou skupin by zůstaly ve sledovaném čase stejné. Shledaný rozdíl mezi předpokládaným a skutečným výsledkem (neboli mezi skupinami na počátku experimentu a v jeho závěru) přičítáme působení sledovaného faktoru. Tímto způsobem můžeme například sledovat efektivitu výuky prostřednictvím nějaké nové metody, kdy měříme rozdíly ve výsledcích ve sledované oblasti u obou skupin na začátku experimentu a na konci experimentu, a rozdíl v přírůstcích přisuzujeme dopadu metody.

Longitudinální designy s výsledky jako závislou proměnnou

Ve společenskovědním výzkumu se pro odstranění nebezpečí reverzní kauzality často používají longitudinální designy, kdy sledujeme respondenty dlouhodobě, a máme tedy opakované informace o dosahovaných výsledcích, ale i o vlivech, které na respondenta působí a výsledky ovlivňují.

V poslední době řada vzdělávacích systémů navazuje na mezinárodní výzkumy vlastními longitudinálními studiemi, ve kterých je sledován přírůstek ve výsledcích žáků a tento přírůstek je vztahován k rodinnému zázemí žáků a k charakteristikám škol a učitelů. S těmito velmi nákladnými výzkumy je identifikace příčinných vztahů daleko jednoznačnější a zřejmější než v průřezových datech z mezinárodních výzkumů.

K využívání mezinárodních výzkumů jako základu k longitudinálním výzkumům začíná

docházet i v České republice. Např. konsorcium NHÚ ČR, NVF a ÚVRV PedF UK využilo prostředků investovaných v ČR do výzkumů TIMSS a PIRLS v roce 2011 a založilo panelový výzkum na skupině žáků, kteří byli v rámci tohoto mezinárodního výzkumu testováni. Třídy testované v roce 2011 byly dotazovány na konci 5. ročníku s cílem zjistit, jací žáci se ucházeli o studium na víceletém gymnáziu a s jakým výsledkem; dále byly opětovně testovány v roce 2012 v 6. ročníku ZŠ a v roce 2016 v 9. ročníku ZŠ. Zároveň byly prostřednictvím dotazníkových šetření zjišťovány rozmanité informace o rodinném zázemí žáků, školách a učitelích. Spolu s tím jsou sledovány gymnaziální třídy, do kterých odešli někteří z respondentů TIMSS a PIRLS 2011, aby bylo možno srovnat vzdělávací pokrok žáků v základních školách a víceletých gymnáziích. Po přechodu do středního vzdělávání bude vytvořen nový panel žáků různých programů středoškolského studia (gymnázií, odborného maturitního a odborného nematuritního studia)⁵⁶.

Německo provedlo v rámci šetření PISA 2000 a PISA 2003 longitudinální šetření na úrovni škol. V roce 2000 byl proveden nadvýběr škol (celkem bylo vybráno 1500 škol) tak, aby byla docílena reprezentativita za jednotlivé spolkové země, v roce 2003 bylo vybráno opět 1500 škol, přičemž 500 škol se v obou výběrech překrývalo. Tento vzorek samozřejmě nebyl reprezentativní, ale umožňoval studovat stabilitu výsledků v jednotlivých školách, případně dopad opatření aplikovaných na úrovni školy. Další příklady využití mezinárodních výzkumů k národním výzkumným aktivitám v Německu uvádí kapitola 5.2.2.

Mezinárodních výzkumů v té podobě, v jaké jsou běžně realizovány, sice není možno využít k longitudinálnímu sledování žáků, tříd ani škol, je však možno je díky měření dlouhodobých trendů využít k longitudinálnímu sledování systémů. V posledních letech se vyskytují zdařilé analýzy, které cyklickou povahu výzkumů používají k vyvozování kauzálních závěrů, a to tak, že vysvětlují změny ve výsledcích žáků v jednotlivých kolech výzkumu prostřednictvím faktorů, v nichž došlo ke změnám. Studie se dělí na ty, které studují dopady faktorů na straně rodiny a školy, a ty, které studují dopady institucionálních faktorů. Ty první jsou logicky založeny na zkoumání změn a rozdílů uvnitř jednotlivých zemí, ty druhé jsou naopak založeny na studiu rozdílů mezi zeměmi.

Analýzy z první skupiny například využívají skutečnosti, že v jednom vzdělávacím systému je řada proměnných stabilní, tedy analýza není negativně ovlivněna tím, že tyto proměnné neumíme kvantifikovat. Velmi pěkným příkladem takové práce je studie Nilsenové a Gustafssona (2014), která se snaží vysvětlit zlepšení výsledků norských žáků ve výzkumu TIMSS pomocí změn, ke kterým došlo v norských školách. Autoři ukázali, že zlepšení ve výsledcích je možno vysvětlit změnami ve vnímání bezpečnosti školy a ve školním důrazu na akademický úspěch (ukazatel *school emphasis on academic success* SEAS charakterizující vnímání profesních schopností ze strany učitelů, důvěru učitelů v žáky a v jejich rodiče a akademická očekávání učitelů). Studie použila kvazilongitudinální design, ve kterém sice nebyli sledováni stejní žáci, ale byly sledovány stejné věkové kohorty v jednom vzdělávacím systému, a prostřednictvím doprovodných dotazníků byly zároveň zachyceny změny v přístupu škol.

Mezinárodní výzkumy jsou považovány za velmi cenné pro sledování vlivů, které není možno identifikovat v rámci jednoho vzdělávacího systému. Jako příklad takového vlivu bývá často uváděno například plošné testování. Vzhledem k tomu, že plošné testování je zpravidla aplikováno na celé systémy, není možno sledovat v rámci jednoho systému jeho dopady (vyjma situace před a po jeho zavedení). Stejně tak není možno v rámci jednoho systému sledovat například dopady povinného předškolního vzdělávání, vysoké diferenciace systému, mechanismu při-

⁵⁶ Více informací o projektu lze získat na <http://pages.pedf.cuni.cz/uvrv/close/>.

dělování učitelů do škol, možnosti svobodné volby základní školy, rozsahu privátního školství a podobně. I když v rámci systému existují nějaké rozdíly (např. mezi regiony), jsou zpravidla daleko nižší než rozdíly mezi zeměmi. V jiných situacích, například při zkoumání soukromých a veřejných škol, sice je možno zkoumat soukromé a veřejné školy v rámci jednoho systému, ale narážíme na problémy s výběrem respondentů. Žáci, kteří navštěvují soukromé školy, se často významně liší od žáků, kteří navštěvují školy veřejné, a to v měřitelných i neměřitelných charakteristikách. Tedy, i když můžeme při srovnání zohlednit řadu měřitelných rozdílů, stále existuje nebezpečí neměřitelných rozdílů, které se zohlednit nepodařilo.

Může se také stát, že přítomnost soukromých škol ovlivní chování veřejných škol, které se nacházejí v okolí a soupeří o stejné žáky. V důsledku tohoto jevu může být prosté porovnání soukromých a veřejných škol zavádějící, neboť neumožňuje zohlednit ovlivnění celého systému přítomností soukromých škol a soutěží mezi soukromými a veřejnými školami. Při agregaci na úroveň systému již tento jev zpravidla zachytit můžeme, neboť můžeme porovnávat systémy s vysokým a nízkým podílem soukromých škol (Schlotter a kol. 2014).

Ve své často citované práci ukázali Hanushek a Woessmann (2011), že země, které aplikují systémy plošného testování (s dopady na školy), mají svobodnou volbu školy a větší lokální autonomii, mají obecně lepší výsledky než země, které tyto postupy neuplatňují. Zároveň ale například ukázali, že lokální rozhodování o platech učitelů je účinné pouze v situaci, kdy je práce učitelů hodnocena systematicky a na základě vhodných kritérií.

I když výsledky podobných studií jsou prezentovány velmi přesvědčivě, je velmi obtížné se ujistit, že byly ošetřeny všechny kulturní a jiné systematické rozdíly, které mohou výsledky zkreslovat. S velkou opatrností je například třeba přistupovat k často citovaným poznatkům, že průměrně vyšší počet žáků ve třídě vede k lepším výsledkům. Vyšší počet žáků ve třídách mají tradičně asijské systémy, které ve srovnávacích tabulkách obsazují jedny z nejvyšších příček. Jejich úspěchy však jsou zapříčiněny primárně jinými faktory než počtem žáků ve třídě (tradičně vysokou disciplínou, všudypřítomným privátním doučováním, významem prisuzovaným vzděláváním, přísnými přijímacími zkouškami na vyšší stupeň škol a podobně).

Výše uvedené studie jsou výjimečné v tom, že jsou umožněny velkými mezinárodními výzkumy – nelze je realizovat pouze s národními daty. Proto představují důležitý doplněk k národním výzkumům. Vzdělávací systémy jsou však velmi složité mechanismy, které jsou pevně ukotveny v kulturách svých zemí. Jednoduché interpretace, které jejich organizátoři mnohdy nabízejí, nemusí být vždy správné a nemusí zachycovat skutečné příčiny, které stojí za vzdělávacími výsledky žáků. Proto je třeba přistupovat k prezentovaným výsledkům s maximální obezřetností.

Longitudinální designy s výsledky jako nezávislou proměnnou

V longitudinálních analýzách, jež jsou diskutované výše, figuruje výsledek žáků jako závislá proměnná. Analýzy se snaží identifikovat faktory, které ovlivňují výsledky žáků. Velmi zajímavé se však ukazují rovněž analýzy, které pracují s výsledky žáků (a dospělých – v případě mezinárodních výzkumů výsledků vzdělávání dospělých) jako s nezávislou proměnnou. V případě průřezových výzkumů, jakým byl například výzkum vědomostí a dovedností dospělých PIAAC, bylo zkoumáno, jak kognitivní dovednosti souvisejí s příjmy, s pracovní pozicí, s dalším vzděláváním, s nekognitivními výstupy typu občanská angažovanost, důvěra ve společnost či subjektivní pocit zdraví. Pro vyvozování kauzálních závěrů realizují vzdělávací systémy návaz-

ná longitudinální šetření, která jim umožňují určit, jak významným prediktorem rozmanitých aspektů životního úspěchu jsou kognitivní dovednosti.

Asi nejznámějším a nejrozsáhlejším longitudinálním výzkumem, který navázal na mezinárodní výzkum výsledků vzdělávání žáků, byl kanadský projekt *Youth in Transition Survey (YITS)*. Ten sledoval žáky, kteří absolvovali šetření PISA, v průběhu dalšího vzdělávání a při přechodu na pracovní trh (OECD 2010b). Kladl si za cíl zmapovat rozhodovací procesy po ukončení povinného vzdělávání a další vzdělávací a pracovní dráhy mladých lidí a porozumět tomu, jaké faktory je ovlivňují. Výzkumníci se zajímali, jak ovlivňují životy mladých lidí rozmanité faktory na straně školy, a snažili se identifikovat vzdělávací dráhy, které vedou k hladkému přechodu na pracovní trh. Studovali úlohu vzdělanostních a profesních aspirací, mechanismy jejich utváření a způsoby získávání pracovních zkušeností. Výzkumníci dále zjišťovali, s jakými vědomostmi, dovednostmi a pracovními zkušenostmi vstupují mladí lidé na pracovní trh a jak se liší situace rozmanitých skupin, zejména mladých lidí, kteří jsou nějakým způsobem znevýhodněni.

Výzkum YITS byl zahájen v roce 2000 spolu s první vlnou výzkumu PISA. Žákovský dotazník byl rozšířen o otázky týkající se vzdělanostních a profesních aspirací a očekávání a počátečních pracovních zkušeností. Kromě žákovského dotazníku byl administrován rovněž dotazník pro rodiče, ve kterém rodiče poskytli doplňující informace o rodinném zázemí žáků. Rodičovský dotazník byl administrován po telefonu. Žáci byli opětovně dotazováni ve dvouletých intervalech prostřednictvím telefonického dotazování. V roce 2000 se do výzkumu zapojilo 30 000 patnáctiletých žáků z 1 200 škol.

Výhodou studie bylo kromě možnosti provazovat faktory měřené v určitém čase s faktory měřenými v budoucnosti a tak vyvozovat kauzální souvislosti také časté dotazování respondentů, které významně zlepšuje kvalitu získaných dat. Respondenti zde byli dotazováni na události z nedávné minulosti, které si dobře pamatují. Výzkumy ukazují, že výpovědi o motivacích, postojích a chování jsou jiné, jsou-li zjišťovány retrospektivně, než jsou-li zjišťovány v době, kdy nás zajímají.

Přínosy kanadské studie často argumentovali představitelé OECD, když se snažili získat zástupce členských zemí v PISA BPC pro realizaci mezinárodního longitudinálního šetření. Takové šetření se však ukázalo být natolik nákladné, že je členské země nebyly ochotny financovat.

Kanada nicméně není jedinou zemí, která provádí návazné národní longitudinální šetření. Longitudinální sledování žáků, kteří se zúčastnili výzkumu PISA 2000, provádí rovněž Dánsko⁵⁷. Dánsko zároveň získalo zajímavá longitudinální data tím, že do výzkumu vědomostí a dovedností dospělých PIAAC v roce 2012 zařadilo 1881 mladých lidí, kteří se zúčastnili výzkumu PISA v roce 2000 a v roce 2012 jim tedy bylo 27 let⁵⁸. První analýzy ukázaly souvislost mezi dobrými čtenářskými dovednostmi v roce 2000 a dobrými čtenářskými dovednostmi v roce 2012, absolvováním terciárního vzdělání, dokončením zvolené vzdělávací dráhy, nepobíráním sociálních dávek a pozdějším manželstvím a rodičovstvím.

Austrálie zahájila v roce 2003 výzkum *Longitudinal Survey of Australian Youth (LSAY)*, který sleduje žáky, kteří se zúčastnili výzkumu PISA v letech 2003, 2006 a 2009 po dobu zhruba 10 let, tedy od konce povinného vzdělávání do samostatného pracovního života. Výzkum zjišťuje, jak se mladí Australané rozhodují o své další vzdělávací a pracovní dráze, a umožňuje monitorovat,

⁵⁷ Data spolu s popisem proměnných jsou k dispozici na <http://cssr.surveybank.aau.dk/webview/>.

⁵⁸ <http://eng.uvm.dk/News/~UVM-EN/Content/News/Eng/2014/140617-PISA-PIAAC-results>

jak se chování stejné věkové skupiny mění v čase. Kromě vzdělávací a pracovní dráhy a jejich prediktorů sleduje rovněž životní spokojenost mladých lidí. Analýzy dat ukázaly, že solidní úroveň matematických a čtenářských dovedností predikuje účast v terciárním vzdělávání, zaměstnanost a dobrý příjem. Dále ukázaly, že zapojení do života školy a sebepojetí predikují pravděpodobnost ukončení střední školy větší měrou než rodinné zázemí. Motivace měla významný dopad na životní úspěch také u žáků, kteří dosáhli v matematickém testu PISA špatných výsledků (Thomson; Hillman 2010).

Longitudinální výzkum navazující na výzkum PISA realizovala i Česká republika. V rámci výzkumu PISA-L, jenž realizoval v letech 2003–2007 tým Sociologického ústavu AV ČR, byli sledováni všichni žáci, kteří se zúčastnili šetření PISA v roce 2003. Nárazné šetření probíhalo v domácnostech prostřednictvím tištěného dotazníku, emailem nebo telefonickým dotazováním dle volby respondentů. Bylo zjišťováno, jak kognitivní dovednosti žáků ovlivnily jejich další vzdělávací dráhu (např. Straková 2010a). Vzhledem k časově omezenému financování bylo možno žáky sledovat pouze do přechodu do terciárního vzdělání, případně na trh práce po ukončení střední školy. I když se organizátoři snažili zvýšit návratnost věcnými odměnami a loteriemi, do kterých automaticky vstupovali ti, co se šetření zúčastnili, došlo v době trvání výzkumu ke značnému úbytku respondentů, kteří studovali v odborném vzdělávání, zejména v učebních oborech. Výsledný datový soubor mohl velmi dobře posloužit ke studiu přechodu do terciárního vzdělávání a vzdělávacích drah v případě gymnazistů. Nepodařilo se však naplnit původní očekávání, že poskytne detailní informace o vazbě mezi výsledky ve věku 15 let, středoškolským vzděláváním a přechodem na pracovní trh pro absolventy různých oborů středního vzdělávání.

ČR navázala longitudinálními výzkumy i na další mezinárodní výzkumy, a to s cílem sledovat následné vzdělávací přechody a návazné vzdělávání a jeho dopady. V již zmiňovaném výzkumu CLoSE bylo například zjišťováno, do jaké míry výsledky TIMSS předurčují, zda se žák uchází v 5. ročníku o studium na víceletém gymnáziu a jak je úspěšný v přijímacím řízení. Analýza ukázala, že kromě kognitivních dovedností je významným prediktorem rodinné zázemí, tedy že děti se stejnými výsledky (a stejným školním prospěchem), jež pocházejí ze vzdělanějších rodin, mají významně vyšší šanci, že je rodiče přihlásí k přijímacím zkouškám a že u přijímacích zkoušek uspějí (Straková; Greger 2013). O těchto poznatcích pojednává podrobněji kapitola 6.

V návazném výzkumu na výzkum PIAAC, který realizuje Národní vzdělávací fond v rámci projektu GAČR, je sledováno, jak kognitivní výsledky predikují účast v dalším vzdělávání a pracovní kariéru mladých dospělých.

Stejně jako v případě studia faktorů ovlivňujících výsledky žáků, i v případě studia kognitivních schopností jako prediktorů se významně uplatňují přednosti vzájemného porovnávání vzdělávacích systémů. Známe jsou například ekonomické studie dopadu úrovně lidského kapitálu (aproximované kognitivními výsledky) na makroekonomické výstupy, jako je například ekonomický růst. V těchto analýzách byly kombinovány výsledky z různých mezinárodních výzkumů. Tyto výzkumy zahrnovaly různé skupiny zemí a respondenti byli testováni různými testy, jejichž výsledky není možno jednoduše propojit. Hanushek a Woessmann (2009) provedli propojení přes jednotnou škálu prostřednictvím výsledků Spojených států, které se účastnily všech šetření a zároveň byly jejich výsledky propojeny s Národním hodnocením vzdělávacího pokroku (NAEP). Prostřednictvím NAEP bylo možno upravit výsledky amerických žáků pro jednotlivá mezinárodní hodnocení (podle věku a testovaného předmětu) a vytvořit časovou řadu výsledků amerických žáků. Na základě těchto výsledků byly pak na stejnou škálu trans-

formovány výsledky všech zemí, které se aspoň jednou zúčastnily mezinárodního šetření. Průměrné výsledky jednotlivých zemí byly pak vztaženy k průměrnému ekonomickému růstu (HDP na hlavu) pro 50 zemí pro léta 1960–2000. To umožnilo odhadnout, jak moc by se zvýšil hrubý domácí produkt s přírůstkem kognitivních dovedností. Pro tvůrce vzdělávacích politik tyto výzkumy přinesly významnou evidenci, že je třeba implementovat opatření na podporu zlepšování vědomostí a dovedností žáků. Podle autorů analýz demonstruje skutečnost, že bylo možno odhalit takové zásadní vazby, nedocenitelnou hodnotu mezinárodních výzkumů výsledků vzdělávání (Hanushek; Woessmann 2013).

Mezinárodní výzkumy výsledků vzdělávání umožňují odpovídat i na takové otázky, jak struktura ekonomiky ovlivňuje odměňování jedinců s rozličnými kognitivními dovednostmi. Tradiční výzkumy v žákovské populaci neumožňují zkoumat význam kognitivních dovedností pro uplatnění na trhu práce. Jeden z nejzajímavějších výsledků analýz výsledků výzkumů v dospělé populaci je, že jednotlivé ekonomiky si cení kognitivních schopností velmi rozdílně, tedy že v některých zemích jsou velké rozdíly mezi platy zaměstnanců s vysokou a nízkou úrovní matematických a čtenářských dovedností. V jiných zemích mají kognitivní dovednosti na výši příjmů jen malý vliv (OECD 2013c).

4.2.4 Využití dat českými výzkumníky

Stejně jako v zahraničí, i v ČR jsou data z mezinárodních výzkumů využívána ekonomy, sociology i pedagogickými výzkumníky. V posledních letech zájem českých výzkumníků o mezinárodní výzkumy a v nich získaná data stoupá. V odborných časopisech jsou nejčastěji zastoupeny analýzy dat z výzkumu PISA. Autoři využívají skutečnosti, že datové soubory obsahují kvalitní indikátor rodinného zázemí (index ekonomického, sociálního a kulturního statusu *escs*, do kterého vstupuje vzdělání a povolání obou rodičů, počet knih v domácnosti a další kulturní a ekonomické statky v rodině), a proto je možno data využívat ke sledování souvislostí mezi rodinným zázemím a výsledky žáků. Čeští autoři využívají ovšem také skutečnosti, že výzkumy asociace IEA je možno využít k analýzám zaměřeným na učitele, neboť obsahují obsáhlé učitelské dotazníky s mnoha zajímavými proměnnými.

Z ekonomických analýz je asi nejznámější studie Münicha a Protivínského (2013), kteří se nechali inspirovat zahraničními analýzami dopadu zlepšení výsledků na HDP zmiňovanými v oddíle 4.2.3. S využitím výsledků českých žáků ve výzkumu PISA 2012 simulovali hypotetický dopad reforem (které by vedly ke zlepšení výsledků vzdělávání) na HDP v České republice v horizontu 80 let. Autoři ukazují, že pokud by se podařilo zvýšit výsledky žáků v celé ČR na úroveň Prahy, zvýšil by se HDP o 6 miliard korun, což činí 149 % HDP z roku 2012. Pokud by se podařilo docílit takového růstu, jakého dosáhlo Polsko mezi lety 1999 a 2009, odhadovaný přírůstek HDP by v následujících 80 letech činil dvojnásobek ročních výdajů na důchody a šestinásobek výdajů na školství.

Data z mezinárodních výzkumů jsou v ČR nejčastěji využívána v sociologických studiích k výzkumu vzdělanostních nerovností. Z posledních let je možno uvést například studii Katrňáka, Simonové a Fónadové (2013), kteří na datech z výzkumů PISA 2000 a 2009 testovali teorie maximálně udržitelné nerovnosti a efektivně udržitelné nerovnosti na úrovni středoškolského studia, tedy zabývali se otázkou, do jaké míry vede expanze vzdělávacích příležitostí ke zvýšení šancí mladých lidí s méně příznivým rodinným zázemím. Matějí a Straková (2003), Straková (2010a) a Straková a Greger (2013) využili dat z mezinárodních výzkumů ke studiu úlohy víceletých gymnázií v českém vzdělávacím systému. Zkoumali, do jaké míry ovlivňuje

pravděpodobnost studia na víceletém gymnáziu rodinné zázemí žáka a míra jeho kognitivních schopností, a jak se liší výsledky žáků na víceletém a čtyřletém gymnáziu po zohlednění rodinného zázemí. Matějů, Smith, Soukup a Basl (2007), Matějů, Smith a Basl (2008) a Simonová a Soukup (2010) se zaměřili na studium vzdělanostních aspirací. Zkoumali, jak se liší formování aspirací v závislosti na navštěvovaném studijním programu a jak vzdělanostní aspirace ovlivňují další vzdělávací dráhu žáků. Analýzy zaměřené na studium nerovností byly publikovány rovněž ve dvou monografiích věnovaných této problematice (Matějů; Straková 2006) a (Matějů; Straková; Veselý 2010).

Řada studií, které pracují s daty z mezinárodních výzkumů, se zabývá rozdíly mezi výsledky chlapců a děvčat (např. Potužníková; Straková 2006, Matějů; Simonová 2013). Studie se zaměřují nejen na rozdíly ve výsledcích v jednotlivých sledovaných oblastech, ale i na souvislost mezi výsledky a školním hodnocením.

Objevují se i studie s méně obvyklými tématy: například Hirschová a Kreidl (2012) zkoumali vliv počtu sourozenců na dosažené výsledky ve výzkumu PISA.

V pedagogických analýzách jsou nejčastěji podrobněji analyzovány výsledky žáků ve vybraných úlohách. Data jsou využívána k analýzám silných a slabých stránek jednotlivých hodnocených oblastí. Např. Rendl a Vondrová (2014) se zaměřili na studium nedostatků českých žáků v matematice. Mnozí pedagogové používají zveřejněné úlohy z mezinárodních výzkumů k provádění vlastních návazných výzkumů. Např. Najvarová (2008) použila čtenářské úlohy PIRLS k vlastnímu výzkumu čtenářské gramotnosti, ve kterém využila rovněž kvalitativní metody, Hrbáčková (2009) využila úlohy PIRLS ke studiu autoregulace a Starý a Laufková (2015) použili čtenářské úlohy z výzkumů PIRLS a PISA k hloubkovému zkoumání porozumění polemickému textu u českých žáků. Kaleja a Zezulková (2015) využili úlohy z výzkumu PIRLS ke zkoumání vzdělávacího pokroku žáků v běžných základních a základních praktických školách.

V českých odborných časopisech se setkáváme rovněž s analýzami, které staví na datech z kontextových dotazníků k analýze postojů učitelů nebo analýze postojů žáků. Např. Münich a Federičová (2015) zkoumali postoje žáků ke škole a ke vzdělávání. Na postoje žáků se zaměřily i práce využívající data z výzkumů občanské výchovy (Basl; Straková; Veselý 2009, Straková; Simonová 2013).

Nejčastěji využívanými technikami jsou u českých výzkumníků (kromě deskriptivních statistik) víceúrovňové modelování, logistická regrese a strukturní modelování.

V českých pedagogických časopisech byly prezentovány rovněž deskriptivní statě prezentující dílčí výsledky, kterým nebyl poskytnut dostatečný prostor v publikacích národních výsledků. Například Mandíková (2009) se zaměřila na postoj žáků k přírodním vědám, Straková (2010b) se věnovala postojům a pedagogickým činnostem českých učitelů. V ČR byla publikována rovněž řada studií přehledových, které představují koncepty, metody a výsledky mezinárodních šetření (např. Straková 2009, Potužníková a kol. 2014). Řada textů dále diskutuje pozitiva a negativa mezinárodních výzkumů. Kritickými texty českých autorů se bude blíže zabývat oddíl 5.1.2.

5. Kritika mezinárodních výzkumů a její reflexe

Tato kapitola se věnuje nedostatkům mezinárodních výzkumů a možnostem, jak tyto nedostatky překonat. Nejprve je podrobně popsána kritika sporných aspektů mezinárodních výzkumů, tak jak se objevuje v mezinárodních i národních odborných polemikách. Velká pozornost je věnována zejména diskusi věnované výzkumu PISA a jeho dopadům na vzdělávací politiku v zúčastněných zemích. Ve druhé části kapitoly je nastíněn pravděpodobný vývoj mezinárodních výzkumů a naznačeny cesty k odstranění kritizovaných nedostatků.

5.1 Pohledy na užitečnost a škodlivost dopadů mezinárodních výzkumů

5.1.1 Diskuse v mezinárodní komunitě

Kořeny kritiky mezinárodních výzkumů

Jak bylo uvedeno v kapitole 4.2.2, v diskusích z posledních let jsou mezinárodní výzkumy a výzkumy efektivity vzdělávání pojednávány jako dva odlišné proudy, které se vyvíjejí paralelně a v optimálním případě se vzájemně obohacují a doplňují. Na přelomu století však byla situace odlišná. Mezinárodní výzkumy výsledků vzdělávání byly pojímány jako integrální součást výzkumů efektivity vzdělávání a kritika, která dnes směřuje specificky k těmto výzkumům, směřovala k výzkumům efektivity vzdělávání obecně. Je užitečné se seznámit s těmito výtkami i s kontextem jejich vzniku, neboť mnohé z nich se objevují i v aktuálních kritikách mezinárodních výzkumů.

Na konci 90. let byla hlavním zdrojem kritických výhrad vůči výzkumům efektivity vzdělávání Velká Británie, neboť tam se koncem minulého století těšily výzkumy efektivity vzdělávání velké oblibě mezi výzkumníky i mezi tvůrci vzdělávací politiky. Kritika těchto výzkumů byla posílena i výhradami vůči využívání testů obecně, které reagovaly na množící se zjištění, že v té době hojně využívané plošné testování všech žáků má řadu negativních dopadů na vyučování a učení.

Výhrady vůči výzkumům efektivity vzdělávání lze rozdělit do čtyř oblastí (Goldstein 2000): 1. zneužití výzkumů vládou, 2. neúnosné zjednodušování komplexních vazeb příčin a následků ve vzdělávacím systému a zaměření se na ligové tabulky, 3. absence teorie, 4. špatná kvalita výzkumu, zpracování dat a jejich interpretací.

1. Zneužití výzkumů vládou

Podle názoru kritiků z řad akademické obce si vláda přisvojila výzkumy efektivity vzdělávání, neboť jí umožnily zbavit se odpovědnosti za kvalitu vzdělávání a vinit školy a učitele ze všech společenských problémů. Terčem kritiky se stali i výzkumníci, kterým bylo vytýkáno, že přijali východisko, že školy a učitelé zodpovídají za ekonomický a sociální rozkvět země, a byli obviňováni z toho, že pracují podle pokynů politiků. To podle kritiků spočívalo zejména v tom, že – ve snaze nalézt klíčové faktory, ovlivňující výsledky vzdělávání – neúnosně zjednodušují realitu. Kritici zdůrazňovali, že primární úlohou akademiků je vytvářet teorie, nikoli se zaměřovat na pragmatická řešení. Tento kritický proud tedy více méně zpochybňoval, že by pedagogický

výzkum měl sloužit jako opora pro tvorbu vzdělávací politiky a že by jeho výsledky měly mít nějaké praktické využití.

2. Neúnosné zjednodušování komplexních vazeb příčin a následků ve vzdělávacím systému a zaměření se na ligové tabulky

Zde směřovala hlavní výhrada k tomu, že školy jsou pojímány jako izolované entity, které spolu nikterak neinteragují. Kritici argumentovali, že v Británii tento předpoklad neplatí, protože školy svým konáním významně ovlivňují okolní školy. Zároveň není možné oddělit fungování škol od fungování celé společnosti. Srovnávání škol (i systémů) na základě výsledků v testech je tedy zavádějící, neboť nezohledňuje skutečnost, že školy fungují v nějakém prostředí, z jehož působení se nemohou vlastním přičiněním vymanit.

3. Absence teorie

Výzkumům byla vytýkána absence teorie, která by provázala všechny důležité aspekty fungování škol včetně společenského kontextu. Kritici uváděli, že ve výzkumech efektivity vzdělávání funguje teorie spíše jako zhmotnění empirických vztahů, přičemž řada faktorů (například klima, étos apod.) není dobře vymezena, tím méně měřena. Kritici výzkumníkům vyčítali, že teorii koncipují na základě získaných poznatků co nejjednodušeji tak, aby potvrdila zamýšlené kroky vzdělávací politiky.

4. Špatná kvalita výzkumu, zpracování dat a jejich interpretací

Řada diskusí se točila kolem statistického modelování, kterému bylo opět vytýkáno zejména to, že realitu nepřipustně zjednodušuje. Obhájci užitečnosti výzkumů připouštěli, že je samozřejmě pravda, že modely nejsou nikdy tak dobré jako data, která se snaží modelovat, a že někdy zjednoduší realitu natolik, že vedou k mylným interpretacím. Zároveň však argumentovali, že na druhé straně kvalitní modely mohou pomoci ukázat složitou povahu jevů a odhalit překvapivé vazby. Připomínali, že většina toho, co víme o vlivu rodinného zázemí na výsledky žáků, pochází z víceúrovňového modelování, které poukázalo na důležitost školního kontextu.

Goldstein (2000) doporučoval vytvořit teorii efektivity vzdělávání a dobře formulovat předpoklady, na kterých je založena. Dále požadoval striktně oddělit výzkum od vzdělávací politiky, která směřuje k hledání rychlých řešení a pracuje v krátkodobém horizontu volebních období, a nezaměřovat se na problémy vymezené vzdělávací politikou, ale na zásadní otázky, které ukáže výzkum.

Nejčastěji kritizované aspekty mezinárodních výzkumů

Na kritiku výzkumů efektivity vzdělávání navázaly po zveřejnění výsledků z prvního a druhého cyklu výzkumu PISA výhrady, které směřovaly specificky k tomuto výzkumu. Kontinuálně na ně navazují další kritická vyjádření. Řada z nich se týká všech mezinárodních výzkumů vědomostí a dovedností. Některé směřují k plošnému testování obecně. Mnohé argumenty jsou zmiňovány opakovaně, aniž by kritici zohlednili debatu, která k těmto argumentům již proběhla a v některých případech vygenerovala pádné protiargumenty. Kritika se stále zaměřuje zejména na výzkum PISA, protože má největší mediální ohlas a jeho tvůrci se často pouštějí do poměrně smělých interpretací. Zároveň se však tento výzkum jeví být díky svému designu a z hlediska identifikace důležitých faktorů ovlivňujících výsledky vzdělávacích systémů nejméně užitečným,

a to nejen kvůli absenci longitudinální složky, která je společná všem mezinárodním výzkumům výsledků vzdělávání, ale i kvůli svému designu.

Kritiku je možno rozdělit do několika okruhů: chybná interpretace získaných poznatků daná zejména absencí longitudinální složky, nedostatečné zohlednění specifik jednotlivých vzdělávacích systémů, nepřipustné zjednodušování zjištěných poznatků, metodologické otázky a pochybení, omezený záběr výzkumů a negativní dopady samotné existence výzkumů na vzdělávací politiky zúčastněných zemí. Následující text se podrobněji zabývá jednotlivými okruhy.

Výhrady k absenci longitudinální složky se týkají všech mezinárodních výzkumů, ale nejčastěji se s nimi setkáváme u výzkumu PISA, kde je situace komplikována i tím, že netestujeme žáky v rámci tříd a ročníků. Zde je výhrady možno shrnout následovně: PISA měří vědomosti a dovednosti, které si žáci osvojili za celou vzdělávací a životní dráhu do věku 15 let. Neříká nic o tom, co z toho se žáci naučili ve škole, v níž se aktuálně nacházejí. Kdybychom chtěli získat informaci o dopadu výuky na dané škole na výsledky žáků, museli bychom změřit vědomosti a dovednosti žáků v době, kdy na danou školu vstoupili, a pak je porovnat s vědomostmi a dovednostmi naměřenými výzkumem PISA, tedy získat informaci o „přidané hodnotě školy“ (viz oddíl 4.2.3). Design výzkumu PISA ovšem neumožňuje změřit vstupní vědomosti a dovednosti. Proto také neumožňuje jednoduše zjišťovat dopad kvality výuky, případně dalších faktorů, na výsledky žáků. Situace je dále komplikovaná tím, že v PISA nejsou hodnoceny celé třídy a že patnáctiletí žáci v mnoha vzdělávacích systémech vstoupili do dané školy teprve nedávno (jako čeští žáci, kteří jsou testováni v prvním roce čtyřletého studia ve středních školách). Je tedy velmi obtížné vyvozovat kauzální závěry o tom, že některé opatření má dopad na výsledky žáků v průřezovém výzkumu, jako je PISA.

Tvůrci výzkumu výše uvedené omezení ne vždy respektují. Jako jeden z poznatků výzkumu PISA bylo například prezentováno zjištění, že školy, které mají lepší výsledky, tyto své výsledky zveřejňují. Z uvedeného poznatku nicméně není možno vyvodit, že zveřejňování dat vede k lepším výsledkům, neboť je stejně tak dobře možné, že existuje obrácená podmíněnost – tedy že školy, které dosahují dobrých výsledků, mají větší tendenci tyto výsledky zveřejňovat. Nejasnosti v příčinách a následcích v tomto případě navíc není možno odstranit tím, že zohledníme socioekonomický status žáků ve školách. Ve zprávě z výzkumu PISA byla však data tímto způsobem interpretována a zveřejňování výsledků bylo doporučeno jako strategie ke zlepšování výsledků žáků (OECD 2007). Podobně byl učiněn závěr, že větší počet žáků ve třídách podmiňuje jejich lepší výsledky, když nebylo dostatečně zohledněno, že větší počet žáků ve třídách bývá zpravidla ve výběrových školách a třídách, kam nastupují žáci s lepšími počátečními znalostmi.

Mezinárodním výzkumům je často vytýkána jejich deskriptivní povaha. Na jejich základě je sice možno popsat silné a slabé stránky týkající se výsledků žáků a podmínek v jednotlivých systémech, jen zřídka však dovednou odpovědět na otázku, jaké jsou příčiny těchto silných a slabých stránek. Rovněž ekonomické analýzy jsou často popisné, protože i ty jsou většinou limitovány průřezovým charakterem dat, a důležité charakteristiky (například kulturní) do nich nejsou zahrnuty. Mnozí kritici připouštějí, že analýzy dat z mezinárodních výzkumů výsledků vzdělávání mohou přispět k naší znalosti vzdělávacích systémů, ale zdůrazňují, že musíme vzít v úvahu jejich omezení. Je možno uvést mnoho příkladů, kdy tvůrci vzdělávacích politik činili jednoduché zkratkovité závěry o příčinách rozdílů ve výsledcích jednotlivých zemí. K řadě z nich zaváděly příčinu i mylné nebo špatně interpretované informace v mezinárodních publikacích výsledků, které mnohdy nedostatečně zohledňují specifika systémů zúčastněných zemí.

Goldstein (2004) výzkumu PISA vytýkal, že stejně jako všechny ostatní mezinárodní srov-

návací výzkumy postrádá longitudinální složku. Rozdíly ve výsledcích jednotlivých zemí tak odrážejí nejen rozdíly ve výsledcích, ale také všechny ostatní rozdíly, které je ovlivňují. Z toho například vyplývá, že z průřezových studií není možno vyvozovat závěry, že výsledky mají dopad na zaměstnanost nebo na zdravotní stav obyvatelstva. Podle jeho názoru by se měl výzkum PISA zaměřit větší měrou na zkoumání rozdílů v kultuře, v obsazích vzdělávání, v organizaci školy, a nikoli na porovnávání výsledků pomocí žebříčků. Zdůrazňoval, že tato žádoucí orientace výzkumu by však vyžadovala jinou a promyšlenější konstrukci dotazníků.

Kritiky upozornily na důležitý fakt, že publikované výsledky a jejich interpretaci by měli čtenáři vždy číst s vědomím specifik vlastního vzdělávacího systému. Například mnohé problémy při interpretaci národní situace v mezinárodním srovnání vznikají obtížnou srovnatelností vzdělávacích systémů, a ne vždy přesným převodem dosaženého vzdělání nebo plánovaného vzdělání z národních na mezinárodní kategorie.

Skutečnost, že v některých zemích, včetně České republiky, jsou testováni žáci v různých typech škol a na různých úrovních vzdělání, vyžaduje, aby výzkumníci a tvůrci vzdělávací politiky v těchto zemích přistupovali s maximální obezřetností zejména k interpretacím týkajícím se dopadu školních faktorů na výsledky vzdělávání. Například mezinárodní zpráva z výzkumu PISA 2012 uvádí, že v ČR má na výsledky škol blahodárný dopad konkurence jiných škol, a to v tom smyslu, že školy, které musí obstát v konkurenci s jinými školami, se více snaží, a tudíž jejich žáci dosahují lepších výsledků (OECD 2013d, s. 54). Podle analýzy OECD dosahovaly školy v ČR, v jejichž spádové oblasti byla alespoň jedna škola, která se ucházela o jejich žáky, lepších výsledků než školy, kde žádná taková škola nebyla, i po zohlednění složení žáků školy. Tento poznatek byl prezentován jako výjimečný, neboť přítomnost konkurence se jevila jako statisticky významná pouze v ČR a v Estonsku. Analýza však nevzala v potaz skutečnost, že v ČR jsou do výzkumu zařazeni žáci různých vzdělávacích programů a různých úrovní vzdělání. Efekt konkurence zmizel, jakmile byl do analýzy zařazen typ školy. Žáci gymnázií a žáci středních odborných škol dosahují v průměru lepších výsledků než žáci základních škol a mezi středními školami panuje větší konkurence (žáci si střední školy pečlivě vybírají). Mezi základními školami nebyl efekt konkurence statisticky významný. Nejedná se tedy o efekt konkurence, ale o efekt střední školy a jejího typu. Nekritické přijetí závěrů v mezinárodní publikaci by mohlo vést k mylným závěrům o fungování českého vzdělávacího systému (více v Dvořák; Straková 2016).

Mezinárodní výzkumy, respektive interpretace jejich výsledků, jsou často kritizovány za to, že nezohledňují skutečnost, že rozmanité charakteristiky systémů, jako je kultura a historie, znemožňují vyvozovat závěry o příčinných vztazích mezi rozličnými aspekty vzdělávacích systémů. Tento problém se vztahuje k opominutým proměnným ve srovnání jednotlivých systémů. Většina těchto problémů se nicméně odstraní při analýze trendových dat, neboť v systémech je řada proměnných neměnných a můžeme je tedy z analýz vynechat a zaměřit se pouze na proměnné, jejichž změny nás zajímají. Tyto analýzy pak umožňují testovat hypotézy o dopadech vzdělávacích politik, jak bylo ukázáno v oddíle 4.2.3.

Již bylo uvedeno, že při interpretaci výsledků mezinárodních výzkumů jsme často svědky zjednodušování. Je to způsobeno snahou o srozumitelnost a atraktivitu pro tvůrce vzdělávacích politik, kteří oceňují jasné a jednoduché návody. Situace je však v řadě případů komplexní a výstupy z mezinárodních výzkumů jsou za tato zjednodušování velmi kritizovány. Například koncept *opportunity to learn* je často interpretován tak, že žáci se naučí to, co jim je učitelem předestřeno. To je podle řady odborníků zapojených v pedagogickém výzkumu ovšem nepřipust-

né zjednodušení, neboť záleží na způsobu práce učitele s daným učivem. Důležitá je struktura vyučovací hodiny, její řízení, srozumitelnost, informativní a podpůrná zpětná vazba. Zároveň motivace a další nekognitivní výstupy jsou utvářeny dalšími faktory, jako je třídní klima a vztahy mezi žáky a učitelem, které podporují vstřícnost, kompetentnost a autonomii. Klieme a kol. (2009) charakterizovali kvalitu výuky pomocí tří faktorů: a) jasné a dobře strukturované řízení třídy, b) podpůrné klima třídy, c) kognitivní aktivace a náročný obsah. Autoři však zdůrazňují, že ani dobrá konceptualizace kvality výuky nemusí vyřešit všechny problémy, neboť různé populace žáků mohou profitovat z různého způsobu výuky. Při analýzách pak musíme tedy zohlednit i interakci mezi způsobem výuky a charakteristikou žáků. Dotazníky v posledních kolech mezinárodních výzkumů obsahují informace, které umožňují zkoumat aspekty práce školy a učitele, které se ukazují jako důležité vzhledem k výsledkům vzdělávání (viz oddíl 3.1.2). Hledání příčinných souvislostí je však stále ztěžováno průřezovou povahou výzkumů.

Mezinárodním výzkumům je dále často vytýkáno, že obsahují jen malý výsek z důležitých vzdělávacích cílů a oblastí vzdělávání. Nejčastěji je kritizována skutečnost, že se realizované studie dosud nezaměřují na cizí jazyky. Taková studie je připravována asociací IEA, což bylo uvedeno v kapitole 1.2.4. Zároveň byla v roce 2011 v 16 evropských zemích realizována Evropská studie jazykových kompetencí, která zjišťovala čtení, poslech a psaní v několika jazycích (*European Survey on Language Competencies*) (European Commission 2012).

Další často diskutovanou slabinou ve vztahu k obsahu výzkumu PISA je absence občanských kompetencí, kterou jsme se zabývali v oddíle 4.1.3. Např. Nussbaum (2011) píše o tom, že jsou ve vzdělávání zanedbávány kompetence, které udržují naživu demokracii. Vzdělávací systémy čím dál tím více směřují k produkci užitečných strojů namísto plně rozvinutých osobností, které jsou schopny autonomního a kritického myšlení. Nedostatečné občanské kompetence mladých lidí jsou ignorovány, zatímco význam kognitivních kompetencí je přeceňován. Mezinárodní výzkumy, které se zaměřují na kognitivní kompetence, k tomuto trendu podle kritiků významně přispívají.

Za pravděpodobně neškodlivější aspekt mezinárodních výzkumů je považováno zacílení na průměrné výsledky a jejich porovnávání prostřednictvím žebříčků. Jeden okruh výhrad směřuje k tomu, že tyto žebříčky nevypovídají věrně o vědomostech a dovednostech žáků v jednotlivých zemích, neboť dostatečně nezohledňují národní specifika (typicky prioritní cíle vzdělávání a kurikulum). Mezinárodní výzkumy jsou kritizovány za to, že „sjednocují“ kurikulum a obracejí pozornost škol a učitelů výhradně k omezenému okruhu vzdělávacích cílů, které jsou měřeny mezinárodními výzkumy. Vzhledem k tomu, že výsledkům výzkumů je přikládána velká váha, ostatní (často důležité) cíle vzdělávání ustupují do pozadí. Kritika se i zde nejčastěji obrací proti výzkumu PISA, kde je rozporováno „právo“ OECD, jejíž primární orientace je ekonomická – prostřednictvím výzkumu PISA určovat priority vzdělávání na celém světě. Vyskytují se rovněž výhrady zaměřené na prezentaci výsledků pomocí průměrů, přičemž kritici uvádějí, že by větší pozornost měla být věnována jejich rozložení v populaci. Je otázka, do jaké míry přispěly mezinárodní výzkumy k tomu, že v souladu s teorií nového institucionalismu (např. Dvořák 2012) se řada pravidel i nepsaných přesvědčení o roli a žádoucí podobě vzdělávání stala celosvětově sdílenou kognitivní normou.

Mnoho kritiků považuje mezinárodní výzkumy za nástroj politik, které jsou souhrnně označovány jako *new public management* (NPM). NPM spočívá v reformách veřejných služeb s cílem snížit veřejné výdaje, zvýšit kvalitu veřejných služeb, zvýšit efektivitu v poskytování těchto služeb a také zvýšit legitimitu a zodpovědnost institucí ve veřejném sektoru. Pod vlivem NPM

došlo k deregulaci, privatizaci a decentralizaci veřejného sektoru a k přejímání mnoha principů řízení známých ze soukromého sektoru (Veselý 2012). V důsledku reformy fungují služby na bázi soutěže, kdy školy soupeří o studenty, nemocnice o pacienty, vědecký ústav o granty a podobně.

Pod vlivem NPM byl systém školství výrazně decentralizován a školy získaly značnou autonomii v otázce výuky. Spolu s tím získaly na váze i myšlenky budování nové profesní identity učitelů a ředitelů. Stejně tak ale začalo nabývat na důležitosti hodnocení výsledků vzdělávání a zodpovědnost škol za tyto výsledky. V této souvislosti hovoříme o akontabilitě, která spočívá především v tom, že se školy zodpovídají za výsledky žáků ve standardizovaných testech.

Veselý (2012) upozorňuje, že problémem spojeným se zaváděním nových forem akontability je přílišný důraz na hodnocení výstupů, a to na úkor hodnocení vstupů a procesů. Máme-li hodnotit výstupy, je nutné předem stanovit konkrétní cíle, kterých má škola dosáhnout a na jejichž základě bude posuzována. I tam, kde lze cíle formulovat poměrně přesně, často vznikají problémy s hodnotícími kritérii – velmi často nelze kvalitu výkonu (a tedy akontabilitu) kvantitativně změřit. I v případě, že cíle a hodnotící kritéria jsou formulována přesně, dochází často k negativním nezamýšleným důsledkům. Pokud jsou totiž instituce a jednotlivci příliš často a detailně měřeni, stávají se sice lepšími z hlediska naplňování těchto požadavků, ale není tomu tak nutně v naplňování jejich veřejné služby. Jednou ze základních otázek všech reform veřejných služeb je nastavení funkčního systému „skládání účtů“, ve kterém si žádný aktér nemůže dovolit jednat nezodpovědně, ale zároveň mu jeho „skládání účtů“ nebrání v efektivním vykonávání jeho práce. Často se totiž stává, že se aktéři bojí poskytovat pravdivé informace o svých výsledcích, protože ty mohou být použity proti nim, nebo se zaměřují pouze na to, aby dostali definovaným požadavkům, a přestávají být invenční a iniciativní. Veselý (2012) v této souvislosti hovoří o „institucionalizaci konformity“. Mezinárodní výzkumy jsou obviňovány z toho, že posilují výše popsané negativní důsledky akontability a ve svém důsledku přispívají k oslabování efektivity fungování vzdělávacích systémů.

V souvislosti s aktivitami OECD bývá ještě často kritizováno, že OECD nabízí své produkty v oblasti vzdělávání nejen státům, ale i jednotlivcům, přičemž šíří svoji „ideologii“ a nehledí na možné negativní dopady. Takto je kritizován produkt *PISA for Schools*, jehož prostřednictvím mohou školy na celém světě otestovat své žáky a porovnat jejich výkony s mezinárodními výsledky, kritizována je i aplikace *PIAAC online*, která slouží k testování dospělých.

Goldstein (2004) zmiňuje i řadu výhrad metodologických, které se týkají překladu výzkumných nástrojů, výběru respondentů a zpracování získaných dat. Jako ústřední nedostatek zpracování dat uvádí předpoklad jednodimenzionality škál, na kterých jsou prezentovány výsledky výzkumu. Argumentuje, že škály nejsou přirozeně jednodimenzionální, ale vícedimenzionální je uměle potlačována výběrem úloh i zpracováním výsledků. Jako problematickou hodnotí zejména skutečnost, že dílčí škály spolu vysoce korelují a že jsou úlohy ke škálám přiřazeny arbitrárně. Domnívá se, že by bylo vhodnější a hlavně poctivější ke konstrukci škál využít faktorovou analýzu, která ukazuje dvě dimenze – kromě obsahu úloh ještě jejich formát. Tento aspekt je při zpracování výsledků z PISA záměrně opominut.

Dalším problémem, na který Goldstein (2004) upozorňoval, byly vyšší hodnoty koeficientu vnitrotřídní korelace, který vypovídá o velikosti rozdílů mezi školami, ve výzkumu PISA než ve výzkumu TIMSS. Tyto rozdíly je ovšem možno vysvětlit zapojením žáků z různých ročníků a škol z různých stupňů vzdělávání, ve kterých jsou vzdělávání patnáctiletí žáci. Výzkum TIMSS zkoumá žáky z jednoho ročníku, takže je nehomogenita charakterizována pouze jakousi horizontální složkou, k níž již nepřistupuje vertikální složka daná odlišnými ročníky a stupni škol.

Část metodologické kritiky mezinárodních výzkumů je pravděpodobně způsobena i tím, že výzkumy jsou metodologicky náročné a tudíž pro řadu aktérů nesrozumitelné. Také často přinášejí nepopulární výsledky, které de facto zpochybňují činnost kritiků, kteří se proti nim vymezují. Řada kritiků nicméně vznáší závažné metodologické výhrady, kterým je třeba věnovat pozornost. Žádná metodologická výhrada se však dosud neprokázala jako natolik pádná, aby výsledky výzkumů zásadním způsobem zpochybnila.

Jako příklad můžeme uvést kritiku použití písemných testů s tím, že jejich prostřednictvím nemůžeme získat validní informaci, neboť výsledky testů neodpovídají výsledkům ústního zkoušení těch samých vědomostí (Gustafsson; Rosén 2014). Experimenty prokázaly, že písemná úloha je pro žáky obecně skutečně obtížnější. Praktické úlohy použité ve výzkumu TIMSS 1995 však prokázaly velkou konzistenci s výsledky písemných testů. Pro mezinárodní výzkumy plyne z této kritiky povinnost zjistit, zda přechod od ústního k písemnému prokazování vědomostí neznevýhodňuje některé žáky více než jiné.

Další metodologická výhrada, která byla široce diskutována v odborné komunitě, směřovala k tomu, že v PISA používaný Raschův model neodpovídá dobře datům (Gustafsson; Rosén 2014). Tato výtky byla opakovaně empiricky potvrzena. Přestože to výsledky pravděpodobně významně neovlivňuje, neboť jsou v čase stabilní, určitě tento poznatek povede ke zdokonalení používaného postupu.

Polemika akademiků s představiteli OECD

Výše uvedené výhrady se objevily v polemice mezi představiteli akademické obce a direktorátu OECD pro vzdělávání, která se uskutečnila v roce 2014. V květnu 2014 byl v britském *Guardianu* uveřejněn otevřený dopis skupiny akademiků adresovaný Andreasu Schleicherovi, který je v OECD zodpovědný za realizaci výzkumu PISA⁵⁹. V dopise mu byla vytykána podpora soutěže mezi vzdělávacími systémy a podpora orientace tvůrců vzdělávacích politik na krátkodobé a velmi úzce zaměřené cíle. OECD bylo dále vytykáno, že jako ekonomická instituce je příliš zaměřena na přípravu mladých lidí pro pracovní trh, přičemž opomíjí obecně kultivační úlohu vzdělávání a přípravu mladých lidí pro aktivní zapojení do osobního života a demokratické společnosti. Na rozdíl od jiných institucí (UNESCO, UNICEF) nemá OECD péči o vzdělání ve své misi. Další výtky směřovaly k tomu, že do výzkumu PISA jsou angažovány komerční firmy, které zejména ve Spojených státech vydělávají na přípravě žáků na testování, jež má neblahý dopad na učení žáků. Podle autorů dopisu poškozuje PISA mladou generaci, protože zvyšuje ve školách na celém světě důraz na testování, zvyšuje míru stresu ve školách a zároveň ubírá učitelům autonomii a chuť do práce. Autoři zároveň zdůrazňují, že na výsledcích výzkumu PISA by neměly být založeny žádné reformy vzdělávacích systémů, neboť se jedná o izolovaná měření, která nezohledňují národní kontext. Dopis požaduje, aby OECD:

1. publikovala výsledky jiným způsobem než prostřednictvím žebříčků,
2. zapojila do konceptualizace výzkumů nejen statistiky a psychometriky, ale také pedagogy, sociology, antropology, historiky a další odborníky z širokého okruhu vědních oborů,
3. zapojila do konceptualizace výzkumu organizace, které se nezabývají formulací standardů a hodnocením, ale zdravím, harmonickým vývojem a spokojeností mladých lidí,
4. zveřejnila náklady na výzkum,

⁵⁹ <http://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics>

5. umožnila nezávislé monitorování všech fází výzkumu,
6. podrobně referovala o zapojení komerčních organizací ve výzkumu,
7. snížila frekvenci testování.

OECD výtky akademiků striktně odmítla. Odpověděla, že PISA svojí pravidelností naopak podporuje dlouhodobé plánování, ale že výsledky některých zemí (např. Polsko a Německo) ukazují, že zlepšení je možno docílit ve velmi krátkém čase. Odmítla nařčení ze zaměření na úzký výsek sledovaných cílů s tím, že PISA měří kromě kognitivních výstupů také postoje a motivace a mnohé další dimenze, a že záběr výzkumu je kontinuálně rozšiřován. Mandát k aktivitám na poli vzdělávání je dle odpovědi poskytován OECD členskými zeměmi, které prostřednictvím řídicího výboru (*Education Policy Committee*) rozhodují o tom, čemu se bude OECD na poli vzdělávání věnovat. Soukromé firmy jsou do aktivit zapojovány pouze na základě otevřených výběrových řízení na vykonání dílčích úkolů, práce jsou řízeny OECD, která za ně nese finální zodpovědnost. Představitelé OECD rovněž odmítli výtku, že PISA zvyšuje nepohodu ve školách, a argumentovali tím, že probíhá jednou za 3 roky na malém výběru škol a žáků a že vlastní testování trvá 2 hodiny. OECD v odpovědi také odmítá nařčení z toho, že zaměřením na žebříčky průměrných výsledků podporuje soutěž a rostoucí nerovnosti. Žebříčky představují jen velmi malou část prezentovaných výsledků. Již od zahájení výzkumu jsou součástí prezentací výsledků podrobné ukazatele ilustrující nerovnosti v zúčastněných zemích. Na výtku, že zařazování rozvojových zemí do žebříčků nedává smysl a zavání kolonialismem, OECD odpovídá, že to je rozhodnutí rozvojových zemí být prezentovány tímto způsobem a že PISA naopak ukazuje, že svět nadále není rozdělen na bohaté a dobře vzdělané a chudé a špatně vzdělané země (OECD 2014d).

Na kritiku obsaženou v otevřeném dopise navázali další akademici, kteří jdou často ještě dále a zpochybňují samotnou měřitelnost a srovnatelnost výsledků vzdělávání (např. Van Avermaet; Sierens, 2010, Davis 2013, Davis a kol. 2015). Davis (2014) vyslovuje domněnku, že některé aspekty hodnocení vzdělávacích výsledků jsou podobné hodnocení uměleckých výkonů, u kterého není požadována konsistentnost hodnotitelů. Pokud upřednostníme konsistentnost, můžeme některé důležité vzdělávací výsledky pominout. Situaci komplikuje podle autora rovněž skutečnost, že jednotlivé výsledky vzdělávání jsou ze své podstaty nesouměřitelné a neporovnatelné. Autor argumentuje také pro holistické hodnocení v neprospěch hodnocení analytického, které je používáno právě pro zvýšení objektivity. Uvádí, že často nelze hodnotit jedno kritérium, aniž bychom pochopili a zohlednili jiné a vzájemnou interakci obou aspektů. Hodnota nějakého prvku žákovské práce může být odlišná v odlišných kontextech a hodnota celku není nutně součtem hodnot dílčích prvků (Davis 2013).

Všechny uvedené kritiky je nutno brát vážně a jsou považovány za cenné podněty pro další vylepšování metodologie i zacílení výzkumů. Jak již bylo uvedeno, kritika mezinárodních výzkumů se soustřeďuje na výzkum PISA, kvůli jeho značnému dopadu na vzdělávací politiku. Část výhrad se však vztahuje i na výzkumy IEA a jistě platí i pro národní testovací programy v řadě vyspělých zemí, což nebývá dostatečně zdůrazňováno.

Debata mezi odborníky, kteří realizují mezinárodní výzkumy a pracují ze získanými daty, a odpůrci mezinárodních komparativních výzkumů pravidelně probíhá na Evropské konferenci pedagogického výzkumu (ECER) na společném sympoziu skupiny *Network 13* zaměřené na filosofii vzdělávání a skupiny *Network 9* zaměřené na hodnocení výsledků vzdělávání. Zde jsou diskutovány nejen povaha výstupů měřených v mezinárodních výzkumech a škodlivé dopady těchto výzkumů, ale také obecné otázky srovnatelnosti a měřitelnosti výsledků vzdělávání.

5.1.2 Kritika mezinárodních výzkumů v českých pedagogických časopisech

Kritika výzkumu PISA našla velkou odezvu i mezi českými a slovenskými výzkumníky. Kašćák a Pupala (2011) publikovali v *Orbis scholae* kritický text, který přehledně shrnul všechny hlavní výhrady publikované v první dekádě 21. století v zahraničních časopisech. Autoři PISA vytýkají zejména to, že dělá vzdělávací politiku; a to výzkumu nepřísluší. Podle jejich názoru deformuje PISA pedagogický výzkum, respektive srovnávací pedagogiku, neboť opomíjí všechny důležité a do té doby zkoumané aspekty, jako je struktura vzdělávacích systémů, kurikula, výukové metody apod., a zaměřuje se pouze na aspekt výstupu. PISA se stává nástrojem *new public managementu*. Do centra pozornosti se dostává výkon a efektivita. Hodnocení se stává součástí globální neoliberální politiky a kontroly. PISA akceleruje orientaci vzdělávacích systémů na výstupy, přičemž tato orientace je považovaná za součást logiky nového institucionálního prostředí, které v globálním tržním prostoru má být schopné udržet kompetitivnost a přitahovat kapitál. Vědomosti se zde mění na komoditu a PISA je nástrojem, který redukuje národní poznatkové továrny tak, aby obstály v mezinárodní ekonomické soutěži a globální produkci. Reformní aktivity, které PISA iniciuje v jednotlivých systémech, se zaměřují na nedůležité aspekty vzdělávacího procesu, přičemž podstatné aspekty jsou zanedbávány. S poukazem na zahraniční zdroje kritizují autoři samotný koncept *evidence based policy* (politiky založené na důkazech), ke kterému se PISA hlásí.

Autoři dále poukazují na to, že organizátoři mezinárodních měření nechávají interpretaci výsledků na jednotlivých zemích, kde se jí zmocňují nepoučení politici, neboť akademici se zdráhají provádět jednoduché závěry. Sdělení se tak redukuje na pořadí v žebříčku.

Neopominutelným aspektem je rovněž okolnost, že PISA představuje velký byznys – nejen na mezinárodní úrovni, kde se na realizaci výzkumu podílí řada soukromých subjektů, ale i na národní úrovni, kde vytváří pracovní místa. Jedinci a skupiny, kteří jsou na výzkumu PISA zainteresováni, pak přirozeně brání kritice, která by uvedla výsledky na pravou míru.

Autoři sumarizují i řadu kritik metodologických, které spočívají v poukazování na nízké rozdíly mezi jednotlivými zeměmi, nezohlednění podílu přistěhovalců či vylučování žáků se specifickými vzdělávacími potřebami v některých zemích. Naznačují dokonce, že některé země v testování záměrně podvádějí. Kritizován je rovněž koncept gramotnosti, který je podle kritiků v rozporu se středoevropskou tradicí a zvýhodňuje anglosaské země, jimž je vlastní. Testové úlohy jsou kulturně kódované angloamerickou tradicí založenou na integraci obsahů vzdělání a na mezipředmětových vztazích bez diferenciací mezi poznávacími režimy jednotlivých disciplín. To se projevuje zřetelně v přírodovědných předmětech, kde je přírodovědná oblast testována jako celek bez rozlišování mezi chemií, fyzikou a biologií, jak je běžné v kontinentální Evropě. Podobné aspekty se projevují i v matematice, kde evropští žáci lépe zvládají matematické operace, ale mají nedostatky v logickém úsudku. Rovněž čtení je pojednáno v rozporu s evropskou tradicí, neboť jsou zařazovány nesouvislé texty, které v Evropě součástí čtenářského výcviku nejsou. Kritizován je rovněž koncept aplikace znalostí v situacích běžného života, kdy je konstatováno, že se nejedná o situace autentické, ale jejich autentičnost je pouze předstírána, a to způsobem, který je často zavádějící z hlediska nalezení správného řešení. Usazování úloh do kontextu běžného života navíc může některé skupiny žáků znevýhodňovat, neboť kontext běžného života se pro různé sociální skupiny žáků liší. Snaha o autentičnost úloh vede u matematických a přírodovědných úloh navíc k velké čtenářské náročnosti.

Autoři dále zpochybňují úspěšnost organizátorů při převádění testů do různých jazyků tak, aby byly jednotlivé jazykové verze skutečně srovnatelné a kulturně zcela neutrální. V závěru

uvádějí argumenty zahraničních výzkumníků pro to, že úlohy se podobají spíše inteligenčním testům než úlohám ověřujícím výsledky vzdělávání a ukazují, že pro úspěch v testu je nejvíce potřebnou zkušeností zkušenost s testováním.

Štech (2011) vychází z textu Kaščáka a Pupalý (2011) a přidává další kritické pohledy. Uvádí, že výzkum PISA je málo užitečný pro oborové didaktiky a pro kultivaci vzdělávání vůbec, neboť pracuje se školami a se žáky jako s nezávislými jednotkami (nezohledňuje společný kontext), čímž vychází z mylného předpokladu. Dále zařazuje PISA do oblasti výzkumů efektivnosti vzdělávání, přičemž tento výzkumný proud zpochybňuje s tím, že se zaměřuje na technologické otázky školní edukace a upozaďuje otázky hodnot a vzdělávacích cílů. Uvádí stanoviska zahraničních badatelů, kteří vyčítají pedagogům, psychologům a sociologům, že dopustili rozvoj výzkumů efektivnosti vzdělávání, který vede k tomu, že výzkumníci dělají pouze tzv. užitečný výzkum. Z kritických „dáblových advokátů“ zaujímajících odstup od zájmů školské decize se mnozí stali experty, kteří slouží tvůrcům vzdělávacích politik. Autor spolu s Kaščákem a Pupalou zdůrazňuje, že vědomosti a dovednosti propagované výzkumem PISA a doprovodná rétorika odpovídají ekonomickému pojetí vzdělávání. Výsledky výzkumu jsou často používány k ospravedlnění opatření vzdělávací politiky. Výsledky PISA mají však spíše heuristický charakter a mohou být v lepším případě použity jako inspirace k dalším rigorózně koncipovaným výzkumům, než aby plnily funkci bezprostředního návodu k didaktickým, natož politickým opatřením.

Ve svém textu z roku 2015 (Štech 2015) se autor ztotožňuje s kritikou výše uvedeného otevřeného dopisu akademiků a odkazuje se rovněž na knihu *Who's Afraid of the Big Bad Dragon: Why China has the Best (and Worst) Education* amerického pedagoga Yong Zhao (Zhao 2014), který uvádí, že PISA se stala jedním z nejdestruktivnějších faktorů v dnešním vzdělávání: vytváří prý iluzorní modely excelence, idealizuje systémy, které mají vážné vzdělávací nedostatky, glorifikuje autoritářské přístupy a pozornost vzdělávacího světa směřuje do minulosti místo do budoucnosti. Štech cituje z textů, které upozorňují na to, že adorace finského systému, ke které došlo v celém vyspělém světě právě v důsledku výzkumu PISA, je nepodložená.

Ve svém textu Štech poukazuje i na to, že kritiky PISA jsou velmi různé povahy. Zatímco někteří kritici se zaměřují na kritiku nácviku a drilování vybraných testovaných operací a na s tím související redukci tvořivého vyučování/ učení objevováním, snížením orientace na řešení problémů a podporou transferu v učení (např. právě signatáři výše zmiňovaného dopisu adresovanému Andreasovi Schleicherovi), druzí zase kritizují testové položky PISA jako psychologicky nepodloženou, ba kontraproduktivní aplikaci převážně problémově formulovaných úloh se silnou kulturně nejednoznačnou kontextovou složkou. Tato kritika směřuje ke školsky neadekvátní orientaci na gramotnostní kompetence.

I v českých a slovenských polemikách se objevuje tvrzení, že kvalitu vzdělávání nelze kvantifikovat. V souvislosti s žebříčky je upozorňováno na škodlivost soutěže a testů s výběrem odpovědi, kdy příprava na testy poškozuje vzdělávací proces. Testy jsou zde spojovány výlučně s výzkumem PISA. Není zohledňována skutečnost, že existovaly již dříve a že jsou široce používány v národních kontextech, kde je jejich dopad často daleko více devastující.

Janík (2015) seznamuje české čtenáře s publikací Liesmanna (2014), ve které je jedna kapitola rovněž věnována kritice výzkumu PISA. Podle Liesmanna je PISA symptomem absurdity, která pohltila vzdělávací systém. PISA podle něj prvořadě mění víru vzdělávacích politiků a žurnalistů v pochybné statistiky. Již dávno se výzkum PISA stal sekulárním náboženstvím, které rozlišuje jen pravověrce a kacíře. Není však pochyb o tom, že PISA má velká vliv na formování vzdělávací politiky.

Janík poukazuje na Liesmannovo přesvědčení, že PISA dobře ukazuje, že ve společnosti vědění je možno mít úspěch i bez vědění, a kritizuje skutečnost, že výzkum konstruuje skryté kurikulum. Kritizuje rovněž zaměření výzkumu na studium nerovností a doporučení odložit diferenciaci systému do pozdějšího věku, přičemž poukazuje na evropskou tradici diferencovaného vzdělávání. Obviňuje výzkum PISA z toho, že konstruuje vzdělávací katastrofy, které následně slouží k prosazování reforem. Podle autora vzdělávací systémy nepotřebují více, ale naopak méně reforem.

5.2 Výhled do budoucna – možné cesty k překonání nedostatků

5.2.1 Úvahy o budoucím směřování mezinárodních výzkumů výsledků vzdělávání

Cyklus výzkumů v základních oblastech vzdělávání Mezinárodní asociace pro hodnocení výsledků vzdělávání vzbudil na celém světě velký zájem. Počet zemí, které se výzkumů účastní, se průběžně zvyšuje. IEA má v současné době 69 členských států. Jak již bylo uvedeno, postkomunistické země se do její práce většinou zapojily na počátku 90. let 20. století. V posledním desetiletí se do pravidelně probíhajících výzkumů zapojily rovněž země bývalého Sovětského svazu a země arabské, africké a asijské. Rozvojové země, pro jejichž žáky je v daném ročníku test příliš obtížný, často testují starší ročník – tedy test, který je určen pro žáky 4. ročníku (žáky desetileté), zadávají žákům v pátém nebo šestém ročníku (žákům jedenáctiletým nebo dvanáctiletým).

Vyspělé země využívají v posledních letech v rámci výzkumů IEA zejména možnosti hodnocení vědomostí a dovedností v hlavních oblastech vzdělávání ve 4. ročnících povinného vzdělávání. Pro primární školu nejsou k dispozici žádná jiná měření, přičemž pro konec povinného vzdělávání je k dispozici výzkum PISA. Tvůrce vzdělávacích politik motivuje potřeba zjistit, do jaké míry predikují výsledky na konci primární školy výsledky na výstupu z povinného vzdělávání, případně chtějí podchytit nedostatky vzdělávacího systému již v raném stadiu povinného vzdělávání.

Vzhledem k tomu, že aktivity OECD představují pro asociaci IEA čím dál tím větší konkurenci, neboť prostředky na výzkum jsou limitované a výzkumy OECD jsou mezi tvůrci vzdělávacích politik velmi populární, je v komunitě zástupců zemí sdružených v IEA věnována velká pozornost přemýšlení o možnostech dalšího rozvoje organizace. Kromě výzkumů TIMSS a PIRLS, které IEA zamýšlí zachovat v uvedené periodicitě, je zvažováno rozšíření do dalších oblastí. Mezi těmito oblastmi je nejčastěji zmiňováno předškolní vzdělávání. V roce 2013 zahájila IEA *Early Childhood Education Study* (ECES), jejímž cílem je prozkoumat, popsat a analyzovat předškolní vzdělávání a zjistit, jak přispívá ke školní připravenosti dětí vstupujících do povinného vzdělávání. Výzkum byl zahájen administrací dotazníku pro tvůrce vzdělávacích politik. Následně měl být vyvinut nástroj pro hodnocení kompetencí žáků před vstupem do povinného vzdělávání. Tento nástroj měl být pilotován ve školním roce 2015/2016 a plošně administrován na podzim 2016. Studie reaguje na rostoucí zájem o kvalitu předškolního vzdělávání v souvislosti s množícími se výzkumnými poznatky o jeho významu pro vzdělávací dráhu dítěte a snižování vzdělanostních nerovností. Realizace výzkumu v plné šíři však naráží na vysokou finanční náročnost a s ní spojenou nízkou ochotu zemí se do studie zapojit, respektive na neschopnost výzkumníků získat potřebné finanční prostředky na její realizaci.

Další směr, kterým se ubírá přemýšlení o dalším vývoji IEA, je snaha zvýšit užitečnost výzkumných aktivit pro vzdělávací politiku a hledání cest ke zkvalitnění vyučování a učení. S tím se pojí záměr navázat na plošná měření kvalitativními výzkumy, které by umožnily lépe porozumět kvantitativním výsledkům na mikroúrovni školy. Dále je uvažováno o realizaci regionálních studií, které by umožnily získat větší vhled do fungování vzdělávacích systémů, protože by umožnily do hloubky studovat i společenské, historické a kulturní kontexty.

Výzkum PISA je sice řízen autonomně vlastní řídicí radou, ale tvoří integrální součást politiky OECD na poli vzdělávání. Ta je momentálně určována strategií OECD, která je zaměřena na rozvoj a aktivaci dovedností a jejich efektivní využívání (*OECD Skills Strategy*). Tato strategie reaguje na skutečnost, že v posledních desetiletích došlo ve světě k velkým změnám, které ovlivnily téměř každý aspekt našeho života: změnil se nejen způsob práce a pracovní činnosti, ale i způsob soukromé i úřední komunikace, nakupování, trávení volného času. Za tyto změny je primárně zodpovědný obrovský rozmach informačních a komunikačních technologií, které rychle pronikly do pracovního i soukromého života.

Řadu profesních dovedností, které byly dříve ceněny, nyní nahrazují informační technologie. Statistiky ukazují pokles pracovních míst, která byla zaměřena na vykonávání rutinních manuálních a informačních služeb. Zároveň ukazují nárůst pracovních míst, která vyžadují komplexní komunikaci. Ta spočívá v získávání a interpretaci důležitých informací a jejich zprostředkovávání druhým a v expertním myšlení, jež spočívá v hlubokém porozumění specifickému oboru, zvědavosti a iniciativním přijímání nových obtížných úkolů. Statistiky dále ukazují, že roste počet pracovních míst s vyššími nároky na vzdělání a flexibilitu. Pracující se stále častěji musejí vyrovnávat s radikálními změnami pracovní náplně danými restrukturalizací pracovišť, s nečekanými změnami pracovního trhu (které jsou provázeny nutností často měnit zaměstnání) a s potřebou držet krok s novými technologiemi a materiály, tedy s potřebou stále se učit novým (a stále složitějším) dovednostem. Ani to však není dostatečné, neboť se zřetelně ukazuje, že pro úspěch na trhu práce již nestačí vědomosti a řemeslo, ale stále více se uplatňuje tvořivost, fantazie a sociální inteligence. Z hlediska rozvoje ekonomik je důležité dbát nejen o rozvoj dovedností, ale též o efektivní využívání těch dovedností, které jsou k dispozici. K tomu často nedochází nejen z důvodů nezaměstnanosti, ale též z důvodů špatného přiřazení pracovníků a pracovních úkolů. Strategie má z toho důvodu tři hlavní součásti: pečovat o rozvoj dovedností v průběhu celého života (*develop skills*), podporovat pracující v tom, aby své dovednosti aktivně nabízeli na trhu práce a aby trh práce předčasně neopouštěli (*supply skills*), a dbát o plné využití dovedností na pracovním trhu (*use skills*)⁶⁰.

Z tohoto důvodu prochází výzkum PISA kontinuálním vývojem a modernizací. Sledované oblasti jsou modifikovány v souladu s aktuálními poznatky, modernímu vývoji je přizpůsobována i administrace testu, čehož je dokladem přechod na elektronickou administraci, i masivní využívání možností počítače při interaktivitě úloh.

Stejně jako výzkumy IEA i výzkum PISA zaznamenal od doby svého vzniku vysoký nárůst zúčastněných zemí. Podle původního záměru byl cyklus šetření PISA určen pouze pro členské země OECD. Mezi nečlenskými zeměmi byl však o výzkum takový zájem, že došlo k jeho zpřístupnění všem vzdělávacím systémům, které jsou schopny ho realizovat v požadované technické kvalitě. Počet zúčastněných zemí rychle roste. Zatímco prvního sběru dat v roce 2000 se zúčastnilo 31 zemí, v současné době je ve výzkumu zapojeno již 79 zemí. Tyto země sice

⁶⁰ Více viz <http://skills.oecd.org/>.

nemají rovná práva⁶¹, ale jsou jim přizpůsobovány obtížnosti úloh. Tvůrci výzkumu se zabývají otázkou, jak zajistit, aby mezinárodní testy poskytovaly rozvojovým zemím hodnotnější informaci, než poskytují dosud. Nabízí se řešení regionálních testů s odpovídající obtížností, které budou dobře rozlišovat v rozvojových zemích s tím, že budou obsahovat úlohy, které umožní propojení na mezinárodní škály. V současnosti používaný způsob méně obtížné sady sešitů, o kterém referoval oddíl 4.1.1, sice umožňuje propojení, ale není schopen plně zohlednit situaci jednotlivých zemí a poskytnout jim dostatečně podrobnou zpětnou vazbu.

Aby bylo možno vytěžít plně mezinárodní výzkumy výsledků vzdělávání, bylo by potřeba výsledky mezi sebou provázat, tedy vytvořit společnou škálu pro výzkumy TIMSS a PISA a případně i pro výzkumy realizované v dospělé populaci. To je důležitý úkol, který vyžaduje spolupráci všech organizátorů.

Jak již bylo zmiňováno, opakovaně je diskutována možnost realizovat mezinárodní longitudinální výzkum, který by umožnil sledovat vzdělávací pokrok studentů v čase, a tedy spolehlivě určit faktory, které ovlivňují výsledky vzdělávání, respektive dopady opatření vzdělávací politiky. V případě longitudinálního výzkumu by se prezentace výsledků nezaměřovala na prosté srovnání, ale byla by založena na konceptu přidané hodnoty. Tento přístup by výrazně omezil možnost špatných interpretací a škodlivých dopadů výzkumů. Realizace longitudinálního výzkumu však naráží na jeho velkou finanční náročnost i na skutečnost, že řada vzdělávacích systémů již longitudinální výzkumy realizuje, a bylo by tedy pro ně obtížné mezinárodní a národní aktivity mezi sebou účelně propojit. Cyklická povaha výzkumů zaměřená na měření trendů (které je realizováno od roku 1995, respektive 2000) nicméně umožňuje testovat hypotézy týkající se dopadů vzdělávacích politik a reforem.

Záměr realizovat longitudinální výzkum byl diskutován v rámci příprav již zmiňovaného projektu šetření vysokoškoláků AHELO. Zde byl původní deklarovaný záměr poskytnout školám zpětnovazební informaci o přidané hodnotě, tedy o přínosu studia v rámci bakalářského cyklu. Zvažováno bylo buď provést testování v prvním a posledním ročníku bakalářského studia u těch samých studentů, nebo provést souběžně testování v obou ročnících. Tento aspekt je pro výzkum AHELO zcela zásadní, neboť jednou z důležitých překážek jeho realizace jsou obavy provázející využití, respektive zneužití získaných dat. Sem patří zejména obava z tvorby žebříčků zemí a univerzit, která by mohla vést k realokaci veřejných zdrojů a vést univerzity k důrazu na výuku k testům, a to na úkor vykonávání dalších významných úkolů, které univerzity mají. S tím se pojí argument, že testy umožňují měřit jen část výstupů, tedy hrozí nebezpečí příliš zjednodušených závěrů. Odpůrci výzkumů mají také obavu z dopadu testování na autonomii vysokých škol, kdy by mohlo docházet k homogenizaci vysokoškolského studia a omezení akademických svobod. Všechny tyto obavy by byly posílením zpětnovazební funkce výzkumu poněkud oslabeny, i když nelze očekávat, že by byly zcela eliminovány.

I v rámci OECD je dlouhodobě diskutována možnost vytvořit srovnávací výzkum pro žáky na vstupu do povinného vzdělávání, který by zachytil míru připravenosti žáků na vstup do povinného vzdělávání a její rozptyl v jednotlivých zemích a zároveň efektivitu povinného vzdělávání v jeho raném období. V této souvislosti byl členům vzdělávacího výboru OECD představen britský projekt IPIPS realizovaný Centrem pro evaluaci a monitoring (CEM) při Univerzitě v Durhamu. Britští výzkumníci vyvinuli adaptivní hodnocení, které umožňuje hodnotit předčtenářské a předmatematické dovednosti dětí na vstupu do povinného vzdělávání

⁶¹ Rozhodují pouze země OECD a země, kterým je přiznáno právo na spolurozhodování.

a v průběhu primárního vzdělávání. Umožňuje sledovat, jak přicházejí různé skupiny dětí do školy připraveny a také jaký pokrok učiní v prvních ročnících povinného vzdělávání. Má tedy longitudinální charakter⁶². Vývoj mezinárodního projektu přechodu z předškolního do primárního vzdělávání však nebyl dosud členskými zeměmi OECD schválen. Jedním z důvodů je obava z dopadů tohoto srovnání na vzdělávací systémy, kdy se odborníci obávají nácvičku na mezinárodní srovnávání již v předškolní kohortě. Další důvod je bezesporu finanční. Hodnocení je třeba s dětmi provádět individuálně prostřednictvím notebooků, což klade velké nároky nejen na technické vybavení, ale rovněž na počet a přípravu administrátorů.

Pro zkoumání významu kognitivních dovedností pro uplatnění na pracovním trhu jsou velmi důležité výzkumy v dospělé populaci. Umožňují studovat, jaké faktory souvisejí s vyšší mzdy, a také jaké aspekty ekonomiky vytvářejí poptávku po lidském kapitálu a dovednostech. O těchto skutečnostech je dosud k dispozici jen velmi málo informací. Tyto analýzy by umožnily ukázat význam měřených dovedností pro ekonomické výstupy, což by zároveň poskytlo informace o testech jako takových. Mnoho aktérů kritizuje testy a poukazuje na nebezpečí, že školy budou připravovat žáky na testy a zanedbávat jiné důležité vzdělávací aktivity. Někteří odborníci soudí, že pokud bychom měli doklady o tom, že testy predikují ekonomický úspěch, pravděpodobně bychom měli méně problémů s tím, že učitelé žáky k takovým testům připravují (Hanushek; Woessmann 2013). Tento doklad však neodstraní výtky vztahující se k tomu, že vzdělávání nepřipravuje pouze pro uplatnění na pracovním trhu. Naopak by tyto výtky mohl posílit.

Velmi důležitou oblastí, která bude v budoucnu rozvíjena, je bezpochyby spolupráce mezi odborníky realizujícími mezinárodní výzkumy výsledků vzdělávání a výzkumníky zabývajícími se výzkumy efektivity vzdělávání, jak bylo podrobně rozvedeno v oddíle 4.2.2. Cílem výzkumů v oblasti efektivity vzdělávání je „určit faktory na úrovni třídy, školy a na vyšších úrovních, které mohou přímo nebo nepřímo vysvětlit rozdíly ve výsledcích žáků při zohlednění žákovských charakteristik, jako jsou vrozené schopnosti, socioekonomický status a předchozí výsledky“ (Creemers; Kyriakides 2008). Modely efektivity vzdělávání tedy zahrnují faktory charakterizující vstupy, procesy a výstupy na všech zmiňovaných úrovních.

Učitelé při aplikaci výukových postupů vždy musí brát v úvahu aktuální výsledky žáka, a proto účinnost opatření nelze posuzovat bez zohlednění vstupních vědomostí a dovedností. Z výše uvedeného je tedy zřejmé, že mezinárodní výzkumy výsledků vzdělávání ve své současné podobě nemohou být základem pro výzkumy efektivity vzdělávání na úrovni jednotlivých zemí, neboť nemají longitudinální charakter (nezjišťují vstupní vědomosti a dovednosti jednotlivých žáků). Výzkumy efektivity vzdělávání přesto mohou profitovat z mezinárodních výzkumů a naopak mezinárodní výzkumy mohou profitovat z výzkumů efektivity. Při realizaci obou je však potřeba implementovat promyšlené a sofistikované postupy. Užitečnost poznatků získaných v rámci mezinárodních výzkumů významně zvyšuje promyšlená koncepce doprovodných dotazníků, kdy jsou doprovodné dotazníky založeny na konceptech, které prokazatelně zvyšují efektivitu vzdělávání (viz např. Creemers a Kyriakides 2010). Tedy koncepty, které jsou operacionalizovány v dotaznících, by měly v lepším případě být založeny na průniku výzkumných poznatků a představ tvůrců vzdělávacích politik. Dotazníky by neměly být v žádném případě založeny výlučně na představách tvůrců vzdělávacích politik, u kterých výzkumy již dávno prokázaly, že nefungují, což se v počátečních etapách výzkumů často dělo. V posledních cyklech výzkumů PISA i výzkumů IEA byly dotazníky koncipovány ve spolupráci s výzkumníky zabývajícími se

⁶² <http://www.cem.org/primary>

efektivitou vzdělávání, což v souvislosti s trendovými měřeními významně zvýšilo potenciál výzkumů zkoumat důležité otázky. Zároveň začaly být ve výzkumech obou organizátorů daleko větší měrou mapovány kontexty vzdělávacích systémů pomocí doprovodných studií i dotazníků pro národní koordinátory nebo vzdělávací experty v zúčastněných zemích.

Vzhledem k systematictějšímu přístupu ke koncepci doprovodných dotazníků bude v budoucnu k dispozici bohatá databáze, která by měla být využívána k testování dopadů efektivity rozmanitých opatření vzdělávací politiky. To je prvořadým úkolem mezinárodních výzkumů, neboť v rámci národních výzkumů je zkoumání dopadů národních politik daleko obtížnější.

Mezinárodní výzkumy mohou být i přes absenci longitudinální složky v rámci jednotlivých zemí užitečné pro zkoumání efektivity vzdělávání například tím, že mohou prostřednictvím korelačních analýz generovat hypotézy, které mohou být následně řádně testovány v rámci robustnějších designů, např. longitudinálních, experimentálních nebo intervenčních. Hypotézy z výzkumů efektivity vzdělávání mohou být v některých případech rovněž testovány v rámci mezinárodních výzkumů. Jedná se typicky o korelační analýzy, které ověřují například, zda výukové metody korelují s výsledky žáků nebo zda podpůrné klima ve třídě koreluje s motivací žáků ke studiu daného předmětu. Mezinárodní výzkumy se k testování podobných hypotéz znamenitě hodí, neboť mají velké reprezentativní výběry, vysokou návratnost a výbornou kvalitu získaných dat. Díky rozmanitosti zemí, které se do nich zapojují, umožňují také ověřovat mezikulturní a mezinárodní validitu poznatků výzkumů efektivity vzdělávání.

Spolupráce s odborníky na výzkum efektivity vzdělávání znamená zaměřit se na studium dynamické povahy školních jevů, tedy více se zajímat nejen o jevy samotné, ale o opatření, jejichž cílem je změna jevů a jejich úspěšnost, jak bylo zmiňováno v oddíle 4.2.2. Výzkumy efektivity vzdělávání také napovídají, že je třeba věnovat větší pozornost samotnému sběru dat, tedy klást si otázky, jaké informace mohou být věrohodně získány od jakých respondentů. Například v řadě případů získáváme od ředitelů škol data, která bychom měli získávat od učitelů nebo která bychom daleko spolehlivěji získali pozorováním ve třídě.

Dalším úkolem mezinárodních výzkumů je hledat souvislosti mezi kognitivními a nekognitivními výstupy a studovat dopad nekognitivních výstupů na další vzdělávací a životní dráhu. Tento výzkum je na samém počátku.

Mnozí odborníci (např. Strietholt a kol. 2014) zdůrazňují, že velmi důležitá bude v budoucnosti spolupráce různých disciplín s odlišnou tradicí. Ekonomové, sociologové a politologové typicky studovali sociální struktury, instituce a další jevy lokalizované na vyšších úrovních, přičemž pedagogové a psychologové studovali individuální rozdíly, a tedy se soustředili na nižší úroveň systému. Ekonometrie usuzuje na příčinné souvislosti. Psychometricki vyvinuli sofistikované modely pro testování kompetencí a postojů. Všechny těchto výdobytků využíváme v mezinárodních výzkumech výsledků vzdělávání a v analýze v nich získaných dat. Opatření vzdělávací politiky a jejich dopady je třeba studovat na úrovni jednotlivců i institucí a je třeba brát v úvahu víceúrovňovou povahu jevů. Tedy musíme kombinovat věcné i metodologické poznatky z rozmanitých disciplín.

Strietholt a kol. (2014) také zdůrazňují, že kromě spolupráce mezi jednotlivými vědeckými disciplínami je potřeba posílit také spolupráci mezi veřejnými a soukromými organizacemi. Mezinárodní testování zajišťují vesměs privátní organizace, jako je ETS, ACER, OECD, IEA. Tyto instituce popisují rozdíly mezi zeměmi, aniž by uspokojivě vysvětlovaly příčiny těchto rozdílů. Spolupráce s univerzitami by mohla posílit výzkumný aspekt budoucích mezinárodních výzkumů a pomoci využít stávající infrastruktury TIMSS a PISA přidáním hodnotných a informativních

národních rozšíření. V zemích, ve kterých nejsou akademická pracoviště zabývající se pedagogickým výzkumem příliš silná metodologicky, může spolupráce při realizaci výzkumů také naopak posloužit jako hodnotná vzdělávací zkušenost.

Jak bylo uvedeno v kapitole 4.1.2, důležitým aspektem, který zcela zásadně ovlivňuje měření vzdělávací cíle, efektivitu měření, jeho dostupnost pro různé skupiny respondentů i využití výsledků, je technologický rozvoj. Vzhledem k zapojení počítačových firem do konceptualizace vzdělávacích cílů (např. *Assessment and Teaching of 21st Century Skills*, o kterém podrobněji informuje oddíl 4.1.3) a vývoje platform pro testování, lze očekávat inovace, které záhy významně posunou možnosti mezinárodních vzdělávacích výzkumů. V tomto ohledu je zejména poučné sledovat oblast měření kompetencí k řešení problémů, kde dochází k překotnému vývoji testovacího prostředí i vlastního pohledu na povahu problémů, které je potřebné a možné simulovat a vyhodnocovat (viz např. Müller a kol. 2014)⁶³. Zároveň je ovšem třeba přemýšlet o aplikaci vyvinutých metod i do jiných oblastí, než je oblast řešení problémů v prostředí informačních technologií.

5.2.2 Využití mezinárodních výzkumů na národní úrovni – inspirace z Německa

Pro zvýšení užitečnosti mezinárodních výzkumů výsledků vzdělávání pro jednotlivé vzdělávací systémy je důležité cílenější a promyšlenější využívání zdrojů, které tyto výzkumy poskytují, na národních úrovních. V budoucnosti lze pravděpodobně očekávat, že budou zúčastněné vzdělávací systémy čím dál častěji obohacovat mezinárodní šetření vlastními longitudinálními komponentami tak, aby zvýšily užitečnost prostředků investovaných do nákladných šetření. V oddíle 4.2.3 byla zmiňována longitudinální sledování žáků s cílem sledovat vlivy kognitivních dovedností na další vzdělávací a životní dráhy žáků i provádění opakovaných měření vědomostí a dovedností.

Z hlediska námětů na využívání mezinárodních výzkumů pro získání informací užitečných pro výzkum a vzdělávací politiku na národní úrovni je možno se poučit v sousedním Německu. V roce 2012 vyšel v Pedagogické orientaci zajímavý text Manfreda Prenzla, který oceňuje pozitivní úlohu mezinárodních výzkumů v Německu, kde se podle jeho názoru díky těmto výzkumům podařilo popsat problémy německého školství. Upozorňuje však, že řadu otázek, které tyto výzkumy otevřely, se dosud nepodařilo zodpovědět, a že je třeba přistupovat k výzkumům a jejich možnostem strážlivě: pedagogická věda je věda jako každá jiná a jako taková má svá omezení. Nerealistická očekávání od výzkumů mohou být nebezpečná. Výzkum by však měl podle Prenzla produkovat solidní a významné poznatky relevantní pro vzdělávací politiku.

Prenzel konstatuje, že díky identifikaci problémů a soustředěné péči o jejich nápravu se podařilo v německém systému docílit mnoha pozitivních změn. Poznatky zároveň iniciovaly řadu dalších výzkumných aktivit, které našly i finanční zajištění. Jako příklad je možno jmenovat například rozšiřující studie spojené s mezinárodními výzkumy nebo německou panelovou

⁶³ Univerzita v Szegedu pořádá každoročně mezinárodní konferenci na téma hodnocení ve vzdělávání (*Conference on Educational Assessment*), před kterou se vždy koná workshop, jehož se pravidelně zúčastňují zástupci špičkových evropských pracovišť (zejména z Německa, Finska a Lucemburku) zabývajících se vývojem nástrojů na měření kompetencí k řešení problémů (*Szeged Workshop on Educational Evaluation*). Na tomto workshopu je možno velmi dobře sledovat, jak rychle se vyvíjí pojetí oblasti i technologické možnosti.

studii *National Education Panel Study (NEPS)*⁶⁴, která zkoumá velmi komplexně a longitudinálně všechny etapy vzdělávání.

Na vývoji NEPS je možno ilustrovat aktuální vývoj německého výzkumu v oblasti vzdělávání. Potřeba poučit se více o fungování systému a rigorózně ověřit skutečnosti, které vyplývaly z mezinárodních výzkumů, vedla v posledních dvou dekadách v Německu k realizaci několika regionálních longitudinálních výzkumů, které byly zaměřeny na sledování rozvoje vědomostí a dovedností, rozhodování o vzdělávacích drahách, na sledování přechodů z primárního do sekundárního vzdělávání, dopadů složení školní třídy na výsledky žáků (např. BIKS v Bavorsku a Hessensku, KOALA-S v Bavorsku a Sasku, MEPS v Severním Porýní Vestfálsku, LAU a KESS v Hamburku, ELEMENT v Berlíně). S cílem získat komplexní informaci o vývoji ve všech spolkových zemích zahájilo Německo v roce 2007 NEPS, který studuje velmi důkladně souvislost mezi rozhodováním o vzdělávací dráze, vzdělávacími procesy v různém vzdělávacím prostředí a rozvojem vědomostí a dovedností v průběhu celé vzdělávací dráhy. Jeho cílem je shromáždit velmi kvalitní a podrobná data, která budou moci být analyzována dalšími výzkumníky v Německu i v jiných zemích a tak výzkumné aktivity zefektivnit a zabránit jejich dublování (Blossfeld a kol. 2009). NEPS se zaměřuje na 8 etap celoživotního vzdělávání⁶⁵:

Etapa 1: Vzdělávání v novorozeneckém a raném věku

Etapa 2: Z mateřské do primární školy

Etapa 3: Z primární školy do nižšího sekundárního vzdělávání

Etapa 4: Z nižšího do vyššího sekundárního vzdělávání

Etapa 5: Gymnázium a přechod do terciárního vzdělávání, učební obor, nebo trh práce

Etapa 6: Z odborného vzdělávání na trh práce

Etapa 7: Z terciárního vzdělávání na trh práce

Etapa 8: Další a celoživotní vzdělávání

Ve všech etapách je zkoumáno pět hlavních teoretických oblastí:

- Rozvoj kompetencí v průběhu životní dráhy
- Vzdělávací procesy ve specifických vzdělávacích prostředích v průběhu životní dráhy
- Sociální nerovnosti a rozhodování o vzdělávacích drahách v průběhu životní dráhy
- Vzdělávání migrantů
- Návratnost vzdělávání v průběhu životní dráhy
- Motivační charakteristiky a osobnostní aspekty v průběhu životní dráhy

To znamená, že ve všech etapách jsou měřeny kognitivní i nekognitivní výsledky vzdělávání a je velmi podrobně mapováno učební prostředí. Děti, žáci a studenti jsou z toho důvodu vybíráni přes školy, aby bylo možno sledovat vliv kvality institucí a organizací. Projekt je zatím plánován do roku 2025. Je koordinován z Univerzity v Bamberku a podílejí se na něm přední výzkumníci z celého Německa tak, aby byla jeho kvalita co nejvyšší. Podoba výzkumných nástrojů je v odborné komunitě široce diskutována (Blossfeld; Schneider; Doll 2009). Výzkum NEPS byl bezesporu ovlivněn poznatky z mezinárodních výzkumů i jejich metodologií. Jeho design je

⁶⁴ <https://www.neps-data.de/en-us/home.aspx>

⁶⁵ Nejedná se tedy o výzkum v pravém slova smyslu longitudinální, kdy by nějací jedinci byli sledováni ve všech životních etapách, ale o výzkum kvazilongitudinální, kdy paralelně sledujeme několik věkových kohort.

však navržen tak, aby reagoval na potřeby německé odborné komunity a německé vzdělávací politiky. Dobře ilustruje dopad mezinárodních výzkumů na národní výzkumné aktivity a je velmi pravděpodobné, že podobným způsobem se bude vyvíjet výzkum v oblasti vzdělávání i v řadě dalších zemí, ve kterých dosud velké výzkumy nenašly své uplatnění.

Mezinárodní výzkumy mají v Německu odraz nejen v dalších výzkumných aktivitách, ale rovněž v programech, které reagují na zjištěné nedostatky s cílem je odstranit. Například v návaznosti na výzkum TIMSS 1995, ve kterém dosáhlo Německo průměrných výsledků v matematice a přírodovědných předmětech, byl vytvořen program SINUS, jehož cílem bylo zlepšit výuku matematiky a přírodovědných předmětů. Program byl nejprve zahájen na druhém stupni povinného vzdělání (1997-2007), následně byl realizován rovněž na prvním stupni (2004-2013). Nositelem programu byl Leibnitzův Institut pro výuku přírodních věd a matematiky (IPN). Ten vytvořil obsah projektu a strukturu pro jeho realizaci ve všech spolkových zemích. Základ programu tvořilo 10 teoreticky a empiricky podložených modulů, které byly nabízeny učitelům k profesnímu rozvoji, a 5 stěžejních úkolů.

Moduly:

1. Kvalitní úlohy.
2. Objevovat, zkoumat, vysvětlovat.
3. Podchytit žákovské představy, rozvíjet základní myšlenky.
4. Rozpoznat učební obtíže – podporovat učení s porozuměním.
5. Objevit a podporovat talenty.
6. Mezipředmětové vyučování.
7. Podchytit zájmy chlapců a děvčat a dále je rozvíjet.
8. Samostatné učení – společné učení
9. Provázet učení – posuzovat učební výsledky.
10. Usnadnit přechody.

Stěžejní úkoly:

1. Proměna vzdělávacích standardů v matematice.
2. Učení o souvislostech v přírodovědě.
3. Podpora žáků se specifickými vzdělávacími potřebami.
4. Usnadnění přechodů.
5. Zpětná vazba založená na datech.

Do programu se zapojilo 3 200 škol. Program poskytoval pedagogům velkou svobodu, neboť jim umožňoval volit si vzdělávací moduly podle svých aktuálních potřeb a zájmů.

Efektivita programu byla vyhodnocována prostřednictvím mezinárodních výzkumů, a to na druhém stupni pomocí výzkumu PISA a na prvním stupni pomocí výzkumu TIMSS (Dalehefte a kol. 2015). Na reprezentativních výběrech škol, které byly zapojeny do programu SINUS, a škol zapojených do mezinárodních výzkumů byly porovnávány nejen výsledky, ale rovněž zájmy o sledované předměty. V případě výsledků nebyly porovnávány pouze průměrné výsledky, ale také rozložení žáků na jednotlivých úrovních způsobilosti, kde bylo sledováno, zda se daří lépe rozvíjet talenty (větší podíl respondentů na nejvyšší úrovni) a zda se daří lépe vycházet vstříc žákům se specifickými vzdělávacími potřebami (nižší podíl žáků na nižších úrovních způsobilosti). Mezinárodní výzkumy tedy nejen iniciovaly realizaci programu, ale sloužily rovněž k vyhodnocení jeho efektivity.

Další důležitou oblastí, ve které docházelo k vzájemné inspiraci mezi národními a mezinárodními výzkumy, byly videostudie výuky. Němečtí výzkumníci se inspirovali videostudiemi TIMSS, ale navázali na ně vlastními šetřeními, která lépe vyhovovala potřebám oborových didaktik. V návaznosti na videostudii TIMSS 1999 byla v Německu realizována videostudie IPN, která podrobně zkoumala výuku fyziky, videostudie DESI, která zkoumala výuku angličtiny jako cizího jazyka, a videostudie Pythagoras, která proběhla i ve Švýcarsku a zkoumala efektivitu výukových metod ve výuce vybraných matematických témat (Janík; Seidel 2009). Němečtí odborníci zároveň svými zkušenostmi významně přispěli k analýze dat z videostudie TIMSS 1999. V České republice na tyto aktivity navázaly videostudie výzkumného týmu z brněnského Institutu výzkumu školního vzdělávání MU, které významně rozšířily povědomí české výzkumné komunity o možnostech videostudií a přinesly řadu pozoruhodných aplikací v aktivitách směřujících ke zkvalitňování výuky (Najvar a kol. 2011). Není pochybností o tom, že při snaze o porozumění charakteristikám kvalitní a efektivní výuky a vysvětlování rozdílů mezi jednotlivými vzdělávacími systémy bude význam videostudií i v budoucnu stoupat.

Při přemýšlení o užitečných poznatcích, na něž by se měl v budoucnosti zaměřit mezinárodní výzkum, rozlišuje Prenzel (2012) mezi věděním o cílech, popisným věděním, kauzálním věděním a věděním pro změnu. Věděním o cílech je podle jeho názoru třeba kultivovat, neboť je zásadní pro ostatní druhy věděním. Za nejrozvinutější považuje Prenzel věděním popisné, i když i zde upozorňuje na skutečnost, že některé důležité problémy jsou ve výzkumu opomíjeny. U vysvětlujícího výzkumu upozorňuje na důležitost správné specifikace ověřovaných modelů, neboť poznání je často zkresleno skutečností, že některé důležité proměnné jsou ve výzkumu opominuty a v modelu chybí. Zdůrazňuje, že kauzální vztahy je třeba ověřovat opakovaně. Za nejdůležitější a zároveň nejzanedbávanější se považuje výzkum pro změnu, který se ovšem musíme naučit lépe realizovat.

Autor se domnívá, že by mělo být iniciováno více projektů a programů, jež mohou přispět k vysvětlujícímu věděním a k věděním pro změnu. Zároveň je třeba reflektovat hranice těchto programů a mít od nich realistická očekávání. Současně připomíná, že i nadále bude mít velký význam výzkum popisný, a proto je třeba ho dále kultivovat. V této souvislosti zdůrazňuje význam promyšleného teoretického modelování a hluboké zorientovanosti ve zkoumaných problémech. Pro budoucnost výzkumu je významné, aby se empirický výzkum vzdělávání zaměřil na důležité problémové oblasti a rozvinul se do šíře – také pokud jde o výzkumné designy a o aplikaci smíšených výzkumných metod, které mají potenciál lépe porozumět sledovaným jevům. Prenzel končí tím, že by bylo zapotřebí více odvahy k tomu vydat se do ještě neprozkoumaných oblastí.

Všechny tyto úvahy mohou stimulovat přemýšlení o národních výzkumných (ale i implementačních) aktivitách, které mohou těžit z mezinárodních výzkumů výsledků vzdělávání i v ČR.

6. Shrnutí poznatků z mezinárodních výzkumů týkajících se českého vzdělávacího systému

Tato kapitola předkládá vybrané poznatky, které jsme získali o českém vzdělávání z mezinárodních výzkumů výsledků vzdělávání od roku 1995 do roku 2015. První část kapitoly se zabývá kognitivními a nekognitivními výsledky, druhá část se zabývá vzdělanostními nerovnostmi a třetí část se zabývá poznatky o školách a učitelích. Přehled účasti ČR v mezinárodních výzkumech výsledků vzdělávání je uveden v příloze. Zde představený výběr zjištění v žádném případě neaspiruje a ani nemůže aspirovat na úplnost. Přináší poznatky, které autorce této publikace připadají důležité a kterým by podle jejího názoru měla být věnována pozornost ze strany výzkumníků i tvůrců vzdělávací politiky. V oblasti nekognitivních výstupů a poznatků o systému se tedy zpravidla jedná o zjištění, která poukazují na jevy, ve kterých se Česká republika dlouhodobě a konsistentně odlišuje od ostatních vyspělých zemí.

6.1 Silné a slabé stránky v měřených oblastech a dlouhodobé trendy

Tento oddíl podává přehled celkových výsledků ve všech věkových skupinách, ve kterých byly realizovány mezinárodní výzkumy. Zabývá se všemi oblastmi, které byly testovány opakovaně, tedy matematikou, přírodními vědami, řešením problémů, informačními technologiemi a občanskou výchovou. Poznatky týkající se výsledků jsou sumarizovány nikoli primárně podle jednotlivých oblastí vzdělávání, ale podle stupňů vzdělávání, protože to lépe odpovídá logice výzkumů.

6.1.1 Kognitivní výsledky

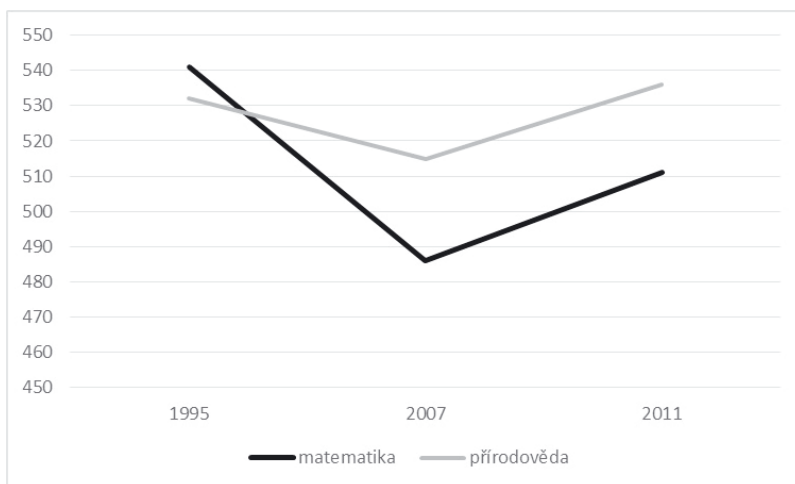
Výsledky žáků 1. stupně

V populaci žáků čtvrtých ročníků proběhlo v rámci výzkumu TIMSS celkem třikrát šetření v matematice a v přírodovědných předmětech: v letech 1995, 2007 a 2011 (výsledky z roku 2015 nebyly v době dokončení této publikace ještě k dispozici). Výsledky jsou vyneseny v grafu 1.

Basl a Tomášek (2013) srovnali vývoj výsledků žáků 4. ročníků v matematice a v přírodovědě ve 14 zemích, které se zúčastnily výzkumu v letech 1995, 2007 a 2011. Při srovnání těchto zemí zaznamenala ČR mezi roky 1995 a 2011 největší zhoršení. Podrobnější pohled na výsledky zároveň ukazuje, že se nejvíc zhoršili žáci s nejlepšími výsledky (jejich podíl nejvíce poklesl).

V přírodovědě jsou zajímavé rozdíly ve výsledcích v jednotlivých kognitivních oblastech. Pro české žáky byly relativně nejobtížnější úlohy zaměřené na zdůvodňování a uvažování. V těchto úlohách dosahovali čeští žáci s dobrými i špatnými výsledky horších výsledků než v přírodovědě jako celku, přičemž se zhoršujícími výsledky se rozdíl mezi celkovým výsledkem a výsledkem v úlohách na zdůvodňování prohluboval. Výzkum ukázal, že žáci s nejslabšími výsledky nedokáží správně argumentovat a neumí dostatečně vysvětlit a zdůvodnit svá tvrzení (Basl; Tomášek 2013).

Graf 1: Výsledky žáků 4. ročníků v šetření v matematice a přírodovědných předmětech (TIMSS, 4. ročník ZŠ)



Zdroj dat: Martin a kol. 2012, Mullis a kol. 2012c

Šetření v oblasti čtenářské gramotnosti proběhlo rovněž třikrát: v letech 1995, 2001 a 2011. V roce 1995 provedla ČR se zpožděním 3 let výzkum IEA Reading Literacy Study, který se v ostatních zemích uskutečnil v roce 1992. Výsledky byly následně přepočítány na škálu, na které byly porovnávány výsledky zemí, jež se zapojily do výzkumu v roce 1992. Podle tehdejší definice testované populace se výzkumu zúčastnili žáci z 3. ročníku ZŠ. Následovaly výzkumy PIRLS, kde již byla definice zpřesněna a odpovídala 4. ročníku ZŠ. Výzkumy RLS a PIRLS měly odlišnou koncepci a nebyla mezi nimi vytvořena přímá návaznost. Zde je tedy velmi obtížné hovořit o trendech. Mezi lety 2001 a 2011, kdy již probíhal výzkum PIRLS koncipovaný pro sledování trendů, došlo však v každém případě v ČR ke statisticky významnému zlepšení (z 537 na 545 bodů) (Mullis a kol 2012b).

Ve výzkumech IEA nemá velký smysl porovnávat výsledky s mezinárodním průměrem, neboť se výzkumu účastní rozmanité množství různorodých zemí. Informace o tom, zda ČR skončila pod průměrem či nad průměrem těchto zemí, nemá tudíž žádnou vypovídací hodnotu. Zajímavé jsou trendy, které ukazují, že v roce 2007 došlo k jakémusi jednorázovému zhoršení. Porovnáme-li pouze výsledky v letech 1995 a 2011, můžeme konstatovat statisticky významné zhoršení v matematice a setrvalý stav v přírodovědných předmětech. Statisticky významné zlepšení ve čtenářské gramotnosti již bylo zmiňováno.

Určitou informaci o mezinárodním kontextu můžeme získat ze srovnání pořadí v rámci určité homogennější skupiny zemí. Tabulka 23 uvádí srovnání se zeměmi OECD a EU. Jejich počty byly v různých šetřeních rovněž různé, ale úroveň vědomostí a dovedností je zde srovnatelnější. První číslo udává pořadí České republiky v žebříčku, druhé číslo udává počet zemí OECD/EU, které se šetření zúčastnily. Z tohoto (velmi hrubého) porovnání je zřejmé, že v počátcích účasti ČR v komparativních výzkumech byla ČR relativně úspěšnější v matematice než ve čtenářské gramotnosti. V posledních výzkumech již se relativní výsledky vyrovnaly a jsou v obou dis-

ciplínách víceméně srovnatelné. Výsledky v přírodovědě se v tomto pohledu jeví dlouhodobě relativně lepší.

Tabulka 23: Pořadí českých žáků 4. ročníků ZŠ v mezinárodních výzkumech

TIMSS	1995	2007	2011
matematika	4/19	18/19	16/28
přírodovědné předměty	7/19	15/19	6/28
RLS/PIRLS	1995	2001	2011
čtenářská gramotnost	12/22	12/21	12/28

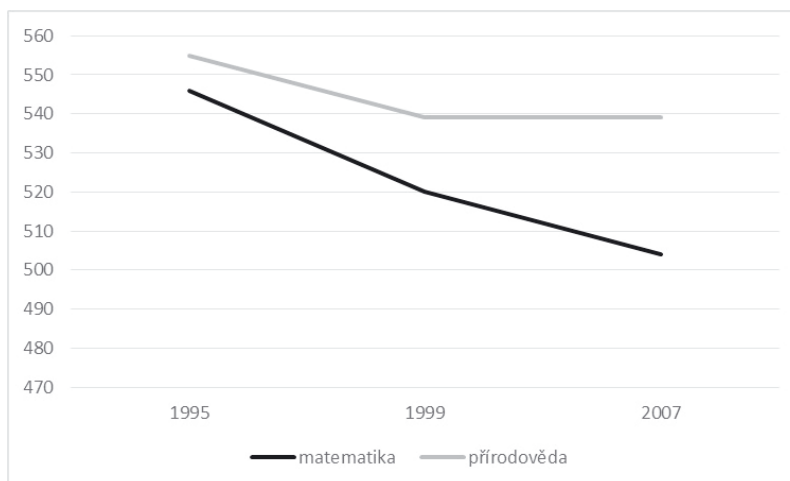
Zdroj: Straková, Veselý 2013

Výsledky žáků 2. stupně

Výše zmiňované trendy, tedy zhoršování v matematice a s tím spojené relativní vyrovnávání výsledků v matematice a ve čtenářské gramotnosti, potvrzují i výsledky žáků 2. stupně povinného vzdělávání. Potvrzují rovněž relativně silnou pozici českých žáků v přírodovědných předmětech.

V grafu 2 uvádíme výsledky žáků 8. ročníků ve výzkumech TIMSS v letech 1995, 1999 a 2007. Zde je zřejmé, že od roku 1995 se výsledky českých žáků v přírodovědě zhoršily, ale v menší míře než výsledky v matematice. Máme k dispozici rovněž jedno měření čtenářské gramotnosti v 8. ročníku z výzkumu RLS, tam se čeští žáci umístili mírně pod průměrem zúčastněných zemí (na 13. místě z 20 zemí EU/OECD).

Graf 2: Výsledky žáků 8. ročníků v šetření v matematice a přírodovědných předmětech (TIMSS, 8. ročník ZŠ)



Zdroj dat: Martin a kol. 2008, Mullis a kol. 2008

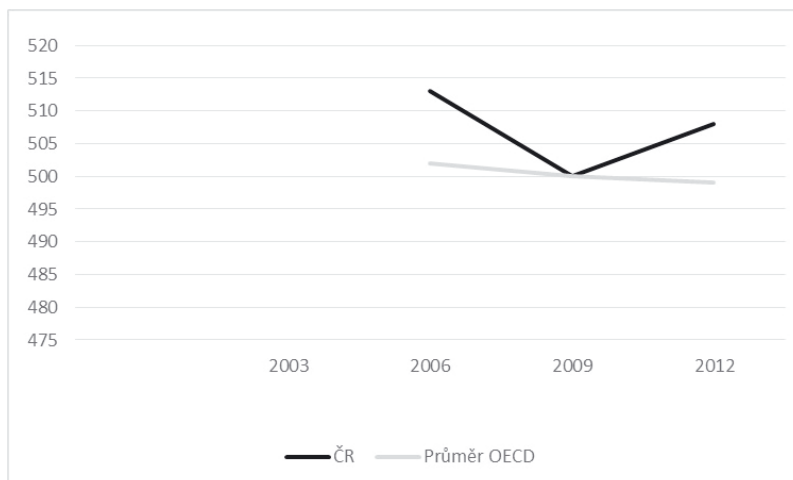
Výsledky šetření IEA potvrzují rovněž výsledky šetření PISA. V grafech 3, 4 a 5 jsou uvedeny postupně výsledky z matematické, přírodovědné a čtenářské gramotnosti. Zde je třeba si připomenout, že výsledky jsou porovnatelné vždy od roku, v němž byla daná oblast hlavní testovanou oblastí. Tedy výsledky v matematice od roku 2003, v přírodních vědách od roku 2006 a ve čtenářské gramotnosti již od roku 2000. V každém z grafů je vždy vyneseno srovnání s průměrem zemí OECD, které se zúčastnily šetření v tomto roce (zpravidla to byly všechny členské země OECD).

Graf 3: Výsledky patnáctiletých žáků v matematické gramotnosti (PISA)



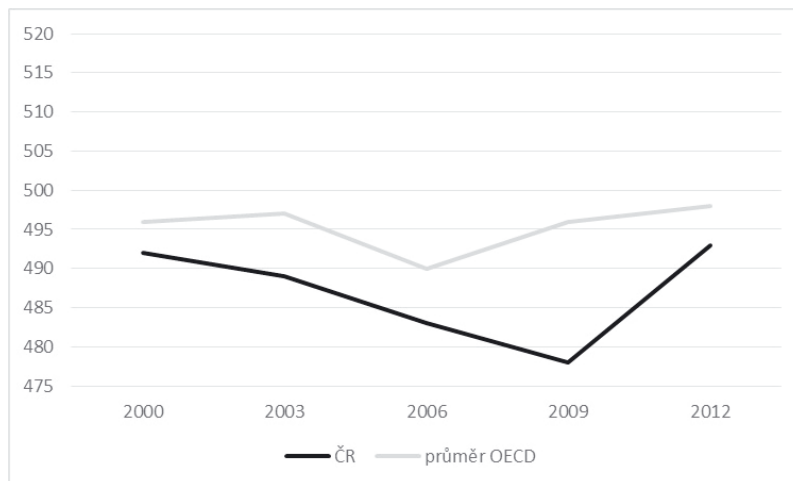
Zdroj dat: OECD 2013b

Graf 4: Výsledky patnáctiletých žáků v přírodovědné gramotnosti (PISA)



Zdroj dat: OECD 2013b

Graf 5: Výsledky patnáctiletých žáků ve čtenářské gramotnosti (PISA)



Zdroj dat: OECD 2013b

V matematice spadli čeští žáci z nadprůměru k průměru zemí OECD, v přírodovědné gramotnosti se po výpadku v roce 2009 vrátili zpět do nadprůměru a ve čtenářské gramotnosti se pohybují rovněž v blízkosti průměru OECD.

Kromě trendů v jednotlivých měřených oblastech se shodují výzkumy IEA v obou populacích s výzkumy PISA rovněž v určitém propadu výsledků v letech 2007, respektive 2009. Tento propad neumíme s jistotou vysvětlit. Je možné, že nemá žádnou spojitost s vědomostmi a dovednostmi žáků, ale že se týká motivace žáků se v mezinárodních výzkumech snažit. Když se zvedla po zveřejnění výsledků v tomto období vlna kritiky týkající se úpadku českého vzdělání, mohla se zvýšit motivace učitelů (a potažmo žáků) ukázat, že kritika není na místě. Toto je ovšem čirá spekulace, exaktní vysvětlení pro tento propad nemáme.

Jak bylo uvedeno v kapitole 1, v rámci mezinárodních výzkumů jsou kromě čtenářských dovedností, matematiky a přírodovědných předmětů zkoumány i další oblasti, ve kterých probíhají šetření s menší četností. Jednou z těchto oblastí je řešení problémů, které bylo sledováno ve výzkumu PISA v letech 2003, 2012 a 2015 (kde jsme však v době dokončení této publikace neměli výsledky k dispozici). V této oblasti není možno přímo sledovat trendy, neboť výsledky z šetření 2003 a 2012 nebyly přepočítány na stejnou škálu. Důvodem byla skutečnost, že v roce 2003 bylo řešení problémů koncipováno jako statické a administrováno prostřednictvím písemného testu. V roce 2012 byl koncept rozšířen o dynamické úlohy, což bylo umožněno počítačovou administrací. Výsledky českých žáků ve srovnání s žáky z ostatních zemí OECD byly v roce 2003 i v roce 2012 nadprůměrné. Mezinárodní zpráva uvádí, že výsledek při počítačové administraci v roce 2012 byl značně ovlivněn schopností žáků pracovat s počítačem. Rozdíly mezi výsledky žáků, kteří doma používají počítač a kteří nikoli, byly v ČR nejvyšší ze všech zúčastněných zemí. Naopak používání počítačů ve škole nemělo na výsledky žáků statisticky významný vliv.

Tento výsledek odpovídá výsledku českých žáků ve studii IEA ICILS, která proběhla v roce 2013 a zkoumala dovednosti žáků 8. ročníků v práci s počítačem. Zde se čeští žáci umístili

v porovnání zúčastněných 19 zemí na prvním místě a předstihli například žáky z Austrálie, Nizozemska, Německa, Norska a Švýcarska. Tento výsledek byl poněkud překvapivý pro kritiky výuky informačních technologií na českých školách. Výsledky výzkumu PISA 2012 však naznačují, že počítačová zdatnost českých žáků se školní výukou příliš nespojuje.

Další zajímavé výsledky přinesly výzkumy občanské výchovy, které se uskutečnily v 8. ročnících povinného vzdělávání v letech 1999 a 2009. Tyto výzkumy zkoumaly znalosti v oblasti fungování demokratické společnosti a postoje žáků. Z hlediska sledovaných vědomostí se čeští žáci umístili mezi zúčastněnými zeměmi v obou šetřeních nad mezinárodním průměrem, mezi lety 1999 a 2009 došlo však ve znalostech českých žáků 8. ročníků ke statisticky významnému zhoršení. V rámci Evropské unie se čeští žáci svými vědomostmi zařadili do průměru, průměrné byly také jejich specifické vědomosti týkající se fungování Evropské unie. Postoji žáků se zabýváme v oddíle 6.1.2., který je věnován nekognitivním výsledkům vzdělávání.

Výsledky žáků středních škol

Vzhledem k tomu, že výzkum PISA testuje žáky prvního ročníku středních škol, nemůže dobře vypovídat o výsledcích žáků ovlivněných středoškolským vzděláním. To je velká škoda, neboť ve většině vyspělých zemí je účast ve středoškolském vzdělání téměř univerzální a o tom, jak je žák vybaven pro vstup na pracovní trh nebo do terciárního vzdělávání, vypovídá nejlépe právě výsledek po ukončení středního vzdělání (úroveň ISCED 3). Organizátoři výzkumů asociace IEA měli zprvu aspiraci populaci žáků středních škol do pravidelného cyklu výzkumů zahrnout, ovšem záhy narazili na problémy související s tím, že účast ve středním vzdělání je v jednotlivých zemích příliš odlišná a že se zároveň od sebe velmi liší jednotlivé středoškolské programy. Usoudili tedy, že rigorózní srovnání celé populace středoškoláků není možné. V současnosti se výzkumy v populaci žáků vyšších ročníků střední školy realizují, ale pouze na části populace s pokročilejší výukou matematiky a fyziky (*TIMSS advanced*). Tohoto šetření se Česká republika dosud nezúčastnila.

Šetření v populaci žáků ukončujících střední školu se tedy v ČR v uplynulých dvaceti letech uskutečnilo pouze dvakrát. Poprvé byli v rámci výzkumu TIMSS v roce 1995 testováni žáci v posledním ročníku středoškolského vzdělání a byla posuzována úroveň jejich matematické a přírodovědné gramotnosti. Zároveň byli z matematiky a fyziky testováni žáci, kteří se na ni nějakým způsobem v rámci středoškolského studia specializovali. V ČR byli do tohoto srovnání zahrnuti žáci gymnázií, neboť cílová skupina měla ideálně tvořit cca 20 % středoškoláků. Z hlediska přírodovědných i matematických vědomostí a dovedností se žáci posledních ročníků všech typů středoškolského studia zařadili ve srovnání 21 vyspělých zemí do mezinárodního průměru. Výsledky gymnazistů v matematice a fyzice byly v mezinárodním srovnání podprůměrné. Zde se ukázalo, že skupina žáků, kteří dosahovali výborných výsledků v matematice a fyzice, je v ČR příliš malá, přičemž v ostatních zúčastněných zemích se dařilo pro tyto disciplíny získat větší podíl žáků, kteří si je osvojili na výborné úrovni (Straková; Tomášek; Palečková 1998).

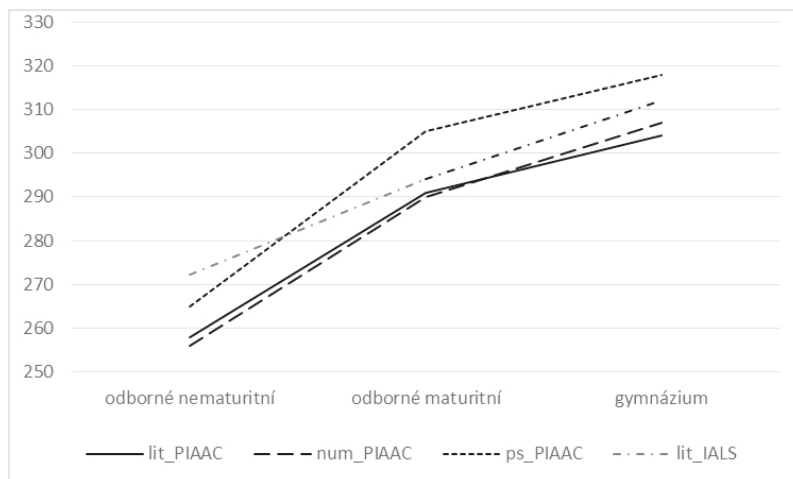
Podruhé se žáci v posledních ročnících středních škol zapojili v roce 1999 do šetření občanské výuky CIVED. Zde se svými znalostmi a dovednostmi zařadili do průměru 16 zúčastněných zemí.

Pro Českou republiku představuje skutečnost, že mezinárodní organizace rezignovaly na šetření celé populace žáků na konci středoškolského vzdělání, velkou nevýhodu nejen proto, že je v ČR míra dokončování středoškolského vzdělání velmi vysoká (jedna z nejvyšších mezi zeměmi OECD), ale také díky struktuře systému středoškolského vzdělání. Je důležité vědět, do jaké míry trpí vědomosti a dovednosti mladých lidí opouštějících regionální školství vysokou

diferenciací systému, tedy do jaké míry uspívají absolventi jednotlivých větví v mezinárodním srovnání. V současné době nemáme ani řádné národní srovnání výstupních vědomostí absolventů jednotlivých oborů středního vzdělání. Představu o rozdílech ve vědomostech maturantů z různých oborů si sice můžeme učinit na základě výsledků z maturitní zkoušky, absolventi učebních oborů však maturitní zkoušku neskládají, tedy nemáme žádnou srovnatelnou informaci o jejich vědomostech a dovednostech v základních oblastech vzdělávání při srovnávání s ostatními středoškoláky.

Tento nedostatek se pokusili alespoň částečně napravit členové českého realizačního týmu mezinárodního výzkumu dospělých PIAAC, kteří v rámci šetření matematické a čtenářské gramotnosti a dovedností řešit problémy v prostředí informačních technologií provedli v ČR nadvýběr mladých lidí ve věku 16-34 let, aby bylo možno srovnat výsledky absolventů jednotlivých typů středoškolského vzdělání. Toto srovnání je uvedeno v grafu 6 pro všechny tři sledované oblasti ve výzkumu PIAAC (2013) a pro čtenářskou gramotnost ve výzkumu IALS (1998). Do výzkumu byli zařazeni mladí lidé ve věku 16-34 let, kteří již ukončili studium.

Graf 6: Srovnání výsledků absolventů jednotlivých typů středoškolského studia



Zdroj: Straková, Veselý 2013

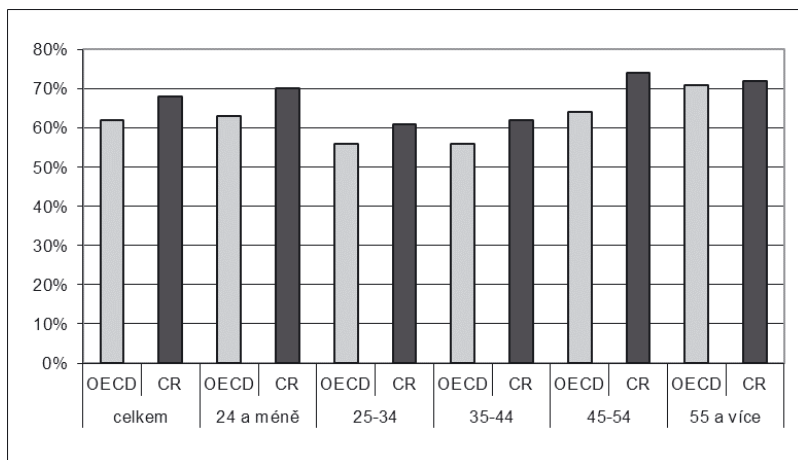
Graf ukazuje, že od roku 1998 do roku 2012 se zvýšily již tak vysoké rozdíly ve výsledcích absolventů maturitních a nematuritních oborů. Ve výzkumu PIAAC byly rozdíly mezi obory zhruba stejně velké ve všech sledovaných oblastech, tedy nezdá se, že by absolventi odborných oborů byli v některé ze sledovaných oblastí více znevýhodněni. Rozdíly mezi absolventy odborných maturitních a odborných nematuritních oborů tvořily ve všech případech zhruba 1 směrodatnou odchylku.

Graf 7 ukazuje podíl absolventů nematuritních oborů (klasifikace ISCED 353), kteří nedosáhli základní úrovně čtenářských dovedností⁶⁶, ve všech věkových kohortách. Vidíme, že tento

⁶⁶ Ve výzkumu PIAAC byla třetí (prostřední) úroveň způsobilosti považována za standard potřebný pro řešení běžných úkolů dospělého života.

podíl je v ČR ve všech věkových kohortách s výjimkou té nejstarší vyšší, než je průměr OECD. Zároveň je zastoupení absolventů nematuritních oborů (ISCED 353) v ČR vyšší, než je průměr v zemích OECD.

Graf 7: Podíl absolventů nematuritních oborů, kteří nedosáhli základní úrovně čtenářské gramotnosti dle věku, PIAAC 2012



Zdroj dat: databáze PIAAC 2012

Toto srovnání je velmi hrubé, ale napovídá, že všeobecným vědomostem a dovednostem českých absolventů nematuritních oborů by bylo vhodné věnovat zvýšenou pozornost. Výzkum PIAAC bohužel neumožňuje provést mezinárodní srovnání maturitních a nematuritních oborů, neboť srovnatelná informace o absolvování jednotlivých středoškolských programů nebyla bohužel v datech k dispozici. Možnost srovnání velice komplikuje skutečnost, že odborné a zejména učňovské vzdělávání se v jednotlivých zemích velice liší z hlediska obsahu, míry diferenciací a prostupnosti na vyšší úroveň vzdělání. Pro ČR je charakteristická vysoká diferenciací, tedy v mezinárodním srovnání vysoký počet úzce specializovaných učebních oborů, a omezená prostupnost (učební obory s výučním listem nemají přímou prostupnost do žádného typu terciárního vzdělání).

Srovnání úrovně absolventů bylo provedeno pro absolventy terciárního vzdělání, kde je srovnatelnost vyšší. Tam se naopak jevila úroveň českých absolventů ve srovnání zemí OECD jako nadprůměrná (Straková; Veselý 2013).

Výsledky v populaci dospělých

Tendence, které pozorujeme v žákovských populacích, potvrzují rovněž výsledky z výzkumů dospělých. Zde máme ovšem k dispozici pouze dva měřící body: rok 1998, kdy proběhl výzkum IALS, a rok 2012, kdy proběhl výzkum PIAAC. Tedy hovořit o trendech v celé populaci by bylo poněkud předčasné. Ve čtenářské gramotnosti se čeští dospělí ve věku 16-65 let za sledovaných 14 let nezhoršili, rozdíl ve výsledku byl v toleranci statistické chyby. V mezinárodním srovnání 24 zemí OECD se v obou sběrech umístili v průměru.

Metodologie výzkumů PIAAC a IALS bohužel neumožňuje porovnávat rovněž výsledky v matematice, neboť koncept kvantitativní gramotnosti sledované ve výzkumu IALS byl natolik odlišný od konceptu numerické gramotnosti sledované v PIAAC, že nebylo možné vytvořit společnou škálu, na které by bylo možno provést přímé porovnání. Můžeme ovšem vycházet z toho, že jak kvantitativní, tak numerická gramotnost zahrnují nějaké matematické dovednosti, a podívat se, jak se změnilo relativní pořadí ČR ve vztahu k matematickým dovednostem ve srovnání s ostatními zeměmi, které se zúčastnily obou výzkumů. Pro úplnost uvádíme rovněž porovnání pořadí na škále čtenářské gramotnosti. Z tabulky je zřejmé, že toto (velmi hrubé) srovnání ukazuje relativní zhoršení v matematických dovednostech a relativní zlepšení v dovednostech čtenářských. V matematice se čeští dospělí umístili v obou výzkumech mezi zúčastněnými zeměmi v nadprůměru.

Tedy i v populaci dospělých zaznamenáváme vyrovnávání rozdílů mezi matematickými a čtenářskými dovednostmi s tím, že v matematice dochází k určitému zhoršení.

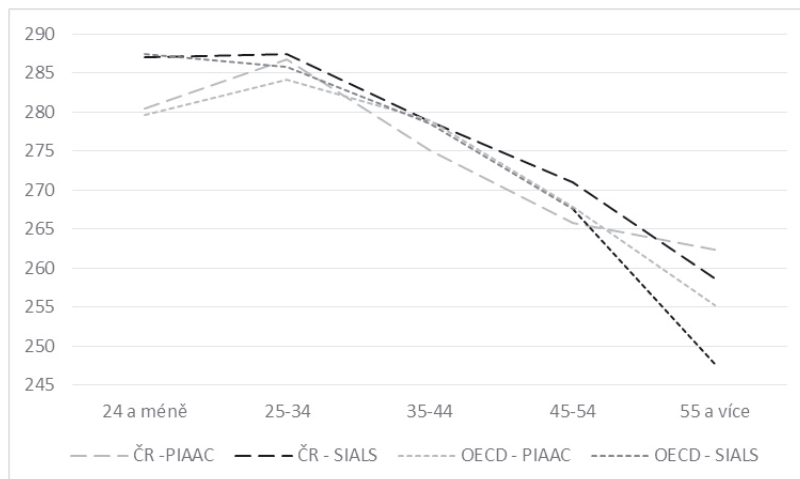
Tabulka 24: Umístění zemí podle výsledků a rozdílů v pořadích ve výzkumech PIAAC a IALS

	numerická gramotnost			čtenářská gramotnost		
	pořadí PIAAC	pořadí IALS	dif	pořadí PIAAC	pořadí IALS	dif
Austrálie	9	10	1	3	11	8
Vlámsko	2	8	6	6	8	2
Kanada	10	9	-1	8	7	-1
Česká republika	7	3	-4	7	9	2
Dánsko	6	2	-4	9	3	-6
Finsko	1	7	6	1	4	3
Německo	8	5	-3	10	6	-4
Irsko	12	12	0	13	12	-1
Nizozemsko	3	6	3	2	5	3
Norsko	5	4	-1	5	2	-3
Polsko	11	13	2	12	13	1
Švédsko	4	1	-3	4	1	-3
Spojené státy	13	11	-2	11	10	-1

Zdroj: Straková, Veselý 2013

Je zajímavé se podívat rovněž na to, jak se od roku 1998 změnilы výsledky jednotlivých věkových kohort. Graf 8 ukazuje, že se významně zhoršily výsledky nejmladší kohorty (16-24 let). Toto zhoršení však v ČR kopíruje zhoršení mezinárodní. V ČR došlo k propadu výsledků střední generace (pro toto zhoršení se nepodařilo nalézt uspokojivé vysvětlení) a k významnému zlepšení výsledků generace nejstarší.

Graf 8: Výsledky ve čtenářské gramotnosti podle věkových kohort, IALS, PIAAC



Zdroj: Straková, Veselý 2013

Třetí oblastí hodnocenou ve výzkumu PIAAC bylo řešení problémů v prostředí informačních technologií. V této oblasti řešili dospělí respondenti úlohy, se kterými se setkávají v běžném životě, s využitím tabulkového a textového procesoru a simulovaného webového prostředí. Čeští dospělí se umístili ve srovnání zemí OECD v průměru (na 12. místě z 19 zemí, které se tohoto volitelného šetření zúčastnily). V nejmladší věkové kohortě (16-24 let) již se však ČR posunula do nadprůměru na 7. místo. Tyto výsledky tedy také podporují poznatky z šetření v žákovské populaci, které ukazují na relativně vysokou zdatnost českých mladých lidí při práci s informačními technologiemi.

6.1.2 Nekognitivní výsledky

V rámci mezinárodních výzkumů je sledována řada výstupů vzdělávání, které nemají kognitivní povahu, například postoje ke škole, ke vzdělávání, sebepojetí, postoje k životnímu prostředí, občanské postoje a podobně. Těmto výstupům je zpravidla věnována při prezentaci poznatků z výzkumů daleko menší pozornost než kognitivním žebříčkům, a přitom mohou prozradit o našem vzdělávacím systému, případně o naší společnosti, leccos zajímavého. Mohou také poskytnout důležité podněty pro přemýšlení o prioritách dalšího rozvoje vzdělávacího systému. Zde jsou uvedeny vybrané indikátory, které šíří informací, které můžeme z mezinárodních výzkumů získat, pouze ilustrují.

Vztah ke škole

V rámci výzkumu PISA bylo opakovaně zkoumáno, jak se žáci cítí ve škole: jak jsou schopni navazovat pozitivní vztahy se svými spolužáky a učiteli, zda se cítí součástí kolektivu, zda je jim ve škole dobře (indikátor *sense of belonging to school* – zpravidla překládáno jako pocit sounáležitosti se školou). Východiskem pro zařazení těchto otázek byl předpoklad, že špatný

vztah ke škole může negativně ovlivňovat sebepojetí žáků, životní spokojenost a motivaci k učení. Indikátor byl konstruován z výroků týkajících se školy, se kterými žáci vyjadřovali míru souhlasu na čtyřstupňové škále. Uváděli, zda: a) se cítí jako outsideri, b) snadno si získávají přátele, c) mají pocit, že do školního prostředí patří, d) cítí se ve škole trapně a nevhodně, e) je ostatní žáci zřejmě mají rádi a f) se cítí osamělí⁶⁷. V roce 2012 pak žáci měli ještě za úkol ohodnotit svoji spokojenost a pohodu ve škole a reflektovat, zda školní prostředí odpovídá jejich představám. Indikátor, který z těchto otázek vznikl, poskytuje dobrou identifikaci toho, jak jsou vzdělávací systémy úspěšné v podpoře celkové spokojenosti (*well-being*) žáků. Vzhledem k tomu, že se jedná o subjektivní výpovědi žáků, je třeba při jejich mezinárodním porovnávání určité obezřetnosti, neboť máme řadu dokladů o tom, že výpovědi jsou ovlivněny kulturou. I když vezmeme v úvahu tato omezení a vycházíme z toho, že čeští žáci mohou být kritičtější než jejich vrstevníci z ostatních zemí OECD, není pravděpodobné, že by případná korekce zcela změnila výsledný obraz. Ten rozhodně nevychází pro ČR příliš příznivě: v roce 2003 dosahoval tento indikátor v ČR jednu z nejnižších hodnot mezi zeměmi OECD a do roku 2012 došlo ještě k jeho statisticky významnému zhoršení, takže byla jeho hodnota mezi zeměmi OECD v ČR nejnižší.

Dalším indikátorem zkoumajícím vztah žáků ke škole byl indikátor charakterizující užitečnost školního vzdělávání z pohledu žáků. Byl zkonstruován z otázek, které zjišťovaly, zda žáci věří: a) že je škola dobře nepřipravila na dospělý život, b) že škola byla ztrátou času, c) že je škola naučila důvěřovat svým rozhodnutím, d) že je naučila věci, které se jim mohou hodit v zaměstnání. V roce 2003 se hodnota indexu českých žáků pohybovala v průměru OECD. Do roku 2012 však došlo ke statisticky významnému zhoršení a ČR se ocitla mezi zeměmi, ve kterých žáci považují školní vzdělání za méně užitečné.

Postoje k matematice

Jedním z dlouhodobě sledovaných ukazatelů ve výzkumu PISA je vnitřní motivace žáků učit se matematiku. Indikátor charakterizující motivaci je založen na deklarovaném zájmu o matematiku a deklarovaném potěšení z řešení matematických úloh. Zde se ČR nachází dlouhodobě pod průměrem zemí OECD, i když mezi lety 2003 a 2012 došlo ke statisticky významnému zlepšení v tomto ukazateli. Instrumentální motivace, která spočívá v tom, že žáci věří, že jim učení se matematice pomůže uspět v dalším studiu a zaměstnání, byla v ČR v roce 2003 v průměru zemí OECD a do roku 2012 se statisticky významně zhoršila tak, že nyní je její hodnota ve srovnání se zeměmi OECD rovněž podprůměrná.

Další indikátor, který charakterizuje vztah žáků k matematice, se zaměřuje na to, zda žáci mají pocit, že jsou v matematice dobří: a) že jim matematika jde, b) že v ní mohou mít dobré známky, c) že se ji učí rychle, d) že je to jeden z jejich nejoblíbenějších předmětů a e) že rozumějí i těm nejobtížnějším úlohám (*self-concept* - sebepojetí). I v tomto indikátoru se ČR nachází pod průměrem zemí OECD. Zde dokonce patří dlouhodobě k zemím, kde je sebepojetí žáků v matematice nejnižší.

Dalším sledovaným indikátorem je obava z matematiky, která je dlouhodobě v průměru zemí OECD, přičemž mezi lety 2003 a 2012 se rovněž statisticky významně zvýšila.

⁶⁷ Otázky charakterizující indikátory jsou uváděny v takovém znění, v jakém figurovaly v žákovských dotaznících.

Postoje ke čtení

Jedním z ukazatelů sledovaných ve čtenářské oblasti je podíl žáků, kteří si čtou pro radost. Hodnota tohoto ukazatele od roku 2000 do roku 2009 významně poklesla. V roce 2009 uvedlo necelých 60 % českých patnáctiletých žáků, že si čtou pro radost, čímž se ČR zařadila mezi země OECD, ve kterých čtou žáci nejméně rádi. Stejný trend ukazuje souhrnný index oblíbenosti čtení, který zahrnuje ještě informace o čtení rozmanitých tiskovin a návštěvě knihoven. V tomto indexu je ČR po Belgii a Norsku zemí s třetí nejnižší hodnotou.

Postoje k přírodním vědám a životnímu prostředí

Ve srovnání s postoji k matematice a čtení se jeví postoje k přírodovědným disciplínám, které byly podrobně zjišťovány v rámci výzkumu PISA v roce 2006, pozitivněji. Sebepojetí českých žáků v oblasti přírodovědných předmětů je v průměru zemí OECD, stejně jako zájem o tyto předměty a jejich oblíbenost.

Pod průměrem zúčastněných zemí se však i zde nachází instrumentální motivace – žáci mají pocit, že jim přírodovědné předměty nebudou příliš užitečné. ČR dále vykazuje jeden z nejnižších podílů mladých lidí, kteří by rádi volili zaměstnání v přírodovědné oblasti.

Povědomí o jevech ohrožujících životní prostředí a zájem o ochranu životního prostředí je u českých žáků ve srovnání zemí OECD zhruba v průměru. Mírně podprůměrný je pak pocit zodpovědnosti za ochranu životního prostředí.

Občanské postoje a postoje k Evropské unii

Mezinárodní výzkumy CIVED a ICCS přinesly unikátní informace o občanských postojích českých žáků (např. Soukup 2010, Straková; Simonová 2013). Obecně je možno říci, že zatímco vědomosti českých žáků v oblasti fungování demokratické společnosti jsou v mezinárodním srovnání průměrné až nadprůměrné, ochota k angažovanosti a postoje ke znevýhodněným skupinám jsou v ČR horší než je průměr ostatních zemí. Toto tvrzení platí obecně pro žáky 8. ročníků a středoškoláky, kteří se zúčastnili šetření v roce 1999, i pro žáky 8. ročníků v roce 2009. Jediný rozdíl, který byl identifikován mezi lety 1999 a 2009, je v pocitu vlastenectví (hrdost na zemi, ve které žijeme). Zatímco v roce 1999 byl pocit vlastenectví v mezinárodním srovnání mezi žáky obou věkových kohort relativně vysoký, v roce 2009 již tomu tak není.

Šetření v roce 2009 poskytuje o postojích českých žáků velmi zajímavé informace, neboť poskytuje srovnání v rámci 24 evropských zemí nejen co se týče jejich obecných postojů, ale rovněž pokud jde o postoje k Evropské unii. Toto srovnání ukazuje ve srovnání s evropským průměrem slabší pocit evropské identity a horší postoj k vlastní zemi, k politickému sjednocení Evropské unie, k evropské měně i k rozšíření EU. V rámci institucí projeví čeští žáci nadprůměrně pozitivní postoj pouze k regionální správě, negativnější postoj než jejich evropští vrstevníci pak projeví k národní vládě, parlamentu, OSN i k Evropské komisi a Evropskému parlamentu. V evropském srovnání negativnější byl také jejich postoj k etnickým menšinám a přistěhovalcům, naopak nadprůměrně silné bylo jejich přání omezit přistěhovalectví (Kerr a kol. 2010).

Jako pozitivní můžeme vyhodnotit nadprůměrnou ochotu českých žáků učit se cizí jazyky. Čeští žáci naopak projeví relativně nízký zájem o politické dění a nízkou ochotu se jakkoli angažovat. V silnější míře než jejich evropští vrstevníci uvedli čeští žáci, že se nezajímají o národní a mezinárodní politiku, že nediskutují o sociálních a politických otázkách mimo školu

a že se neúčastní obecně prospěšných aktivit. Výrazně horší ve srovnání s žáky z ostatních evropských zemí je i jejich ochota se v budoucnu účastnit voleb.

6.2 Rozložení výsledků v populaci

O výkonu vzdělávacího systému nesvědčí pouze průměrné výsledky a jejich vývoj, kterými se zabývala kapitola 6.1, ale také rozložení výsledků v populaci. Za úspěšnější vzdělávací systémy jsou považovány ty, které jsou schopny zabránit vzniku velkých vzdělanostních nerovností, tedy situaci, kdy jsou mezi žáky velké rozdíly ve výsledcích, respektive kdy nějaké skupiny žáků dosahují velmi špatných výsledků. Proto se v této kapitole budeme zabývat tím, co víme o rozložení výsledků v jednotlivých věkových kohortách, o vazbě mezi výsledky a rodinným zázemím žáků a o rozdílech mezi vybranými skupinami žáků. Téma je zde pojednáno relativně podrobně, protože mezinárodní výzkumy představují pro ČR v tomto ohledu zcela výjimečný informační zdroj. Vzhledem k tomu, že zde dosud neprobíhala žádná jiná šetření na reprezentativních výběrech, nemohli bychom informaci o rozložení výsledků a souvislosti výsledků a rodinného zázemí získat žádným jiným způsobem⁶⁸. Mezinárodní výzkumy díky své pravidelné realizaci navíc umožňují se nejen poučit o aktuálním stavu, ale také sledovat dlouhodobý vývoj.

6.2.1 Rozložení výsledků a socioekonomické souvislosti

Jako ukazatel rozdílů mezi žáky v jednotlivých populacích můžeme využít širší rozdělení testových výsledků. Široké rozdělení svědčí o tom, že mezi žáky jsou velké rozdíly. Porovnáme-li širší rozdělení (hodnoty 5. a 95. percentilu ve výzkumech IEA, respektive hodnoty 10. a 90. percentilu ve výzkumu PISA), zjistíme, že je relativně vyšší v matematice a přírodovědných předmětech než ve čtenářské gramotnosti. V matematické a přírodovědné gramotnosti se pohybuje ve 4. ročníku ZŠ zhruba v průměru zúčastněných vyspělých zemí, u patnáctiletých žáků mírně nad průměrem zemí OECD. Ve čtenářské gramotnosti se pohybuje ve 4. ročníku pod průměrem vyspělých zemí, u patnáctiletých v blízkosti průměru. Šíře rozdělení jsou v čase víceméně stabilní.

V populaci dospělých se nicméně ve srovnání zemí, které se zúčastnily výzkumu PIAAC, ČR vyznačuje jedním z nejužších rozdělení mezi zúčastněnými zeměmi v matematické i čtenářské gramotnosti, přičemž šíře rozdělení je v čase rovněž stabilní. Důležité jsou vysoké hodnoty 5. percentilu, které ukazují, že se v ČR daří udržovat relativně vysokou úroveň čtenářských a matematických dovedností u osob, u kterých jsou tyto dovednosti nejnižší. Úzké rozdělení dále ukazuje vysokou homogenitu – mezi vědomostmi a dovednostmi dospělých v oblasti matematické a oblasti čtenářské gramotnosti nejsou mezi dospělými v České republice velké rozdíly⁶⁹.

⁶⁸ Soukromé organizace mají velké databáze dat z testovacích programů nabízených školám i jednotlivcům (nejznámější jsou pravděpodobně SCIO a Kalibro, i když organizací, které testování nabízejí, je více). Tato data jsou však získávána od škol, které projevíly zájem se do programů zapojit, výběry tedy nemohou být považovány za reprezentativní. Testování žáků v povinném vzdělávání prováděná státem byla v minulosti omezena pouze na některé regiony, nebo nezahrnovala informace o rodinném zázemí. Výstupy tedy neumožňovaly získat informace o souvislosti výsledků a rodinného zázemí, případně dalších proměnných. Již zmiňovaný výzkum CLoSE je výjimečný v tom, že je prováděn na reprezentativním výběru a kromě výsledků vzdělávání a jejich vývoje podrobně mapuje i rodinné zázemí.

⁶⁹ Skutečnost, že šíře rozdělení v žákovské populaci je v mezinárodním srovnání průměrná a šíře rozdělení v dospělé populaci podprůměrná, může být způsobena tím, že ČR má vysokou účast v základním i středním vzdělávání. K širší rozdělení v dospělé populaci tedy již nepřispívají jedinci, kteří se při měření v žákovské populaci nacházejí mimo systém. Srovnání v žákovské a dospělé populaci také neprobíhalo na identické skupině zemí.

Síla vazby výsledku a socioekonomického statusu

Tabulka 25 ukazuje pro země OECD, které se zúčastnily výzkumu PISA v letech 2003 a 2012, souvislost mezi výsledkem v matematickém testu a indexem *escs*, který charakterizuje rodinné zázemí, v obou sběrech dat. Pro každý rok je v levém sloupci uveden celkový přírůstek ve výsledku v testu, který odpovídá jednotkovému navýšení hodnoty indexu *escs* na úrovni žáka. V dalších dvou sloupečcích je pak přírůstek „rozložen“ na žákovskou a školní složku (hodnota vychází z regresních koeficientů dvouúrovňového modelu). Země jsou v tabulce seřazeny sestupně podle hodnoty celkového individuálního přírůstku v roce 2012.

Z tabulky je zřejmé, že ve srovnání zemí OECD je síla závislosti mezi výsledkem v testu z matematiky a rodinným zázemím v České republice jednou z nejvyšších. Velmi vysoká je zejména vazba na úrovni školy. To znamená, že v ČR se od sebe jednotlivé školy velmi odlišují z hlediska složení žáků, které vzdělávají, a z hlediska jejich výsledků. Relativně vysoké hodnoty rozdílů mezi školami v ČR, které byly identifikovány ve výzkumu PISA, jsou jistě způsobeny i tím, že do výzkumu PISA jsou zapojeny školy na úrovni základního i středního vzdělávání, tedy k rozdílům mezi školami se přidávají rozdíly mezi ročníky na dvou různých úrovních vzdělávání. Tak tomu není ve všech zemích – v řadě zemí již jsou všichni patnáctiletí žáci na středních školách a zpravidla také jeden z ročníků bývá zastoupen výrazně více než ostatní ročníky. Tato skutečnost tedy může relativní nehomogenitu mírně zvyšovat, v následujícím oddíle ale ukazujeme, že v ČR jsou vysoké rozdíly i mezi základními školami. Zkreslení tudíž není tak velké, aby změnilo celkový obraz. Důležité je rovněž zjištění, že se rozdíly mezi školami dále zvyšují (nárůst v závislosti výsledku na souhrnném indexu mezi roky 2003 a 2012 je statisticky významný). Porovnáme-li stejné hodnoty ve čtenářské gramotnosti mezi lety 2000 a 2009, kdy byla čtenářská gramotnost hlavní testovanou oblastí, zjistíme, že i z hlediska čtenářské gramotnosti patří ČR k zemím s nejsilnější závislostí výsledků na rodinném zázemí na úrovni žáka celkově i na úrovni školy. Zde se však rozdíly mezi školami statisticky významně nezvýšily⁷⁰.

Analogické porovnání souvislosti výsledků a rodinného zázemí bohužel nemůžeme provést na prvním stupni, neboť nemáme k dispozici informaci o rodinném zázemí žáků. Jediný údaj, kterým je možno u žáků 4. ročníku aproximovat rodinné zázemí, je počet knih v domácnosti. Tato proměnná představuje pro mezinárodní srovnání přece jen poněkud příliš hrubý ukazatel.

Rozdíly mezi školami

Jak bylo uvedeno výše, o velkých rozdílech mezi školami, co se týče jejich výsledků, vypovídají i analýzy národních dat pořízených na prvním a druhém stupni povinného vzdělávání ve výzkumech IEA i ve výzkumu PISA. V tabulce 26 jsou uvedeny koeficienty vnitrotřídní korelace⁷¹ pro výsledky v matematickém testu pro 4. ročník (testovaný v letech 1995, 2007 a 2011) a 8. ročník (testovaný v letech 1995, 1999 a 2011). Výsledky napovídají, že ve 4. ročnících základní školy dosud, i přes rostoucí diferenciaci primárního vzdělávání, není pozorován nárůst rozdílů ve výsledcích žáků jednotlivých škol. V 8. ročníku jsme do roku 2007 pozorovali mírný nárůst těchto rozdílů.

⁷⁰ <http://www.oecd.org/pisa/pisaproducts/pisa2009resultsovercomingsocialbackgroundinequityinlearningopportunitiesandoutcomesvolumeii.htm>

⁷¹ Koeficient vnitrotřídní korelace je podíl rozptylu výsledků mezi školami ku součtu rozptylu výsledků mezi školami a rozptylu výsledků mezi žáky uvnitř těchto škol. Koeficienty vnitrotřídní korelace uvedené v této kapitole byly vypočteny z nulového modelu odhadnutého v software HLM 6, který umožňuje rozložit rozptyl na úroveň školy a úroveň žáka. Software zároveň umožňuje řádně pracovat s plausibilními hodnotami. Data byla vážena pomocí žákovských vah.

Tabulka 25: Souvislost výsledku v matematickém testu a rodinného zázemí žáků, PISA 2003 a 2012
(v závorce jsou uvedeny standardní chyby)

	PISA 2003				PISA 2012				Změna 2012–2003			
	Celkový přírůstek na úrovni žáka (s.e.)	Přírůstek na úrovni žáka (s.e.)	Přírůstek na úrovni školy (s.e.)		Celkový přírůstek na úrovni žáka (s.e.)	Přírůstek na úrovni žáka (s.e.)	Přírůstek na úrovni školy (s.e.)		Změna v celkovém přírůstku na úrovni žáka (s.e.)	Změna v přírůstku na úrovni žáka (s.e.)	Změna v přírůstku na úrovni školy (s.e.)	
Slovenská republika	48 (2,5)	20 (1,7)	79 (4,5)		54 (2,9)	21 (2,1)	86 (6,4)		6 (3,8)	0 (2,7)	7 (7,8)	
Nový Zéland	44 (1,7)	35 (1,8)	53 (5,8)		52 (1,9)	36 (2,3)	66 (6,8)		8 (2,5)	1 (2,9)	13 (8,9)	
Česká republika	46 (2,0)	17 (1,6)	94 (5,3)		51 (2,7)	14 (1,7)	127 (6,5)		5 (3,4)	-3 (2,4)	33 (8,4)	
Belgie	51 (1,9)	22 (1,3)	88 (5,8)		49 (1,7)	19 (1,5)	102 (6,0)		-2 (2,6)	-4 (2,0)	12 (8,4)	
Maďarsko	50 (2,1)	12 (1,6)	86 (3,6)		47 (2,8)	6 (1,4)	98 (4,9)		-3 (3,5)	-5 (2,1)	12 (6,1)	
Rakousko	42 (2,2)	9 (1,4)	100 (5,8)		43 (2,2)	15 (1,6)	85 (5,9)		2 (3,1)	7 (2,1)	-15 (8,3)	
Německo	44 (1,6)	15 (1,5)	82 (4,6)		43 (2,0)	11 (1,4)	103 (5,6)		-1 (2,5)	-4 (2,1)	21 (7,2)	
Korea	38 (2,8)	12 (1,5)	78 (6,9)		42 (3,3)	14 (2,0)	114 (10,2)		5 (4,3)	2 (2,5)	36 (12,3)	
Austrálie	40 (1,8)	26 (1,1)	54 (4,1)		42 (1,3)	25 (1,3)	64 (4,1)		2 (2,2)	-1 (1,7)	10 (5,9)	
Polsko	40 (1,6)	33 (1,7)	26 (4,7)		41 (2,4)	32 (1,8)	36 (7,4)		1 (2,9)	-2 (2,5)	10 (8,8)	
Japonsko	43 (4,5)	5 (1,8)	121 (11,4)		41 (3,9)	4 (1,7)	150 (8,2)		-2 (6,0)	-1 (2,5)	29 (14,0)	
Nizozemsko	39 (2,2)	12 (1,4)	105 (5,5)		40 (3,1)	9 (1,6)	147 (10,9)		0 (3,8)	-3 (2,2)	41 (12,2)	
Dánsko	39 (1,8)	32 (1,7)	28 (5,0)		39 (1,7)	31 (1,7)	38 (4,2)		1 (2,5)	-1 (2,4)	10 (6,5)	
OECD (průměr 2003)	39 (0,4)	20 (0,3)	59 (1,1)		39 (0,4)	20 (0,3)	71 (1,2)		0 (0,6)	-0,8 (0,5)	12,0 (1,6)	
Irsko	36 (1,7)	27 (1,5)	40 (5,1)		38 (1,8)	26 (1,7)	52 (4,7)		2 (2,5)	-1 (2,3)	13 (6,9)	
Švýcarsko	41 (1,8)	23 (1,8)	62 (7,2)		38 (1,8)	25 (1,5)	66 (8,9)		-3 (2,6)	2 (2,4)	4 (11,4)	

Lucembursko	35	(1,2)	19	(4,4)	72	(7,2)	37	(1,2)	20	(2,7)	68	(4,9)	2	(1,7)	1	(5,2)	-4	(8,7)
Švédsko	36	(1,9)	32	(1,8)	28	(6,9)	36	(1,9)	28	(2,0)	41	(7,3)	-1	(2,7)	-4	(2,6)	12	(10,0)
Spojené státy	42	(1,4)	29	(1,5)	46	(4,3)	35	(1,7)	24	(1,9)	41	(5,4)	-7	(2,2)	-6	(2,4)	-5	(6,9)
Portugalsko	28	(1,2)	18	(1,1)	37	(3,9)	35	(1,6)	23	(1,4)	33	(4,0)	7	(2,0)	5	(1,8)	-4	(5,5)
Řecko	36	(2,0)	17	(1,7)	58	(6,0)	34	(1,8)	18	(1,5)	55	(5,1)	-2	(2,8)	0	(2,3)	-3	(7,9)
Španělsko	27	(1,4)	18	(1,3)	31	(3,5)	34	(1,1)	27	(1,0)	26	(3,2)	6	(1,8)	9	(1,6)	-5	(4,7)
Finsko	28	(1,4)	28	(1,3)	0	(5,6)	33	(1,8)	29	(1,5)	22	(6,9)	5	(2,3)	2	(2,0)	22	(8,9)
Norsko	41	(1,8)	38	(2,0)	20	(7,0)	32	(2,4)	27	(2,2)	49	(7,8)	-8	(3,1)	-11	(3,0)	29	(10,5)
Turecko	50	(5,0)	11	(1,5)	87	(7,0)	32	(2,4)	6	(1,0)	83	(7,4)	-18	(5,6)	-5	(1,8)	-4	(10,2)
Kanada	30	(1,3)	23	(1,0)	31	(3,7)	31	(1,2)	23	(1,2)	41	(5,0)	1	(1,8)	0	(1,6)	9	(6,2)
Island	26	(1,6)	26	(1,7)	8	(5,7)	31	(2,1)	25	(2,9)	45	(8,3)	5	(2,6)	-1	(3,4)	36	(10,1)
Itálie	31	(1,8)	8	(1,0)	69	(5,5)	30	(1,2)	7	(0,7)	83	(4,1)	-1	(2,2)	-1	(1,2)	14	(6,9)
Mexico	30	(1,9)	7	(0,9)	53	(3,3)	19	(0,8)	5	(0,5)	29	(1,4)	-11	(2,0)	-2	(1,1)	-24	(3,6)

Zdroj: <http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-ii.htm>

Tabulka 26: Vývoj koeficientu vnitrotržní korelace (vyjádřeném v %) pro matematický test, TIMSS

	1995	1999	2007	2011
4. ročník	21,0		16,3	20,7
8. ročník	27,4	28,7	30,4	

Zdroj dat: národní datové soubory TIMSS 1995, 1999, 2007, 2011

Informaci o dalším vývoji v posledních ročnících ZŠ můžeme získat z tabulky 27, kde jsou uvedeny hodnoty koeficientu vnitrotržní korelace ve všech třech testovaných oblastech pro žáky 9. ročníku, a to pro všechny sběry dat výzkumu PISA od roku 2000 do roku 2012. Zde se dopouštíme určité nepřesnosti, neboť výběr ve výzkumu PISA je reprezentativní za populaci patnáctiletých, nikoli za 9. ročník. Žáků 9. ročníku je ve vzorku nicméně polovina a porovnáme-li parametry skupiny žáků 9. ročníku vybraných do výzkumu PISA s parametry celkové populace žáků 9. ročníku, zjistíme, že si dobře odpovídají. O velikosti odchylky způsobené tím, že provádíme výpočet na takto vybrané skupině, jsme se mohli poučit například ve výzkumu PISA 2006, kde byl kromě reprezentativního výběru patnáctiletých proveden i reprezentativní výběr žáků 9. ročníků. Hodnoty koeficientu vnitrotržní korelace vyjádřené v procentech a vypočtené na tomto výběru byly 33,4 % pro matematický test, 32,5 % pro čtenářský test a 31,2 % pro přírodovědný test. Porovnání s hodnotami uvedenými v tabulce 27, které byly získány u skupiny žáků 9. ročníku z výběru patnáctiletých, ukazuje, že hodnoty vypočtené na datech z výběru patnáctiletých jsou v některých případech mírně nadsazené. To ukazuje i porovnání s hodnotami koeficientu v 8. ročníku v roce 2007, který je o 4 % nižší než hodnota odhadnutá v PISA pro rok 2006. I při zohlednění tohoto pravděpodobného nadsazení je však zřejmé, že rozdíly mezi školami v posledních ročnících povinného vzdělávání jsou relativně vysoké a nadále se významně zvyšují. To znamená, že čím dál tím víc záleží na tom, v jaké škole dítě absolvuje povinnou školní docházku.

Tabulka 27: Koeficient vnitrotržní korelace (vyjádřený v %), výzkum PISA

	2000	2003	2006	2009	2012
matematika	28,8	33,6	34,3	42,4	49,7
čtení	28,1	27,9	34,1	39,1	44,8
přírodověda	21,0	28,1	31,2	47,7	46,8

Zdroj dat: národní datové soubory PISA 2000 - 2012

Vzdělanostní přechody

Již bylo uvedeno, že data z mezinárodních výzkumů představují zcela unikátní zdroj informací o vzdělanostních nerovnostech v České republice a o jejich vývoji. Výzkumy ukázaly, že v ČR je více než v řadě jiných vyspělých zemí vzdělání mladého člověka podmíněno tím, z jaké rodiny pochází a jakou školu navštěvuje. Analýzy dat z mezinárodních výzkumů doplnily informaci o relativně silné závislosti dosaženého vzdělání na rodinném zázemí, kterou přinesly ve 2. polovině minulého století sociologické studie, a způsobily, že se čeští výzkumníci začali touto

problematikou intenzivně zabývat se zřetelem k vysoké diferenciaci českého vzdělávacího systému.

Jak pomohly mezinárodní výzkumy zmapovat nerovnosti na přechodu do víceletého gymnázia, ukazuje výzkum CLoSE prostřednictvím longitudinální komponenty navazující na výzkum TIMSS a PIRLS 2011. Sledoval žáky, kteří byli testováni v roce 2011, na přechodu na druhý stupeň ZŠ, respektive do víceletých gymnázií. Analýza ukázala, že i při kontrole školních výsledků a studijních předpokladů (aproximovaných testem TIMSS) mají žáci pocházející ze vzdělaných rodin vyšší šanci, že budou studovat na víceletém gymnáziu. Je to způsobeno i tím, že v ČR je rozhodnutí, zda bude dítě skládat přijímací zkoušky na víceleté gymnázium, plně v kompetenci rodiny. Logistická regrese, jejíž výsledky jsou uvedeny v tabulce 28, ukazuje, na čem záviselo, zda se žák ucházel o studium na víceletém gymnáziu. Vyplývá z ní, že největší vliv na to, zda žák aspiroval ke studiu na víceletém gymnáziu, měl jeho školní prospěch; zhruba poloviční vliv pak mělo vzdělání jeho rodičů a výsledek v matematickém testu TIMSS⁷².

Tabulka 28: Aspirace na studium na víceletém gymnáziu, CLoSE 2012

	B	s.e.	sig.
průměrný prospěch	0,501	0,053	0,000
výsledek v matematickém testu	0,197	0,035	0,000
vzdělání rodičů	0,227	0,034	0,000
R-SQUARE			0,531

Zdroj dat: TIMSS 2011, CLoSE 2012

Tabulka 29 ukazuje, na čem závisí, zda žák bude na víceletém gymnáziu studovat. Zde poněkud oslabil význam školního prospěchu a mírně posílil význam studijních předpokladů (výsledku v matematickém testu) a vzdělání rodičů. Šetření ukázalo, že pro úspěch v přijímacích zkouškách je nezbytná pečlivá domácí příprava a že žáci se nejčastěji připravují se svými rodinnými příslušníky nebo navštěvují placené kurzy. Tedy vzdělanější rodiče nejen, že žáky k přijímacím zkouškám přihlásí, ale také je na ně buď sami připraví, nebo jim mohou zaplatit přípravné kurzy. Žáci méně vzdělaných a méně motivovaných rodičů jsou v procesu přijímání znevýhodněni hned dvakrát: rodiče je ke studiu nepřihlásí, a i kdyby je přihlásili, nemohou jim zajistit přípravu (Straková; Greger 2013).

Závislost na rodinném zázemí pozorujeme rovněž na přechodu do středního vzdělání. Tabulka 30 ukazuje na datech z výzkumů IALS a PIAAC, jaké faktory ovlivňují studium v maturitním oboru a jak se význam těchto faktorů změnil mezi lety 1998 a 2012. Je z ní zřejmé, že dívky a mladí lidé z rodin s vyšším socioekonomickým statusem mají vyšší šanci na studium v maturitním oboru i při kontrole studijních předpokladů (zde výsledek v testu čtenářské gramotnosti). Šance na přijetí byla vyšší v roce 2012 než v roce 1998. Analýza ovšem zároveň ukázala pozitivní interakci roku a pohlaví, a roku a socioekonomického statusu. To znamená, že znevýhodnění

⁷² Proměnné byly standardizovány, jejich vliv je tedy možno přímo porovnávat. Analýza byla provedena v software M-plus a využitím replikačních vah. Průměrný prospěch byl průměrem hlavních známek na pololetním vysvědčení v 5. ročníku, vzdělání rodičů bylo počítáno jakou součet kategorií na škále 1- základní, 2 - střední bez maturity, 3- střední s maturitou, 4-vysokoškolské, výsledek v matematickém testu byl charakterizován první plausibilní hodnotou.

chlapců a dětí s nízkým socioekonomickým statusem na přechodu do maturitního oboru se v čase zvyšuje.

Tabulka 29: Šance na studium na víceletém gymnáziu, CLoSE 2012

	B	s.e.	sig.
průměrný prospěch	0,466	0,066	0,000
výsledek v matematickém testu	0,222	0,041	0,000
vzdělání rodičů	0,262	0,045	0,000
R-SQUARE			0,552

Zdroj dat: TIMSS 2011, CLoSE 2012

Tabulka 30: Šance na studium středoškolského oboru s maturitou, IALS a PIAAC

	model 1			model 2		
	B	s.e.	sig.	B	s.e.	sig.
výsledek v testu čtenářské gramotnosti	0,986	0,002	0,000	0,983	0,002	0,000
pohlaví (ž)	0,321	0,001	0,000	0,328	0,001	0,000
SES	0,819	0,002	0,000	0,843	0,002	0,000
rok (2012)	0,558	0,001	0,000	0,575	0,001	0,000
rok * SES				0,095	0,002	0,000
rok * pohlaví (ž)				-0,068	0,002	0,000
R-SQUARE			0,399			0,402

Zdroj dat: IALS 1998, PIAAC 2012

Mezinárodní výzkumy zároveň umožňují pomocí údajů z žakovských a ředitelských dotazníků porovnat podmínky ke vzdělávání v různých programech a tím ukázat, že žáci, kteří jsou svým rodinným zázemím znevýhodněni v přechodu do výběrových programů, jsou dále znevýhodňováni při studiu na nevýběrových programech (ZŠ, nematuritní obory), neboť tyto obory jim poskytují horší vzdělávací podmínky. Kromě méně motivovaných spolužáků je to například horší vybavení, méně kvalifikovaní učitelé a vyšší četnost kázeňských problémů (viz např. Straková 2011).

6.2.2 Poznatky týkající se rozdílů mezi skupinami

Ve většině vzdělávacích systémů vyspělých zemí dochází k monitorování výsledků vzdělávání a vzdělávacího pokroku specifických skupin, aby se zamezilo tomu, že některé skupiny v systému systematicky neprospívají, aniž by tato skutečnost byla reflektována a byla přijata nápravná opatření. Hlavními sledovanými skupinami jsou zpravidla dívky a chlapci, etnické, kulturní a náboženské menšiny, žadatelé o azyl, uprchlíci, žáci s potřebou podpory v jazyce, v němž se vyučuje, žáci se zdravotním postižením a znevýhodněním a žáci se specifickými

vzdělávacími potřebami. Z mezinárodních výzkumů se můžeme poučit pouze o rozdílech mezi chlapci a dívkami, respektive muži a ženami, dále o výsledcích žáků cizinců. Určité omezené poznatky máme v ČR rovněž o vzdělávání romských dětí. I když tyto informace jsou v případě posledních dvou skupin pouze přibližné a bylo by je třeba ověřit dalšími cílenými výzkumy, jsou důležité, neboť o kognitivních výsledcích těchto skupin máme jen velmi limitované poznatky z jiných zdrojů⁷³.

Rozdíly podle pohlaví

Ve věkové kategorii žáků 4. ročníků dosahují čeští chlapci opakovaně lepších výsledků než dívky v matematice a přírodovědných předmětech, přičemž rozdíly mezi oběma pohlavími jsou v ČR mezi zúčastněnými zeměmi jedny z nejvyšších (Tomášek a kol. 2012). Ve čtenářské gramotnosti dosahují naopak lepších výsledků dívky, zde však patří rozdíly v ČR mezi zúčastněnými zeměmi k nejnižším (Kramplová a kol. 2012).

Mezi patnáctiletými žáky zkoumanými v rámci výzkumu PISA pozorujeme dlouhodobě statisticky významně lepší výsledky dívek ve čtenářské gramotnosti a naopak statisticky významně lepší výsledky chlapců v matematické gramotnosti. Velikost rozdílů je nicméně ve srovnání zemí OECD spíše nižší. V přírodovědné gramotnosti jsou rozdíly nevýznamné. Velikost rozdílů je v čase značně stabilní (OECD 2013b). Statisticky významné rozdíly mezi chlapci a děvčaty nebyly zaznamenány ani v řešení problémů v roce 2012.

Ve vztahu k těmto (velmi konsistentním) poznatkům je zajímavé zjištění, že dívky mají lepší průměr známek na vysvědčení v českém jazyce i v matematice (v rámci výzkumu PISA jsou zjišťovány známky z hlavních předmětů na posledním vysvědčení).

Dále je zajímavé, že přestože v žákovské populaci dosahují dívky lepších výsledků v testech čtenářské gramotnosti než chlapci, v populaci dospělých dosahují muži statisticky lepších výsledků než ženy jak ve čtenářské, tak v matematické gramotnosti. Rozdíly mezi muži a ženami zůstávají v čase stabilní. V řešení problémů v prostředí informačních technologií nebyly rozdíly mezi oběma pohlavími identifikovány.

Výsledky žáků cizinců

V mezinárodních zprávách PISA jsou od počátku pravidelně prezentovány výsledky pro první a druhé generace přistěhovalců v zemích, ve kterých tvoří přistěhovalci více než 3 %. Jako přistěhovalci jsou zkoumány děti, které se narodily v jiné zemi, nebo jejichž rodiče se narodili v jiné zemi než v zemi testování. ČR byla do tohoto srovnání poprvé zařazena v roce 2009. Z poznatků v letech 2009 a 2012 vyplývá, že v ČR se průměrné výsledky patnáctiletých žáků přistěhovalců statisticky významně neliší od výsledků žáků s českým mateřským jazykem.

⁷³ V roce 2015 byla zveřejněna výzkumná zpráva z projektu ESF Sociologický monitoring vzdělanostních vstupů a výstupů dětí a žáků, včetně dětí a žáků se speciálními vzdělávacími potřebami v České republice s názvem Sociologická analýza zaměřená na porovnání dosažených výsledků žáků se speciálními vzdělávacími potřebami v různých typech škol. Projekt byl realizován Slezskou univerzitou v Opavě. Jedním z cílů výzkumu bylo monitorovat pokrok žáků s mentálním postižením, s hraničním intelektem a žáků intaktních ve školách hlavního vzdělávacího proudu a školách praktických a speciálních. I když by projekt mohl potenciálně přinést informace o výsledcích vzdělávání romských žáků, jeho součástí nebyla identifikace těchto žáků. Projekt dále zkoumal výsledky žáků pouze v průběhu jednoho ročníku školní docházky (3. ročník ZŠ) a testy navíc nebyly koncipovány tak, aby umožňovaly přímé měření pokroku jednotlivců ve sledovaných vzdělávacích oblastech.

Velikost rozdílu mezi českými žáky a žáky cizinci se příliš nezmění ani po zohlednění socioekonomického statusu. Srovnání bylo provedeno ve čtenářské a matematické gramotnosti (OECD 2010a, OECD 2013b).

Sonda provedená v ČR v roce 2009 v rámci výzkumu PISA, kdy byl proveden nadvýběr 200 žáků cizinců, napovídala, že se od sebe neodlišují ani výsledky jednotlivých skupin přistěhovalců podle země/regionu původu (byli srovnáváni přistěhovalci ze zemí bývalého Sovětského svazu, z asijských zemí a z afrických zemí). Tento poznatek by však bylo třeba ověřit na vyšším počtu respondentů.

Heath a Kilpi-Jakonen (2012) srovnávali v PISA 2009 výsledky přistěhovalců podle věku přistěhování. Definovali skupinu žáků, kteří se přistěhovali v předškolním věku, a tedy absolvovali v dané zemi celé povinné vzdělávání (časné přistěhování), a těch, kteří absolvovali v dané zemi maximálně 3 roky (pozdní přistěhování). Prostřední skupinu tvořily děti, které navštěvovaly školu v dané zemi více než 3 roky, ale nastoupily do ní později než v 1. ročníku. V ČR byly rozdíly mezi prostřední skupinou a časným přistěhováním statisticky nevýznamné. Přistěhovalci 2. generace však měli statisticky lepší výsledky než časní přistěhovalci. Výsledky majority se od výsledků časných přistěhovalců neliší (Linhartová; Horáčková 2015).

Srovnání provedená v rámci výzkumu PISA však vypovídají o situaci dětí cizinců v České republice pouze omezeným způsobem. Zařazení či nezařazení žáků do výzkumu bylo v případě žáků se speciálními vzdělávacími potřebami, mezi které patří i nedostatečná znalost jazyka testu, v kompetenci učitelů. Řada žáků tedy nebyla zařazena z důvodů nedostatečné znalosti českého jazyka. Organizátoři výzkumu mají k dispozici informace o tom, kolik žáků bylo z tohoto důvodu vyloučeno a v mnoha případech vědí i o tom, jaký mají tito žáci mateřský jazyk, respektive z jaké země pocházejí. Bylo by vhodné tyto informace podrobit podrobnějšímu zkoumání, neboť představují unikátní informaci o žácích-cizincích v českém vzdělávacím systému a jejich elementární jazykové zdatnosti (podle hodnocení jejich učitelů).

Výsledky romských žáků

V PISA 2009 bylo spolu s nadvýběrem dětí cizinců provedeno i speciální šetření v 18 školách v blízkosti vyloučených lokalit, které byly identifikovány společností GAC. Dětem na těchto školách byly administrovány výzkumné nástroje PISA s cílem porovnat výsledky romských a majoritních žáků ve školách hlavního vzdělávacího proudu s převahou majoritních žáků, ve školách hlavního vzdělávacího proudu s převahou romských žáků a ve školách praktických. Žáci byli pro účely výzkumu identifikováni prostřednictvím otázky na jazyk komunikace v rodině (příčemž byli dotazováni nejen na svůj komunikační jazyk, ale i na jazyk komunikace rodičů a prarodičů). Kontrolní identifikace byla provedena učitelem.

Šetření ukázalo rozdíly ve výsledcích romských a neromských žáků ve prospěch žáků neromských ve všech typech škol (tedy i ve školách praktických) i po zohlednění rodinného zázemí. Výzkumný vzorek byl však natolik malý, že není možno zjištění zobecňovat. Pro potvrzení získaných výsledků by bylo potřeba provést další šetření na reprezentativním výběru romských žáků (Straková; Tomášek 2013). Toto šetření u nás dosud nebylo provedeno, přestože výsledkům vzdělávání romských žáků v ČR je věnována velká pozornost ze strany zahraničních organizací a že je do něj investováno velké množství finančních prostředků z Evropského sociálního fondu. Absence dat znemožňuje rigorózně posoudit efektivitu vynakládaných prostředků.

6.3 Další poznatky z mezinárodních výzkumů výsledků vzdělávání žáků

Jak bylo uvedeno v kapitole věnované doprovodným dotazníkům, obsahují tyto dotazníky velké množství zajímavých údajů, které však zůstávají často výzkumně nevytěženy. Z žakovských dotazníků můžeme získat kromě údajů o rodinném zázemí žáků (které je podrobně zkoumáno zejména ve výzkumu PISA) rovněž informace o nekognitivních výstupech vzdělávání, kterým byla věnována pozornost v oddíle 6.1, a informace o průběhu výuky z pohledu žáků. Informace o výuce byly dosud analyzovány zpravidla se zřetelem k četnosti různých výukových aktivit. Zde se největší pozornosti dostávalo přírodovědným předmětům, ve kterých se opakovaně ukazovalo, že čeští žáci ve srovnání s žáky z jiných zemí málo samostatně experimentují, mají malou příležitost navrhnout experimenty, formulovat a ověřovat hypotézy, tedy si osvojit základy vědecké práce (viz např. Straková 2010b). Opakovaně byla sledována rovněž nízká četnost skupinové práce a individualizace výuky. Tato srovnání byla ovšem čistě deskriptivní a nezahrnovala kontext, tedy nevztahovala se k pedagogickým tradicím jednotlivých zemí, ani nebyla opřena o žádný komplexní koncept kvalitní výuky.

Na základě kritik mezinárodních výzkumů zmiňovaných v kapitole 5, které na tento nedostatek upozorňovaly, roste v posledních letech v mezinárodních výzkumech důraz na teoretickou bázi doprovodných dotazníků tak, aby nebyly podkladem pro jednoduché, vzájemně nesouvisející deskriptivní poznatky, které nejsou ukotveny v žádném kontextu, ale aby naopak sloužily pro komplexnější analýzy souvislostí založené na teoretických modelech. Nejhojněji se nyní uplatňuje model kvalitní výuky, která je charakterizována dobrým řízením třídy, podpůrným prostředím, srozumitelností a kognitivní náročností (Klieme; Paul; Reusser 2009, Klieme 2013). Tento model je od roku 2012 využíván ve výzkumu PISA jako vodítko pro konstrukci žakovských dotazníků a ve výzkumu TIMSS jsou na jeho základě konstruována v některých zemích i národní přídatná šetření.

Potenciálně cenným a málo využívaným zdrojem informací jsou ředitelské dotazníky a učitelské dotazníky administrované v rámci výzkumů IEA. Ředitelské dotazníky mohou být využity v národních analýzách a poskytnout užitečné deskriptivní informace zejména o tom, jakým obtížím čelí ředitelé při své práci a jaká je struktura jejich pracovních povinností. Určitou nevýhodou je relativně malý počet respondentů (150 až 250)⁷⁴.

Ve vztahu k mezinárodním analýzám školních dat je naopak třeba určité obezřetnosti. Otázky jsou formulovány tak, aby mohly být použity v celé řadě velmi různorodých systémů, a tím pádem hrozí určité nebezpečí, že budou v jednotlivých systémech různé pochopeny. Ve výzkumu PISA pak často mezinárodní analýzy nedávají pro Českou republiku dobrý smysl, neboť v ČR, jak již bylo uvedeno, jsou do výběru zahrnuty základní a střední školy. Proměnné týkající se organizace školy, výběrovosti školy, jejího zaměření, přijímacího řízení a podobně poskytují v případě těchto dvou stupňů škol odlišnou informaci a jejich společná analýza není příliš užitečná. Již byla zmiňována skutečnost, že ve zprávě z výzkumu PISA 2012 byla uveřejněna informace, že ČR je spolu s Estonskem jedinou zemí, ve které konkurence mezi školami přispívá ke kvalitě výuky. Když byl do analýzy zahrnut typ školy, souvislost mezi výsledkem a konkurencí vymizela.

⁷⁴ V mezinárodních výkumech jsou poznatky z ředitelských a učitelských dotazníků vždy prezentovány prostřednictvím žáků, které ovlivňují. Byly tedy převažovány žakovskými vahami a prezentovány jako „poznatky ředitelů/učitelů daného procenta žáků v dané zemi“.

Takových příkladů by bylo možno nalézt více. Školní analýzy ve výzkumu PISA je tedy třeba přijímat spíše s nedůvěrou (Dvořák; Straková 2016).

Učitelé dotazníky mohou poskytnout dílčí informace o výukových aktivitách z pohledu učitelů, o zodpovědnostech učitelů, o obtížích, se kterými se setkávají ve své práci, o zpětné vazbě, které se jim dostává od vedení škol, o dalším vzdělávání. Zde se například ukázalo, že si čeští učitelé ve srovnání s učiteli z jiných zemí velmi málo vzájemně hospitují, nebo že v dalším vzdělávání upřednostňují kurzy zaměřené na informační technologie před pedagogickými tématy, jako je individuální podpora či formativní hodnocení. I v případě učitelů však platí námitka týkající se relativně nízkého množství respondentů.

Velmi důležité informace, které poskytují učitelé dotazníky, se týkají postojů učitelů a vztahů učitelů k žákům. Zde je naopak možnost mezinárodního srovnání velkou výhodou, neboť bez tohoto srovnání je obtížné naměřené hodnoty interpretovat. Subjektivní výpovědi však samozřejmě hrozí i zde určitým kulturním zkreslením.

Ve výzkumu TIMSS 2011 byly postoje učitelů analyzovány s oporou o koncept akademického optimismu (Hoy a kol. 2000, McGuigan; Hoy, 2006, Woolfolk Hoy; Hoy; Kurz 2008), na jehož základě byl konstruován indikátor SEAS (*School Emphasis on Academic Success* – důraz školy na akademický úspěch). Akademický optimismus zdůrazňuje potenciál školy překonat znevýhodnění způsobené socioekonomickými i dalšími faktory, které negativně ovlivňují školní výsledky žáků. Sestává ze tří složek, kterými jsou subjektivně vnímané vlastní schopnosti (*sense of efficacy*), důvěra v žáky a v jejich rodiče a cílené vytváření vstřícného a podpůrného prostředí, které klade na žáky vysoké akademické nároky.

Mezinárodní analýza (Martin; Mullis 2013) zkoumala vztah mezi důrazem školy na akademický úspěch a výsledkem žáků. V této analýze byla ČR jednou z mála zemí, kde vazba mezi těmito proměnnými nebyla identifikována. Hodnota indexu však zároveň patřila v České republice mezi zúčastněnými zeměmi k nejnižším. Odpovědi učitelů ukázaly, že čeští učitelé mají ve srovnání s učiteli v ostatních zemích obecně nižší sebedůvěru, přičemž relativně nejnižší je jejich důvěra ve své schopnosti žáky zaujmout. V mezinárodním srovnání podprůměrná byla i jejich důvěra v žáky a rodiče, přičemž relativně nejsilnější byla nedůvěra ve snahu žáků dosahovat ve škole dobrých výsledků. V mezinárodním srovnání podprůměrné bylo i jejich hodnocení nároků kladených na žáky a spokojenost s profesí učitele. Tím potvrdily informace získané ve výzkumu TIMSS nepříznivé nálezy z výzkumu OECD TALIS (*Teaching and Learning International Survey*), které ukázaly relativně nízké sebepojetí českých učitelů, přičemž nejhůře hodnotili čeští učitelé svoji schopnost přesvědčit žáky, že mohou mít dobré výsledky, že mohou pomoci žákům uvědomit si, jakou má učení hodnotu a rovněž motivovat žáky, kteří nemají zájem o školní práci (Greger; Simonová; Straková 2015).

I když se ve výpovědích učitelů, jak bylo zmiňováno výše, pravděpodobně projeví určité kulturní odlišnosti spočívající v tendenci k negativnímu hodnocení, rozdíly ve výpovědích českých učitelů a učitelů z jiných vyspělých zemí byly natolik vysoké, že je těmito poznatkům rozhodně třeba věnovat pozornost. Tím spíše, že na problematice vztahy mezi žáky a učiteli v českých školách ukazují i výpovědi z žakovských dotazníků z výzkumu PISA.

Index vztahů mezi žáky a učiteli je dlouhodobě jedním z nejnižších mezi zeměmi OECD. V míře souhlasu s tvrzením *Žáci vycházejí s většinou učitelů* se ČR umístila na 12. místě odspoda z 34 zúčastněných zemí OECD, u tvrzení *Většinou učitelů záleží na tom, aby se žákům vedlo dobře* na 25. místě, u tvrzení *Většina učitelů skutečně naslouchá tomu, co říkám* na 5. místě, u tvrzení *Když potřebuji pomoc, učitelé mi ji poskytnou* na 10. místě a u tvrzení *Většina učitelů je ke mně*

spravedlivá na 12. místě. Vztahy žáků k učitelům jsou ve výzkumu PISA pomocí těchto výroků sledovány dlouhodobě a v ČR se příliš nemění. Znamená to, že žáci sice věří, že to s nimi učitelé myslí dobře, ale zároveň se domnívají, že je učitelé neberou vážně a že se jim daří s nimi navazovat dobré vztahy a účinně jim pomáhat pouze v omezené míře.

Závěrem: Poznatky a příležitosti pro českou odbornou komunitu

Tato publikace si kladla za cíl ohlédnout se za dvacetiletou zkušeností masové realizace mezinárodních výzkumů výsledků vzdělávání, zasadit tyto výzkumy do širšího kontextu výzkumů efektivit vzdělávání, shrnout současnou diskusi kolem jejich předností a nedostatků a předestřít perspektivy jejich dalšího vývoje. Byla motivována vírou, že reflexe vývoje a kontext mezinárodní diskuse může pomoci českým výzkumníkům přistupovat k mezinárodním výzkumům i k jejich kritice poučeně a strážlivě a účelně využívat zdroje, které poskytují. K tomu je ovšem potřeba se podrobně seznámit i se způsobem provedení výzkumů a metodologií zpracování dat. Z tohoto důvodu je také v publikaci věnován technickým aspektům výzkumů a zpracování dat velký prostor.

Z přehledu uvedeného v této publikaci je zřejmé, že mezinárodní výzkumy výsledků vzdělávání procházejí kontinuálním vývojem. Hledají cesty jak poskytnout spolehlivé, citlivé a srovnatelné měření v rozšiřující se škále vzdělávacích systémů s velmi různou úrovní vědomostí a dovedností jejich žáků. Vyrovnávají se s požadavkem ověřování nekognitivních vzdělávacích výstupů, jakými jsou personální a sociální dovednosti, které jsou bezesporně pro život mladých lidí velmi důležité, ale obtížně měřitelné. Reagují na bouřlivý rozvoj informačních technologií a snaží se jej využít nejen k ověřování nových dovedností, ale zejména k zefektivnění vlastního měření.

Kolem mezinárodních výzkumů výsledků vzdělávání se soustřeďují nejlepší světoví odborníci působící v oblasti psychometrie, měření výsledků vzdělávání, výběrů a statistického zpracování. Vzhledem k tomu, že jejich výsledky jsou bedlivě sledovány a že musí čelit velmi přísné kritice ze strany příslušníků akademické obce, prochází metodologie výzkumů neustálým zdokonalováním. Ze stejných důvodů je ovšem věnována čím dál větší péče tomu, aby k metodologickým zdrojům získal přístup co nejširší okruh výzkumníků, aby tito výzkumníci rozuměli používané metodologii alespoň na uživatelské úrovni a byli schopni s daty řádně a poučeně pracovat. Metodologické aspekty jsou pro užitečnost a relevanci výzkumů velice důležité a není možno je ignorovat. Na druhé straně je chvályhodné, že se organizátoři výzkumů snaží uživatelům přístup k potřebným zdrojům co nejvíce usnadnit.

I když je pravděpodobné, že se v budoucnosti prodlouží periodicita výzkumů a že zveřejňování výsledků nebude přitahovat tak velkou pozornost jako dosud a stane se víceméně rutinní záležitostí, nelze očekávat, že by mezinárodní výzkumy ztratily na významu. Tvůrci vzdělávacích politik se naučili využívat jejich výsledků a není pochyb o tom, že je budou i nadále financovat. Vzhledem ke kvalitě a objemu získaných dat a k rostoucí dostupnosti specializovaných nástrojů pro jejich analýzu budou tato data rostoucí měrou využívána badateli v oblasti společenských věd a jejich analýzy budou i nadále hrát důležitou úlohu v pedagogickém výzkumu.

Za dvacet let masové realizace přinesly výzkumy celou řadu užitečných poznatků, nepodařilo se jim však přinést univerzální odpověď na to, jaké jsou charakteristiky úspěšných vzdělávacích systémů. V tomto smyslu výzkumy nenaplnily očekávání svých zakladatelů. Toto „selhání“ však bylo stimulem pro rozvoj analytických přístupů a velké invence na straně předních pedagogických výzkumníků i statistiků, kteří hledají cesty, jak ze získaných dat získat věrohodné informace, které pomohou odpovědět na důležité otázky a reagovat na zásadní problémy, kterým vzdělávací systémy aktuálně čelí.

Není pochyb o tom, že některé prezentace výsledků z výzkumu PISA jsou přinejmenším problematické a mohou být velmi zavádějící. Zároveň je čím dál tím zřejmější, že výsledky jsou zatíženy velkou chybou. V tomto ohledu je velmi užitečná neustálá a hlasitá kritika, které jsou výzkumy podrobovány, neboť vybízí k větší obezřetnosti při nakládání s výsledky a eliminuje unáhlené kroky tvůrců vzdělávací politiky. Tato kritika by však neměla být důvodem pro paušální odmítnutí mezinárodních výzkumů výsledků vzdělávání a zdrojů, které nabízejí. Poskytují datové soubory, které svojí velikostí i kvalitou získaných dat přesahují data, jež jsou ve většině vzdělávacích systémů (zejména v menších zemích) získávána v rámci národních výzkumů. Výzvou pro pedagogické výzkumníky zůstává, aby se těchto zdrojů naučili využívat v rámci národních analýz a přídatných výzkumů, které je dále zhodnotí.

Při pohledu na dvacetiletou historii mezinárodních výzkumů v ČR je užitečné se zamyslet nad tím, jaké byly přínosy těchto výzkumů a zda měla jejich realizace rovněž nějaké dopady neblahé.

Výzkumy přinesly pohled na vývoj vědomostí a dovedností českých žáků v povinném vzdělávání. Vzhledem k opakovaným měřením, která umožňují odhalit dlouhodobé trendy, můžeme nyní s dostatečnou jistotou prohlásit, že dochází ke zhoršování vědomostí a dovedností v matematice, že čeští žáci mají stabilní a v mezinárodním srovnání nadprůměrnou úroveň vědomostí a dovedností v přírodovědných předmětech a že čtenářské dovednosti českých žáků jsou ve srovnání vyspělých zemí zhruba průměrné a mají spíše vzestupnou tendenci.

Nyní již také víme, že musíme vždy vyčkat na dlouhodobé trendy a nemůžeme vynášet silné soudy na základě jednotlivých měření. Dosud neumíme řádně zdůvodnit propad, ke kterému došlo ve výsledcích českých žáků v letech 2007 a 2009. Nevíme, zda výsledky odrážely skutečný propad ve vědomostech a dovednostech či zda rozhodující úlohu hrála motivace žáků a učitelů ve vztahu k výzkumu. Propad ve vědomostech a dovednostech mohl být způsoben například určitým chaosem po uzákonění kurikulární reformy, kdy učitelé hledali, jak se s ní ve svých školách a třídách vypořádají. Mohlo se však jednat pouze o absenci motivace snažit se v testování podat dobrý výkon, což je faktor, který výsledky významně ovlivňuje a výzkumníci mají jen malé možnosti ho ovlivnit.

V každém případě je však měření zatíženo velkou chybou a při interpretaci výsledků a trendů je třeba mít tuto skutečnost stále na paměti. To je poučení nejen pro Českou republiku, ale i pro ostatní země, které se mezinárodních výzkumů výsledků vzdělávání účastní. V prvních fázích výzkumů byly daleko čtenější situace, kdy tvůrci vzdělávacích politik zvažovali reformu systémů po každém dílčím zhoršení v některé z měřených oblastí.

Skutečnost, že se propad ve výsledcích v letech 2007 a 2009 nepodařilo uspokojivě vysvětlit, je nicméně přece jen poněkud zneklidňující. Bylo by velmi užitečné provést systematické zkoumání výkyvů v zúčastněných zemích a schopnosti zemí tyto výkyvy vysvětlit, abychom lépe porozuměli jejich zdrojům a míře jejich závažnosti. Zároveň by bylo určitě zajímavé pokusit se hledat vysvětlení v širším společenském kontextu, který přesahuje vzdělávací systém (např. změna společenského klimatu, výchovných metod apod.). Užitečné informace o rozdílech v jednotlivých kohortách a jejich příčinách mohou poskytnout také výzkumy v dospělé populaci⁷⁵.

Je zajímavé se podívat na vývoj vědomostí a dovedností žáků v jednotlivých oblastech

⁷⁵ Zajímavý příspěvek na toto téma pronesl švédský výzkumník Jan-Eric Gustasson na konferenci ECER v roce 2015. Ve svém výzkumu srovnával výsledky stejných věkových kohort ve výzkumu PISA a ve výzkumu PIAAC a snažil se vysvětlit mezikohortní rozdíly ve Švédsku s oporou o existující národní datové zdroje (<http://www.eera-ecer.de/ecer-programmes/conference/20/contribution/34990/>).

v souvislosti s mírou přijetí mezinárodních konceptů těchto oblastí v České republice. Co se týče matematiky, zdá se, že matematické vzdělávání je u nás natolik pevně ukotveno a má natolik pevnou tradici, že konceptualizace matematiky ve výzkumu TIMSS, respektive matematické gramotnosti ve výzkumu PISA neměla na jeho uchopení českými odborníky na matematické vzdělávání a samotnými vzdělavateli žádný dopad. O příčinách zhoršení panuje v komunitě sdílený konsensus, že je způsobeno rozvolněním školního vzdělávání a snahou učinit pro děti školu zábavnou. Není již v silách učitelů donutit žáky k tomu, aby drilovali matematické výpočty a úkony, tyto pak nejsou dostatečně zautomatizovány a to se negativně projevuje ve snížené schopnosti žáků řešit matematické úlohy. Za povšimnutí ovšem také stojí skutečnost, že čeští žáci mají relativně vysokou míru neřešenosti úloh. To může svědčit o tom, že si netroufají řešit úkoly, u kterých na první pohled nevidí způsob řešení, že nejsou zvyklí řešení „vymýšlet“. Tento přístup může svědčit také pro obavu žáků z chybování. Zkoumání příčin zhoršení v matematice je v každém případě důležité a zajímavé, protože může přispět k porozumění tomu, jak se žáci matematice učí, a odhalit i jiné aspekty jejich přístupu ke školní práci.

Relativně dobré výsledky v přírodovědných předmětech můžeme určitě vysvětlit do jisté míry tradičně vysokou pozorností, která je přírodovědným předmětům v českém vzdělávacím systému věnována, s čímž souvisí i jejich relativně štedrá časová dotace. Podrobnější pohled na výsledky českých žáků v přírodovědných předmětech ukazuje, že za relativně dobrými celkovými výkony českých žáků stojí zejména dobré faktické znalosti. Čeští žáci dosahují relativně lepších výsledků v oblasti vědomostí z přírodovědných disciplín, přičemž horší vědomosti vykazují v oblasti vědomostí o přírodních vědách, v rozpoznávání přírodovědných otázek a používání vědeckých důkazů (Palečková a kol. 2007). Koncept přírodovědy používaný ve výzkumu TIMSS a zejména koncept přírodovědné gramotnosti uplatňovaný ve výzkumu PISA se snaží nalézt rovnováhu mezi vědomostmi z jednotlivých přírodovědných oblastí a porozuměním vlastní podstatě přírodovědného bádání. Tento koncept ovšem nenalezl v České republice širšího uplatnění. Žáci mají v přírodovědných předmětech malou příležitost okusit půvab přírodovědného bádání, samostatně tvořit hypotézy, navrhnout experimentální cesty k jejich ověření, realizovat experimenty a vyhodnocovat jejich výsledky. Experimentální práce, které provádějí, mají v nejlepším případě povahu ověřování zákonitostí, případně nácviku rutinních operací. I tyto experimenty mají samozřejmě pro žáky velký význam. Nicméně absence skutečného bádání ve většině českých škol může být jedním z důvodů, proč čeští žáci projevují nízkou ochotu uvažovat v souvislosti s budoucím profesním uplatněním o přírodovědných oborech. Badatelské aktivity by mohly mít nejen motivační efekt, ale mohly by zároveň napomoci rozvoji dovednosti řešit problémy, kritického myšlení, fantazie a dalších oblastí, které jsou aktuálně považovány za hodny kultivace v rámci školního vzdělávání. Zároveň by představovaly i základ pro bádání ve společenskovědních oblastech, pro které žáci rovněž nejsou ve většině škol připravováni. Je důležité si připomenout, že nízké příležitosti k samostatné vědecké práci se týkají všech typů škol včetně gymnázií, která připravují žáky ke studiu na univerzitách.

Existují samozřejmě pedagogové, pro které výše uvedené tvrzení neplatí a kteří badatelské dovednosti rozvíjejí. Pozoruhodná je například iniciativa Heuréka a další aktivity pracovníků katedry didaktiky fyziky MFF UK, které rozvíjejí experimentální přístup k výuce fyziky. Rozhodně však nemůžeme hovořit o širším přijetí konceptu přírodovědné gramotnosti a snaze o její rozvoj ve všech přírodovědných oborech. Zdá se, že tento koncept je v přílišném rozporu s českým pojetím přírodovědného vzdělávání a naráží rovněž na striktní rozdělování přírodovědných disciplín. Relativně nadějnější se jeví přijetí konceptu vyučujícími na prvním stupni. Zde jeho realizace však naráží na jejich nedostatečnou metodologickou, experimentální a mnohdy

i znalostní vybavenost. Vyučující by potřebovali větší odbornou podporu a měla by jim být poskytnuta, protože mnozí z nich by se rozvíjení přírodovědné gramotnosti v plné její šíři rádi a se zájmem věnovali.

Na rozdíl od konceptů matematických a přírodovědných vědomostí a dovedností v mezinárodních výzkumech se v české komunitě velmi dobře ujala konceptualizace čtenářské gramotnosti. Široce užíván je zejména koncept čtenářské gramotnosti výzkumu PIRLS. Slouží výzkumníkům jako teoretická základna pro zkoumání rozmanitých aspektů čtenářské gramotnosti u prvostupňových žáků, přičemž ve výzkumech je využíván nejen koncept, ale také uvolněné úlohy. Na čtenářské úlohy také velmi příznivě reagují vyučující na prvním stupni povinného vzdělávání, které uvolněné úlohy rády a hojně využívají ve výuce. Vstřícné přijetí konceptu má pravděpodobně několik důvodů. Cílený rozvoj čtenářské gramotnosti tak, jak je vymezena ve výzkumu PIRLS, zřejmě dobře doplňuje čtení čítankových textů, které se vždy v české primární škole těšilo velké pozornosti, tedy není v rozporu s žádnými národními zvyklostmi. Ucelený srozumitelný koncept čtenářské gramotnosti zde dosud chyběl, nicméně česká pedagogická veřejnost byla dlouhodobě dovedně vedena k uznání čtenářské gramotnosti jako důležité složky základního vzdělání a k porozumění jejím aspektům lektory společnosti Kritické myšlení. Společnost Kritické myšlení ve své práci vycházela ze zahraničních zdrojů, ze kterých vycházel i konceptuální rámec výzkumu PIRLS, čímž se oba proudy vzájemně podporovaly.

Na vyšších stupních vzdělávání již je přijetí čtenářské gramotnosti v podání mezinárodních výzkumů nižší. Koncept PISA se nepotkává s potřebami pedagogů druhého stupně a středních škol tak, jako koncept PIRLS s potřebami učitelů prvního stupně. Pravděpodobně zde již také tradiční pojetí výuky upřednostňuje jiné aspekty osvojení mateřského jazyka než rozvoj čtenářských dovedností.

Zcela nezastupitelný význam mají mezinárodní výzkumy při zkoumání vzdělanostních nerovností a jejich vývoje. I když čeští sociologové zkoumali souvislosti mezi rodinným zázemím a dosaženým vzděláním a jejich vývoj v čase již před zapojením ČR do mezinárodních výzkumů, nerovnosti v regionálním školství nebyly do té doby sledovány. Výzkumy IEA a zejména výzkum PISA poukázaly na skutečnost, že v České republice existuje relativně silná závislost mezi rodinným zázemím a výsledky vzdělávání a že jsou zde velké rozdíly ve výsledcích škol z hlediska jejich výsledků a složení žáků, kteří je navštěvují.

Tento poznatek je překvapivý ve světle charakteristik české společnosti, která se vyznačuje zejména vysokou ekonomickou homogenitou, se kterou souvisí i velmi nízký podíl rodin žijících pod hranicí chudoby. ČR se dále vyznačuje také homogenitou kulturní (podíl Romů je odhadován na 1,4 – 2,8 % populace, podíl cizinců tvoří cca 4 % populace, z čehož pětinu tvoří občané Slovenské republiky, další početné skupiny tvoří Ukrajinci, Vietnamci a Rusové, kteří se do české společnosti relativně dobře integrují). Poznatek, že tyto rozdíly jsou vysoké nejen v kontextu ostatních vyspělých zemích, které často vzdělávají heterogennější žákovské populace, ale zejména, že se dále zvyšují, vedl k zevrubnějšímu zkoumání této problematiky českými výzkumníky. Důsledkem je vyšší povědomí české pedagogické komunity i české společnosti o této oblasti. Tvůrci vzdělávací politiky mají tím pádem také k dispozici dostatek empirické evidence, která ukazuje, že situace z hlediska nerovností ve vzdělávacím systému vyžaduje jejich pozornost.

Podrobnější zkoumání však zároveň ukázalo, že v české pedagogické komunitě i v české společnosti je výrazně zakořeněno přesvědčení, že nerovnosti, které pozorujeme, jsou víceméně v pořádku a nemá cenu, ba nebylo by žádoucí snažit se je eliminovat. Drtivá většina učitelů

i rodičů se domnívá, že je správné, aby děti byly vzdělávány ve školách praktických a ve školách výběrových, a diferenciací systému v rámci veřejného školství (výběrové školy, do kterých jsou žáci často již na prvním stupni přijímáni na základě přijímacích zkoušek a kde si rodiče platí za nadstandardní služby) je přijímána veřejností spíše vstřícně. Na otázku nerovností v ČR tak poukazují zejména zahraniční instituce, přičemž předmětem kritiky bývá zpravidla diferenciací na 2. stupni povinného vzdělávání, na kterou opakovaně upozorňují představitelé OECD, a špatné podmínky ke vzdělávání a špatné výsledky vzdělávání romských žáků, které jsou kritizovány Evropskou komisí. Tuto kritiku však vnímá česká společnost většinově jako nezaslouženou, a případná opatření, která z ní pramení, jako vnucená a neužitečná.

Prosazovat opatření na zvýšení spravedlivosti je samozřejmě obtížné ve všech zemích, neboť společenské skupiny, které těží z výběrových vzdělávacích příležitostí, jsou zároveň ty, které se těší ve společnosti velkému vlivu a jsou proto schopny zavádění opatření na eliminaci nerovností poměrně účinně bránit. V ČR je však odpor k opatřením na zvýšení spravedlivosti mimořádně silný a vyjadřují jej i společenské skupiny, které by z nich mohly profitovat. Tolerance české společnosti ke vzdělanostním nerovnostem má pravděpodobně kořeny v odporu k rovnostářství propagovanému v socialistické éře a částečně může pramenit také z toho, že čeští občané nevidí přímou souvislost mezi vzděláním a budoucím příjmem (výzkum PIAAC ukázal, že v ČR jsou vědomosti a dovednosti relativně málo oceňovány – ve výzkumu PIAAC patřila ČR k zemím s relativně nejnižšími rozdíly v příjmech pracujících s různou úrovní matematických a čtenářských dovedností). Tolerance společnosti k ekonomickým nerovnostem je v české společnosti podstatně nižší.

Ve vztahu k postojům ke vzdělanostním nerovnostem jsou velmi zářející zejména postoje učitelů. Zde máme opakované doklady o tom, že učitelé silně podporují oddělené vzdělávání žáků s různými kognitivními schopnostmi, tedy na jedné straně ve školách praktických a na druhé straně ve školách výběrových (např. FACTUM INVENIO 2009, Straková a kol. 2013). Zároveň mají nízkou důvěru v možnosti vzdělávání žáků s nějakým typem znevýhodnění, a zejména v možnosti jejich vzdělávání v hlavním vzdělávacím proudu (Jarkovská a kol. 2015, Kaleja; Zelzková 2015). Mezinárodní výzkumy ukázaly, že postoje českých učitelů se jeví velmi negativní i v mezinárodním srovnání, kdy učitelé vyjadřují relativně vysokou nedůvěru k žákům a jejich rodičům i ve své schopnosti a vyjadřují také nízkou spokojenost se svým povoláním.

Těmto poznatkům je třeba v každém případě věnovat další výzkumnou pozornost a měli by jim věnovat pozornost rovněž tvůrci vzdělávací politiky a vzdělavatelé učitelů. Pokud se nepodaří přesvědčit učitele o tom, že jejich úkolem je motivovat k učení všechny žáky včetně žáků pocházejících ze znevýhodněného prostředí a že tento úkol zvládnou, a pokud budou i nadále přistupovat k žákům a jejich rodičům s nedůvěrou, bude se klima na našich školách jen obtížně měnit. K tomu je ovšem potřebné vyjádřit učitelům podporu a důvěru, poskytnout jim potřebnou pomoc a také jejich práci řádně ohodnotit – čeští učitelé jsou jedněmi z nejhůře placených učitelů v zemích OECD. O tom, že je třeba věnovat klimatu českých škol pozornost, svědčí i výpovědi žáků, které ukazují relativně negativní vztah ke škole a k učení a velké rezervy ve vztazích mezi žáky a učiteli.

Z výše uvedeného přehledu je zřejmé, že mezinárodní výzkumy přinesly českým výzkumníkům i tvůrcům vzdělávací politiky celou řadu užitečných poznatků. Tyto poznatky je třeba nepřeceňovat, ale byla by velká škoda jich nevyužít jako podkladu pro další zkoumání a monitorování vývoje českého vzdělávacího systému a jeho výsledků.

Ještě větší pozornost než dosud by měla být věnována možnostem využití výzkumů při

monitorování opatření implementovaných v českém vzdělávacím systému, případně rozšířením, která by zkoumala aspekty fungování systému, které se těší aktuální pozornosti výzkumníků nebo tvůrců vzdělávacích politik.

Nezdá se, že by mezinárodní výzkumy napáchaly v českém prostředí nějaké škody. I když prezentace výsledků v letech 2007 a 2009 byly z dnešního pohledu zbytečně katastrofické, vzdělávací politika na ně nereagovala žádnými systematickými reformními kroky. Nereagovala ani na ostatní poznatky, které jsou konsistentní a péči by si zasloužily – například výše zmiňovaný nárůst nerovností či negativní postoje žáků a učitelů. Je možné, že informace o zhoršování výsledků v tomto období posílila snahy o tvorbu standardů a zavádění plošného testování. Nicméně tyto snahy se v české vzdělávací politice objevily již před zveřejněním výsledků a neměly na reálnou situaci v systému žádný konkrétní dopad. Byly vytvořeny jakési standardy, u nichž nemáme žádné doklady o tom, zda a jak ovlivňují práci učitelů. Plošné testování je čas od času zaváděno s tím, že stále není jasné, jaké cíle by mělo ve vzdělávacím systému plnit. I v české komunitě se však objevují hlasy, které přičítají výzkumu PISA vinu za rozmanitá rozhodnutí tvůrců vzdělávacích politik⁷⁶.

Kritici, kteří považují výzkumy za neužitečné – například z toho důvodu, že málo vycházejí z české reality a potřeb našeho vzdělávacího systému, by mohli jistě zpochybňovat vynaložené finanční prostředky a čas žáků, který by mohl místo pro testování být využit pro výuku. Finanční argument ovšem v období čerpání prostředků z ESF neobstojí, neboť tyto prostředky jsou zcela evidentně vynakládány na řadu aktivit, které jsou podstatně problematičtější než aktivity výzkumné, neboť mohou mít na vzdělávací systém trvalé neblahé dopady.

V ČR také neplatí, že by pojetí vzdělávacích cílů deformovalo výuku. Jak bylo uvedeno výše, koncept matematiky a přírodovědné oblasti český vzdělávací systém nikterak nepoznamenal, koncept čtenářské gramotnosti je vnímán jako užitečný a pojí se s testovými úlohami, které jsou učitelkami široce užívány. Nejsme svědky žádné systematické přípravy na mezinárodní testování a je velmi pravděpodobné, že kdyby tato příprava nastávala, byla by ve čtenářské a přírodovědné gramotnosti pro žáky spíše užitečná a v matematice by odpovídala běžné výuce.

Ve vztahu k dopadům mezinárodních výzkumů na český vzdělávací systém se nejvíce nabízí otázka, co by se dalo udělat pro to, aby zdroje, ke kterým má ČR v souvislosti s těmito výzkumy přístup, byly využity lépe, a to zejména pro osvětu a další vzdělávání české pedagogické komunity. Potužníková a kol. (2014) formulovali tři oblasti, kterým je třeba se věnovat, chceme-li využít poznatků z mezinárodních výzkumů pro zkvalitňování výuky:

1. Propojit mezinárodní výzkumy s domácími didaktickými výzkumy. Jedná se zejména o rozšiřování a prohlubování české realizace ve fázi jejich přípravy, a to zejména v rovině oborově didaktické. To mimo jiné znamená přizvat k realizaci domácího šetření odborníky na didaktiku, kteří by například rozšířili dotazníky, případně zjemnili kategorie chybných žákovských odpovědí a učinili tak výsledky užitečnější pro oborově didaktiky.
2. Podrobit existující datové zdroje sekundárním analýzám s cílem získat podrobnější informace o silných a slabých stránkách českých žáků v jednotlivých předmětech.
3. Využít získaných poznatků k utváření nové kultury učebních úloh, se kterými se žáci dosud ve výuce běžně nesetkávají, a k zařazení těchto úloh do učebnic a do výuky.

⁷⁶ Stejně hlasy se objevují i v akademické komunitě. Např. Holubová (2015) ve své bakalářské práci viní výzkum PISA nejen z nárůstu testování v ČR, ale zejména z toho, že v jeho důsledku došlo k novelizaci školského zákona směrem k inkluzivním přístupům.

I když se v posledních letech setkáváme s rostoucím využíváním dat z mezinárodních výzkumů českými výzkumníky, řada zajímavých aspektů zůstává v české výzkumné komunitě nepovšimnuta. V ČR také chybí souborné analýzy podobné těm, které vycházejí v jiných vyspělých zemích (viz např. již zmiňovaná Wendt a kol. 2016). Setkáváme se buď s izolovanými vědeckými články z pera akademiků, nebo s deskriptivními zprávami z dílny realizátorů výzkumů, přičemž jen zřídka dochází k jejich propojení. Chybí tedy kombinace teoretického ukotvení, aktuálních a důležitých výzkumných otázek, metodologické erudice a praktické aplikace výsledků.

Již zde byla řeč o uplatnění konceptů jednotlivých hodnocených oblastí v českém prostředí. Způsob, jak jsou uchopeny vzdělávací cíle, ale zejména jak se vymezení oblastí průběžně aktualizuje, vypovídá hodně o tom, jak se vyvíjí náhled na testované oblasti ve světě. Bylo by dobré, kdyby čeští odborníci, kteří utvářejí koncepty těchto oblastí pro český vzdělávací systém, měli příležitost nejen se seznámit s výslednými koncepty, ale také se zúčastnit mezinárodních diskusí o těchto konceptech. Toho je možno docílit zapojením odborníků do realizace výzkumů a delegovat je do expertních skupin, které jsou k jednotlivým oblastem zřizovány. Tím bychom docílili toho, že by naši odborníci přispěli do mezinárodní debaty svými vědomostmi a měli by příležitost ovlivnit výslednou podobu konceptů a nástrojů tak, aby nám lépe vyhovovala. Zároveň by se však mohli poučit z mezinárodní debaty a obohatit o získané zkušenosti debatu národní.

Již bylo zmiňováno, že realizace výzkumů je v různých zemích pojímána různě. V některých zemích jsou na každý běh výzkumu vypisována výběrová řízení, a jednotlivé cykly realizují různé instituce. To se může jevit jako neefektivní, protože každý nový subjekt se musí seznámit s řadou rutinních postupů, které se opakují, a vybudovat infrastrukturu k řádné realizaci výzkumu. Výhodou tohoto přístupu je však skutečnost, že se o realizaci výzkumů poučí řada odborníků a řada odborníků v průběhu realizace výzkumů profituje ze spolupráce s mezinárodním koordinačním centrem. V některých systémech dochází k tomu, že infrastruktura zůstává stejná, ale dochází k povinné spolupráci různých pracovišť a odborníků tak, aby docházelo k maximální výměně mezi národními a mezinárodními zdroji v dílčích oblastech.

V ČR jsou výzkumy bezesporu realizovány ve vynikající kvalitě, neboť jejich realizaci zajišťují odborníci, kteří mají v této oblasti velkou zkušenost. Určitě by však bylo možno věnovat větší péči propojení národního koordinačního centra s akademickými pracovišti, která se zabývají výukou hodnocených oblastí a zkoumáním postojů a chování pedagogických pracovníků. Zapojení akademiků do realizace výzkumů by jim umožnilo navazovat nové kontakty se zahraničními kolegy, přispět svojí erudicí ke zvyšování kvality výzkumů doma i v zahraničí a zároveň zprostředkovat českým kolegům poznatky z mezinárodních diskusí o rozmanitých aspektech hodnocení výsledků vzdělávání v národních i mezinárodních šetřeních.

Samostatnou otázkou samozřejmě zůstává institucionální ukotvení mezinárodních výzkumů pod Českou školní inspekcí. Za jeho výhodu lze považovat garanci návratnosti (účast se stává pro školy v zásadě povinnou) i profesionální administraci inspektory, kteří požívají ve školním prostředí autoritu a umějí se v něm pohybovat. Na druhé straně tím aktivita z pohledu škol pozbývá výzkumného charakteru, neboť je realizována kontrolní institucí, což může mít teoreticky dopad na odpovědi respondentů. Od ČŠI rovněž nelze očekávat koncepční uchopení realizace mezinárodních výzkumů z hlediska monitoringu celého vzdělávacího systému, neboť takovýto komplexní monitoring není jejím úkolem.

Je tedy nutno se zabývat otázkou, jak zajistit, aby byla data využita k hlubším analýzám a tedy více než nyní přispívala k poznání týkajícímu se českého vzdělávacího systému a jeho vývoje a odpovídala na aktuální otázky české vzdělávací politiky. Vzhledem k výše zmiňova-

nému institucionálnímu ukotvení a počtu studií a počtu pracovníků, kteří mají zodpovědnost za českou implementaci výzkumů, nelze očekávat od realizačního týmu žádné sofistikované analýzy. I když data jsou k dispozici dalším výzkumníkům jak na stránkách mezinárodních výzkumů, tak na stránkách České školní inspekce⁷⁷, jsou již výzkumníci limitováni při jejich zpracování způsobem výběru respondentů a volbou národních proměnných. Ty jsou většinou voleny poněkud nahodile bez jasného analytického záměru. Potenciál výzkumů pro zodpovězení otázek vzdělávací politiky na národní úrovni je tak do jisté míry promrhán.

Přístup tvůrců vzdělávací politiky k mezinárodním výzkumům se bohužel pojí s celkově nekoncepčním přístupem k monitoringu vzdělávacího systému a získávání zpětné vazby o jeho fungování a vývoji, který je pro českou státní správu charakteristický a který je opakovaně reflektován jak národními, tak zahraničními odborníky (např. Santiago a kol. 2012). S tím souvisí i ad hoc způsob rozhodování o účasti ČR v nabízených výzkumech, který je v posledních letech ještě deformován skutečností, že jsou výzkumy realizovány v rámci projektů hrazených z prostředků Evropského sociálního fondu.

Za téměř trestuhodné je možno označit nevyužití vzdělávacích možností, které mezinárodní výzkumy poskytují v oblasti metodologie měření výsledků vzdělávání. Z pojetí testových šetření, která jsou realizována v ČR, jako je maturitní zkouška nebo občasná testování 5. a 9. ročníků, je zřejmé, že jejich realizátoři by mohli značně profitovat z bohatých zdrojů, které mezinárodní výzkumy nabízejí. Je velká škoda, že národní a mezinárodní testování fungují v ČR z metodologického hlediska jako zcela oddělené světy a že realizátoři národního testování nepovažují za potřebné osvojit si metody používané v zahraničí a nevyužívají vzdělávacích možností vytvářených při mezinárodních výzkumech.

Jako vysoce žádoucí se například jeví využívání maticového designu při realizaci výběrových šetření. Česká republika má velmi málo informací o tom, jak se daří naplňovat rámcové vzdělávací programy v jednotlivých oblastech vzdělávání. Nevíme nejenom nic o tom, co umějí čeští žáci například v dějepise a v zeměpise, ale také v jiných aspektech matematiky, českého jazyka a přírodovědných předmětů než těch, které jsou ověřovány v mezinárodních výzkumech. Tyto informace jsou nicméně důležité pro přemýšlení o úpravách kurikula, o zvyšování, respektive snižování náročnosti vzdělávacích cílů a podobně. Podrobnější přehled o vědomostech a dovednostech žáků v jednotlivých vzdělávacích oblastech však nemůžeme získat prostřednictvím jednotného testu. Takový test může obsahovat jen omezené množství úloh, které v žádném případě nemohou obsáhnout celou šíři vzdělávacích cílů, které jsou obsaženy v rámcových vzdělávacích programech. Administrovat jednotný test je neefektivní a neúčelné v situaci, kdy potřebujeme ověřit široký rozsah vzdělávacích cílů a nepotřebujeme porovnávat výsledky jednotlivých žáků. Aby bylo možno administrovat různým žákům různé sady úloh, je třeba se naučit agregovat získané výsledky tak, aby bylo možno je prezentovat na společné škále. K tomu je možno využít metodologie IRT, která byla podrobněji popsána v oddíle 3.3.1. V rámci mezinárodních výzkumů jsou k této problematice pravidelně organizována školení.

Další oblastí, ve které by bylo možno využít zdrojů nabízených mezinárodními výzkumy, je měření výsledků vzdělávání prostřednictvím počítačů. Zde se jedná v první řadě o oblast adaptivního testování, ve které je opět třeba prezentovat na společné škále výsledky z různých sad úloh. Důležité poučení je však možno získat i o konstrukci úloh a způsobech jejich vyhodnocování.

Velkou inspirací pro českou odbornou komunitu by se mělo stát pečlivé vymezení testovaných

⁷⁷ Zároveň je nutno připomenout, že tam nejsou umístěna všechna data: chybí data z počátečních cyklů výzkumů.

oblastí, které umožňuje uživatelům testů si vytvořit názor na to, do jaké míry test hodnotí ty vzdělávací cíle, které i oni považují za důležité, tedy zda pro ně má nějakou vypovídací hodnotu. V České republice se s pečlivým popisem cílů vzdělávání, které test ověřuje, setkáváme zřídka. Tím tvůrci testu opomíjejí příležitost vzdělávací cíle konkretizovat prostřednictvím testových úloh a tím kultivovat dovednosti v oblasti hodnocení výsledků vzdělávání. Zároveň podporují představu, že test je jakási černá skříňka, která jaksi automaticky měří to, co je potřeba. Tato skutečnost velmi vyniká například v situaci, kdy některé školy používají k přijímacím zkouškám testy CERMAT a jiné testy agentury SCIO. Tyto testy se velmi liší obsahem a vzdělávacími cíli, které ověřují. Ani jedna z obou organizací však neposkytuje uživatelům podrobnou informaci o ověřovaných cílech a jejich zdůvodnění. Bylo by zajímavé zjistit, na základě jakých kritérií se školy rozhodují, jaké testy pro své přijímací řízení vyberou.

Velmi důležitou metodologickou inspirací by se mělo stát propojování výsledků z jednotlivých kol testování. V České republice se bohužel stále ještě setkáváme s tím, že jsou na základě národního testování vyvozovány závěry o vývoji výsledků žáků v situaci, kdy není kontrolována srovnatelná obtížnost testů, respektive kdy výsledky nejsou prezentovány na stejné škále. Tedy například hovoříme o zhoršování žáků v průběhu dvou let, kdy test ve druhém roce byl obtížnější než v prvním a výsledky žáků v něm tedy byly logicky horší. To je velmi matoucí pro odbornou a širokou veřejnost. Státní instituce by měly být schopny vždy srovnatelnost zajistit a doložit, jak to činí. Měly by tak aktivně přispívat ke zvyšování „testové gramotnosti“ odborné komunity a ne ještě posilovat stávající diletantismus.

Jak již bylo uvedeno v kapitole 5, v České republice hodně rezonují kritické hlasy, které se (zejména ve vztahu k výzkumu PISA) ozývají ze zahraničí. Mnohé z těchto hlasů jsou však založeny na nedostatečné znalosti metodologických aspektů výzkumů. I když k mezinárodním výzkumům samozřejmě existují důležité a případné připomínky, žádná z těchto připomínek se dosud neukázala natolik pádná, aby výsledky výzkumů zásadněji znevěrohodnila. Tyto připomínky však mají velmi sofistikovaný charakter. Metodologie mezinárodních výzkumů je velmi dobře promyšlená a založená na aplikaci nejnovějších vědeckých poznatků v oboru. Na druhou stranu povrchní metodologická kritika, která nevychází z důkladné znalosti používaných postupů, není zpravidla relevantní.

PISA je často kritizována za to, že nerespektuje národní tradice jednotlivých vzdělávacích systémů a snaží se všem státům vnutit (nedostatečně odborně podložené) koncepty gramotností. Někteří jí vyčítají, že podporuje odklon od tradičního předávání poznatků a podporuje v povrchnosti, jiní naopak, že obrací pozornost vzdělavatelů výlučně k poznatkům a vede k zanedbávání jiných důležitých vzdělávacích cílů, jako jsou například sociální a personální kompetence, emoční inteligence a podobně. Stejně rozporuplné jsou pak výhrady týkající se vztahu výzkumu PISA a spravedlivosti ve vzdělávání. Někteří mu vyčítají, že prostřednictvím testování a porovnávání výsledků podporuje konkurenci, která vede k nežádoucí diferenciaci systémů a růstu nerovností, jiní se naopak rozhořčují nad tím, že PISA upozorňuje na nerovnosti a podsouvá tvůrcům vzdělávacích politik, že by měli odsunout rozdělování dětí do co nejvyššího věku, čímž podporuje systémy v sociálním inženýrství. Někteří kritici pak zpochybňují samu ideu, že je možno definovat kvalitu vzdělávacího procesu a jeho výstupů a tuto kvalitu měřit, a zdůrazňují jedinečnost individuálních interakcí žáků a učitelů a individuálního rozvoje každého dítěte v jeho jedinečných podmínkách.

V České republice, jak již bylo uvedeno, našly mezinárodní kritiky velkou odezvu. V českém prostředí rezonují zejména výhrady k neblahé úloze OECD, která chce vnutit všem zemím stejné vzdělávací standardy, a k nekompatibilitě konceptu gramotností s naší vzdělávací tradicí.

Na stejném základě byly a jsou kritizovány také klíčové kompetence figurující v rámcových vzdělávacích programech.

Je však zajímavé, že i když se čeští kritici v případě mezinárodních výzkumů připojují rovněž k výhradám týkajícím se testování a jeho negativních dopadů na vzdělávací systémy, k domácím testovacím aktivitám se vyjadřují minimálně. Akademická obec mlčí k obsahu testů i k jeho možným dopadům přesto, že v České republice je nad čím se zamýšlet. Zatímco v českých pedagogických časopisech je diskutováno, jak PISA negativně ovlivňuje vzdělávání tím, že se zaměřuje na omezené cíle, nesetkáváme se s texty, které by se zabývaly tím, jak poškozuje vzdělávací systém například koncepce maturitní zkoušky z českého jazyka. Stejně tak se kriticky zabýváme metodologickými aspekty výzkumu PISA, ale metodologické aspekty maturitní zkoušky, které přímo ovlivňují osudy tisíců mladých lidí, zůstávají v české pedagogické komunitě nepovšimnuty stejně jako metodologie plošných testů, kterým jsou čas od času vystavováni žáci 5. a 9. ročníku. Česká akademická obec se zabývá negativními dopady výzkumu PISA na pohodu mladých lidí, ale necítí potřebu se zabývat dopadem plánovaných přijímacích zkoušek do maturitních oborů.

Tyto přijímací zkoušky přitom mohou mít zcela zásadní dopad na kvalitu vzdělávání mladých lidí, kteří se nebudou do maturitních oborů hlásit. Je velmi pravděpodobné, že základní školy budou implicitně nebo explicitně hodnoceny na základě úspěšnosti svých žáků v těchto přijímacích zkouškách. Z toho důvodu budou v posledních ročnících povinného vzdělávání věnovat zvýšenou pozornost přípravě žáků, kteří budou tyto zkoušky skládat. Žáci, kteří se hlásí do učebních oborů bez maturity, tak budou znevýhodněni hned dvakrát: za prvé tím, že je rodiče do maturitního oboru nepřihlásí a za druhé tím, že jejich vzdělávání bude pro jejich učitele méně důležité.

Na jedné straně je výborné, že se čeští pedagogové zajímají o mezinárodní výzkumy a diskusi, která k nim probíhá v zahraničních časopisech. Na druhé straně je skutečně zarážející, že nepovažují za potřebné přispět svojí erudicí a kritickým pohledem do diskuse, která probíhá v České republice o hodnocení, jež mají rozhodující dopad na osudy mladých lidí a zcela zásadním způsobem ovlivňují i to, na co budou učitelé klást ve škole důraz.

Bylo by poctivé a užitečné, kdyby kritici mezinárodních výzkumů (čeští i zahraniční) spolu s kritikou vždy uvedli i východiska, z kterých přistupují k mezinárodním výzkumům a k výzkumům efektivit vzdělávání. Zejména by měli uvést, zda se domnívají, že je možno konceptualizovat a měřit kvalitu vzdělávacích procesů a výsledků, a zda považují za žádoucí, aby pedagogický výzkum sloužil jako poznatková báze pro zvyšování kvality vzdělávání. I když je jistě zajímavé se poučit i o východiskách kritiků, kteří na obě výše uvedené otázky odpovídají záporně, tyto odborníci nemohou svými připomínkami přispět k tomu, aby se zvyšovala užitečnost a kvalita mezinárodních výzkumů výsledků vzdělávání nebo národních aktivit, které jsou od nich odvozeny. S kritiky, kteří odpovídají pozitivně, je možno se domlouvat o tom, co lze objektivně měřit a jakým způsobem, co je žádoucí měřit a jak zvýšit užitečnost výzkumné činnosti doma i v zahraničí a účinně eliminovat její případné negativní dopady.

Literatura

- Basl, J.; Straková, J.; Veselý, A. (2009). Rozvoj občanské společnosti a demokracie pohledem mladých Čechů. *Pedagogika* 4/2009, 365-379.
- Basl, J.; Tomášek, V. (2013). Vývoj výsledků v čase. 6-21 in TIMSS 2011 PIRLS 2011. *Vybraná zjištění*. Praha: ČŠI.
- Baumert, J.; Fend, H.; O'Neil, H.F.; Peschar, J.L. (1998). *Prepared for life- long learning: Frame of reference for the measurement of self-regulated learning as a cross-curricular competency (CCC) in the PISA project*. Paris: OECD.
- Beaton, A. E.; Mullis, I.V.S.; Martin, M.O.; Gonzalez, E. J.; Kelly, D. L.; Smith, T.A. (1997a). *The Mathematics Achievement in the Middle School Years*. Boston: TIMSS International Study Center.
- Beaton, A. E.; Martin, M.O.; Mullis, I.V.S.; Gonzalez, E.J.; Smith, T.A.; Kelly, D. L. (1997b). *The Science Achievement in the Middle School Years*. Boston: TIMSS International Study Center.
- Beller, M. (2013). Technologies in Large-Scale Assessments: New Directions, Challenges, and Opportunities. 25-46 in von Davier, M., Gonzales, E., Kirch, I., Yamamoto, K. (Eds.) *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research*. Dordrecht: Springer.
- Binkley, M., Sternberg, R., Jones, S. Nohara, D. (1999). An overarching framework for understanding and assessing lifeskills. In *Frameworks: Working drafts*. Briefing materials for National Study Managers meeting, Luxembourg.
- Bryk, A. S.; Raudenbush, S. W. (1992). *Hierarchical linear models in social and behavioral research: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Blossfeld, H.P.; Schneider, T. & Doll, J. (2009) Methodological Advantages of Panel Studies. Designing the New National Educational Panel Study (NEPS) in Germany. *Journal for Educational Research Online* 1(1), 10-32.
- Burdett, N.; Sturman, L. (2013). *A Comparison of PISA and TIMSS against England's National Curriculum*. http://www.iea.nl/fileadmin/user_upload/IRC/IRC_2013/Papers/IRC-2013_Burdett_Sturman.pdf
- Carstens, R.; Pelgrum, J.W. (2009). *SITES 2006 Technical Report*. Amsterdam: IEA.
- Coleman, J.S.; Campbell, E.Q.; Hobson, C.F.; McPartland, J.; Mood, A.M.; Weinfeld, F.D., et al. (1966). *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office.
- Creemers, B.P.M.; Kyriakides, L. (2008). *The dynamics of educational effectiveness: a contribution to policy, practice and theory in contemporary schools*. London and New York: Routledge.
- Creemers, B.P.M.; Kyriakides, L. (2010). *Using the Dynamic Model to develop an evidence-based and theory-driven approach to school improvement*. *Irish Educational Studies*, 29, 1, 5-23.
- Csapó, B. (2012). Developing a framework for diagnostic assessment of early science. 55-78 in Bernholt, S., Neumann, K., Nentwig, P. (Eds): *Making it tangible - Learning outcomes in science education*. Waxmann, Münster.
- Csapó, B. (2014). Long-term impact of school readiness: Results from a ten-year longitudinal study. 43 in Korom, E., Pásztor A. (Eds): PÉK 2014. XII. *Pedagógiai Értékelési Konferencia. Program - Előadás-összefoglalók*. Szeged, 1.-3.5. 2014. Szegedi Tudományegyetem, Szeged.
- Csapó, B.; Molnár, Gy.; Nagy, J. (2014). Computer-based assessment of school readiness and early reasoning. *Journal of Educational Psychology*. 106. 2. 639-650. doi: 10.1037/a0035756.
- ČŠI. 2014. *Výsledky PISA 2012: Kreativní řešení problémů Schopnosti žáků řešit problémy z reálného života* (V. díl). Praha: ČŠI.
- Davis, A. (2015). *PISA: Is it Really Possible to Test All Educationally Significant Achievements with High Levels of Reliability?* Příspěvek prezentován na konferenci ECER 2015.

- Dalehefte, I.M.; Rieck, K.; Wendt, H.; Kasper, D.; Köller, O.; Bos, W. (2015). Mathematische Kompetenzen von Lernenden aus SINUS-Grundschulen im Vergleich zu TIMSS 2011. 185-200 in Wendt, H., Stubbe, T. C., Schwippert, K., Bos, W. (Eds.) *10 Jahre international vergleichende Schulleistungsforschung in der Grundschule. Vertiefende Analysen zu IGLU und TIMSS 2001 bis 2011*. Münster: Waxmann.
- Davis, A. (2013) How far can we aspire to consistency when assessing learning? *Ethics and Education* 8, 3, 217-228.
- Davis, A.; Winch, C. ; Lum, G. (2015) *Educational Assessment on Trial*. London: Bloomsbury.
- Durlak, J. A.; Weissberg, R.P.; Dymnicki, A. B.; Taylor, R. D.; Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development* 82(1), 405-432.
- Dvořák, D. (2012). Nový institucionalismus v pedagogice. *Studia paedagogica*, 17, 2, 9-26.
- Dvořák, D.; Straková, J. (2016). Konkurence mezi školami a výsledky žáků v České republice: pohled zblízka na šetření PISA 2012. *Pedagogika*, 66, 2, 206-229.
- Elias, J.M. (2003). *Academic and social-emotional learning*. Brusel: International Academy of Education.
- Elley, W.B. (Ed.). (1992). *The IEA Study of Reading Literacy: Achievement and Instruction in Thirty-Two School Systems*. Oxford: Pergamon Press.
- European Commission. (2012). *First European Survey on Language Competences*. http://ec.europa.eu/languages/policy/strategic-framework/documents/language-survey-final-report_en.pdf.
- EURYDICE. (2002). *Key competences: A developing concept in general compulsory education*. Eurydice. <http://www.eurydice.org>.
- FACTUM INVENIO. (2009). *Analýza předpokladů a vzdělávacích potřeb pedagogických pracovníků pro zkvalitňování jejich pedagogické práce*. Praha: Factum Invenio.
- Fay, K.; Wendt, H.; Kasper, D. (2015). Soziale Kompetenz von Grundschulkindern in Deutschland. 55-75 in Wendt, H., Stubbe, T.C., Schwippert, K., Bos, W. (Eds.) *10 Jahre international vergleichende Schulleistungsforschung in der Grundschule. Vertiefende Analysen zu IGLU und TIMSS 2001 bis 2011*. Münster: Waxmann.
- Fraillon J.; Ainley, J.; Gebhart, E.; Schulz, W. (2013). *Measuring Computer and Information Literacy across Countries*. http://www.iea.nl/fileadmin/user_upload/IRC/IRC_2013/Papers/IRC-2013_Fraillon_etal.pdf
- Goldstein, H. (1987). *Multilevel models in educational and social research*. New York: Oxford University Press.
- Goldstein, H.; Woodhouse, G. (2000). School Effectiveness Research and Educational Policy. *Oxford Review of Education*, 26 (3+4), 353-363.
- Goldstein, H. (2004). International comparisons of student attainment: some issues arising from the PISA study. *Assessment in Education*, 11 (3), 319-330.
- Gonzales, E. (2014). Calculating Standard Errors of Sample Statistics when Using International Large-Scale Assessment Data. 59 - 73 in Strietholt, R., Bos, W., Gustafsson, J.E., Rosén, M. (Eds.) *Educational Policy Evaluation through International Comparative Assessments*. Göttingen: Waxmann.
- Greger, D.; Simonová, J.; Straková, J. (2015). Poznatky o pedagogických přesvědčeních českých učitelů. 136-152 in Greger, D., Simonová, J., Straková, J. (Eds.) *Spravedlivý start? Nerovné šance v předškolním vzdělávání a při přechodu na základní školu*. Praha : Pedagogická fakulta UK.
- Greiff, S.; Wüstenberg, S.; Funke, J. (2012). *Dynamic Problem Solving: A new measurement perspective*. *Applied Psychological Measurement*, 36(3), 189-213.

- Gustafsson, J.E.; Rosén, M. (2014). Quality and Credibility of International Studies. 19-32 in Strietholt, R., Bos, W. Gustafsson, J.E., Rosén, M. (Eds.) *Educational Policy Evaluation through International Comparative Assessments*. Göttingen: Waxmann.
- Hanushek, E. A.; Woessmann, L. (2009). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *NBER Working Paper 14633*. (January). Cambridge: National Bureau of Economic Research.
- Hanushek, E. A.; Woessmann, L. (2011). The economics of international differences in educational achievement. 89-200 in Hanushek E. A., Stephen, M., Woessmann, L. (Eds.) *Handbook of the economics of education*, Vol. 3. Amsterdam: North Holland.
- Hanushek, E. A.; Woessmann, L. (2013). The Role of International Assessments of Cognitive Skills in the Analysis of Growth and Development. 47 - 66 in von Davier, M., Gonzalez, E., Kirsch, E., Yamamoto, K. (Eds). *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research*. Dordrecht: Springer.
- Heath, A.; Kilpi-Jakonen, E. (2012). Immigrant Children's Age at Arrival and Assessment Results, *OECD Education Working Papers*, No. 75, OECD Publishing. <http://dx.doi.org/10.1787/5k993zsz6g7h-en>.
- Hiebert, J., R.; Gallimore, H.; Garnier a kol. (2003). *Teaching Mathematics in Seven Countries - Results from the TIMSS 1999 Video Study*. Washington: NCES.
- Hirschová, M.; Kreidl, M. (2012). Vliv počtu sourozenců na matematickou, čtenářskou a přírodovědnou gramotnost v ČR. *Sociologický časopis / Czech Sociological Review*, 697-720, 4/2012.
- Holubová, J. 2015. *PISA testování a novela školského zákona v perspektivě antropologie vzdělávání*. https://is.muni.cz/th/413294/fss_b/.
- Hoy, W. K.; Tarter, C. J.; Woolfolk Hoy, A. (2000). Academic optimism of schools: A force for student achievement. *American Educational Research Journal*, 43, 425-446. doi:10.3102/00028312043003425.
- Hrbáčková, K. (2009). Autoregulace procesu čtenářského rozvoje žáků na 1. stupni základní školy. *Pedagogická orientace* 19(4), 74-91.
- Hutchison, D.; Schagen, I. (2006). *Comparisons Between PISA and TIMSS - Are We the Man with Two Watches?* http://www.iea.nl/fileadmin/user_upload/IRC/IRC_2006/Papers/IRC2006_Hutchison_Schagen.pdf.
- Chudowsky, N.; Pellegrino, J.W. (2003). Large-scale assessments that support learning: What will it take?. *Theory Into Practice* 42, 1, 75-83.
- Chvál, M.; Procházková, I.; Straková, J. (2015). *Hodnocení výsledků vzdělávání didaktickými testy*. Praha: Česká školní inspekce.
- Jarkovská, L.; Lišková, K.; Obrovská, J.; Suralová, A. (2015). *Etnická rozmanitost ve škole - Stejnost v různosti*. Praha: Portál.
- Johnson, E.G.; Rust, K.F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics*, 17(2), 175-190.
- Janík, T.; Seidel, T. (Eds.) (2009). *The power of videostudies in investigating teaching and learning in the classroom*. Göttingen: Waxmann.
- Janík, T. (2015). *O praxi nevzdělanosti s K.P. Liessmannem*. <http://www.ucitelske-listy.cz/2015/10/tomas-janik-o-praxi-nevzdelanosti-s-k-p.html>.
- Johnstone, C.; M. Thurlow; J. Altman; J. Timmons ; K. Keto (2009). Assistive technology approaches for large-scale assessment: Perceptions of teachers of students with visual impairments. *Exceptionality* 17(2), 66-75.

- Judkins, D.R. (1990). Fay's Method of Variance Estimation. *Journal of Official Statistics*, 6(3), 223-239.
- Johnson, E.G.; Rust, K.F. (1992). Population inferences and variance estimation for NAEP data. *Journal of Educational Statistics*, 17(2), 175-190.
- Kaleja, M.; Zezulková, E. (2015). *Sociologická analýza zaměřená na porovnání dosažených výsledků žáků se speciálními vzdělávacími potřebami v různých typech škol*. Opava: Centrum empirických výzkumů Fakulty veřejných politik Slezské univerzity v Opavě.
- Kaščák, O.; Pupala, B. (2011). Pisa v kritické perspektive. *Orbis Scholae* 2011/1, 53-68.
- Katrňák, T.; Simonová, N.; Fónadová, L. (2013). Od diferenciacie k diverzifikaci: test MMI a EMI v českém středním vzdělávání v první dekádě 21. století. *Sociologický časopis / Czech Sociological Review*, 4/2013, 491-520.
- Kerr, D.; Sturman, L.; Schulz, W.; Burge, B. (2010). *ICCS 2009 European Report Civic knowledge, attitudes, and engagement among lower secondary students in 24 European countries*. Amsterdam: IEA.
- Klieme, E.; Pauli, C.; Reusser, K. (2009). The Pythagoras Study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. 137-160 in Janik, T., Seidel, T. (Eds.) *The power of video studies in investigating teaching and learning in the classroom*. New York: Waxmann.
- Klieme, E. (2013). The Role of Large-Scale Assessments in Research on Educational Effectiveness and School Development. 115-148 in von Davier, M., Gonzales, E., Kirch, I., Yamamoto, K. (Eds.) *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research*. Dordrecht: Springer.
- Koenig, J.A. (Ed.). (2001). *Assessing 21st Century Skills*. http://www.nap.edu/catalog.php?record_id=13215
- Koretz, D. (2005). Alignment, High Stakes, and the Inflation of Test Scores. *Yearbook of the National Society for the Study of Education*. 104, 2, 99-118.
- Kozma, R. (2009). Assessing and teaching twenty-first century skills: A call to action. 13-23 in Schueremann, F., Bjornsson, J. (Eds.). *The transition to computer-based assessment: New approaches to skills assessment and implications for large scale assessment*. Brussels: European Communities. <http://www.worldclassarena.net/doc/file17.pdf>.
- Kramplová, I. a kol. (2012). *Národní zpráva PIRLS 2011*. Praha: ČŠI.
- Křížová, I. a kol. (2001). *Znalosti, dovednosti a postoje čtrnáctiletých žáků v oblasti výchovy k občanství*. Praha: ÚIV.
- Kyriakides, L.; Charalambos Y. Ch. (2014). Educational Effectiveness Research and International Comparative Studies: Looking Back and Looking Forward. 33-50 in Strietholt, R., Bos, W. Gustafsson, J.E., Rosén, M. (Eds.) *Educational Policy Evaluation through International Comparative Assessments*. Göttingen: Waxmann.
- Levin, H., N. (2013). The Utility and Need for Incorporating Noncognitive Skills Into Large-Scale Educational Assessments. 67 - 86 in von Davier, M., Gonzalez, E., Kirsch, E., Yamamoto, K. (Eds.) *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research*. Dordrecht: Springer.
- Liessmann, K. P. (2014). *Geisterstunde. Die Praxis der Unbildung. Eine Streitschrift*. Wien: Paul Zsolnay Verlag.
- Linhartová, T.; Horáčková, K. (2015). Děti a žáci s odlišným mateřským jazykem v předškolním a primárním vzdělávání. 38-50 in Greger, D., Simonová, J., Straková, J. (Eds.) *Spravedlivý start?*

- Nerovné šance v předškolním vzdělávání a při přechodu na základní školu.* Praha: Pedagogická fakulta UK.
- Loony, J., W. (2011). Integrating Formative and Summative Assessment. *OECD Working Paper No. 58*. http://www.oecd-ilibrary.org/education/integrating-formative-and-summative-assessment_5kghx3kbl734-en.
- Mandíková, D. (2009). Postoje žáků k přírodním vědám – výsledky výzkumu PISA 2006. *Pedagogika*, 59(4), 380-394.
- Mandíková, D.; Palečková, J.; Tomášek, V. (1996). *Praktické úlohy TIMSS*. Praha: VÚP.
- Martin, M.O.; Mullis, I. V.S.; Beaton, A. E.; Gonzalez, E. J.; Smith, T. A.; Kelly, D. L. (1998). *The Science Achievement in the Primary School Years*. Boston: TIMSS International Study Center.
- Martin, M.O.; Mullis, I.V.S.; Foy, P. (with Olson, J.F., Erberber, E., Preuschoff, C., & Galia, J.). (2008). *TIMSS 2007 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Boston: TIMSS & PIRLS International Study Center.
- Martin, M.O.; Mullis I.V.S. (Eds.). (2012). *Methods and Procedures in TIMSS and PIRLS 2011*. Boston: TIMSS & PIRLS International Study Center. <http://timss.bc.edu/methods/>.
- Martin, M.O.; Mullis, I.V.S.; Foy, P.; Stanco, G.M. (2012). *TIMSS 2011 International Results in Science*. Boston: TIMSS & PIRLS International Study Center.
- Martin, M.O.; Mullis, I.V.S. (Eds.). (2013). *TIMSS and PIRLS 2011: Relationships Among Reading, Mathematics, and Science Achievement at the Fourth Grade—Implications for Early Learning*. Boston: TIMSS & PIRLS International Study Center.
- Martin, M. O.; Mullis, I.V.S.; Hooper, M. (Eds.). (2016). *Methods and procedures in TIMSS 2015*. Boston: TIMSS & PIRLS International Study Center.
- Masters, G.N.; Forster, M. (2000). *The Assessment We Need*. Camberwell: Australian Council for Educational Research.
- Matějů, P.; Straková, J. (2003). Role rodiny a školy v reprodukci vzdělanostních nerovností. Sociologický pohled na úlohu víceletých gymnázií ve světle výzkumu PISA 2000. *Sociologický časopis / Czech Sociological Review*, 5/2003, 625–652.
- Matějů, P.; Straková, J. (Eds.) (2006). *(Ne)rovné šance na vzdělání. Vzdělanostní nerovnosti v České republice*. Praha: Academia.
- Matějů, P.; Smith, M.; Soukup, P.; Basl, J. (2007). Determination of College Expectations in OECD Countries: The Role of Individual and Structural Factors. *Sociologický časopis / Czech Sociological Review*, 6/2007, 1121–1148.
- Matějů, P.; Smith, M.; Basl, J. (2008). Rozdílné mechanismy – stejné nerovnosti. Změny v determinaci vzdělanostních aspirací mezi roky 1989 a 2003. *Sociologický časopis / Czech Sociological Review*, 2/2008, 371–400.
- Matějů, P.; Straková, J.; Veselý, A. (Eds.) (2010). *Nerovnosti ve vzdělávání. Od měření k řešení*. Praha: SLON.
- Matějů, P.; Simonová, N. (2013). Koho znevýhodňuje škola: chlapce, nebo dívky? Rozdíly v dovednostech, školních výsledcích a vzdělanostních aspiracích dívek a chlapců devátých tříd základních škol, *Orbis scholae*, 7 (3), 107–138.
- McGuigan, L., & Hoy, W. K. (2006). Principal Leadership: Creating a Culture of Academic Optimism to Improve Achievement for All Students. *Leadership and Policy in Schools*, 5(3), s.203-229. doi: 10.1080/15700760600805816.

- Messick, S.; Beaton, A.; Lord, F. (1983). National Assessment of Educational Progress reconsidered: A new design for a new era (*NAEP Report 83-1*). Princeton, NJ: National Assessment of Educational Progress.
- Mislevy, R. J.; Beaton, A.; Kaplan, B.A.; Sheehan, K. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29(2), 133-161.
- McKinsey & Company. (2007). *How the World's Best-Performing School Systems Have Come Out on Top*. McKinsey & Company.
- McKinsey & Company. (2010). *How the World's Most Improved School Systems Keep Getting Better*. McKinsey & Company.
- Morris, A. (2011) Student Standardized Testing: Current Practices in OECD Countries and a Literature Review. *OECD Education Working Papers*, No. 65.
- Mullis, I.V.S.; Martin, M.O.; Beaton, A.E.; Gonzalez, E. J.; Kelly, D. L.; Smith, T.A. (1998). *The Mathematics Achievement in the Primary School Years*. Boston: TIMSS International Study Center.
- Mullis, I.V.S.; Martin, M.O.; Foy, P. (with Olson, J.F., Preuschoff, C., Erberber, E., Arora, A., Galia, J.). (2008). *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Boston: TIMSS & PIRLS International Study Center.
- Mullis, I.V.S.; Martin, O.M.; Kennedy, A.M.; Trong, K.L.; Sainsbury, M. (2009). *PIRLS 2011 Assessment Framework. TIMSS 2011 Assessment Frameworks*. Boston: TIMSS & PIRLS International Study Center. http://timssandpirls.bc.edu/pirls2011/downloads/PIRLS2011_Framework.pdf.
- Mullis, I.V.S.; Martin, M.O.; Minnich, C.A.; Stanco, G.M.; Arora, A.; Centurino, V.A.S.; Castle, C.E. (Eds) (2012a). *TIMSS 2011 Encyclopedia: Education Policy and Curriculum in Mathematics and Science, Volumes 1 and 2*. Boston: TIMSS & PIRLS International Study Center.
- Mullis, I.V.S.; Martin, M.O.; Foy, P.; Drucker, K.T. (2012b). *PIRLS 2011 International Results in Reading*. Boston: TIMSS & PIRLS International Study Center.
- Mullis, I.V.S.; Martin, M.O.; Foy, P.; Arora, A. (2012c). *TIMSS 2011 International Results in Mathematics*. Boston: TIMSS & PIRLS International Study Center.
- Mullis, I.V.S.; Martin, O.M. (Eds.) (2013). *Timss 2015 Assessment Frameworks*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA).
- Müller, J.C.; Kretzschmar, A.; Wüstenberg, S.; Greiff, S. (2014). Extending the Assessment of Complex Problem Solving to Finite State Automata: Embracing Heterogeneity. *Paper presented at the 12th Szeged Workshop on Educational Evaluation, Szeged, April 2014*.
- Münich, D.; Protivínský, T. (2013). *Dopad vzdělanosti na hospodářský růst: ve světle nových výsledků PISA 2012*. Praha: IDEA. http://idea.cerge-ei.cz/files/IDEA_Studie_10_2013.pdf.
- Münich, D.; Federičová, M. (2015). Srovnání žákovské oblíbenosti školy a matematiky pohledem mezinárodních šetření. *Pedagogická orientace*, 2015, 25, 4, 557-582.
- Najvar, P.; Najvarová, V.; Janík, T.; Šebestová, S. (2011). *Videostudie v pedagogickém výzkumu*. Brno: Paido.
- Najvarová, V. (2008). Čtenářská gramotnost žáků 1. stupně základní školy. *Pedagogická orientace*, 18(1), 5-19.
- NEA. (2010). Global Competence Is a 21st Century Imperative. An IEA Policy Brief www.nea.org/assets/docs/HE/PB28A_Global_Competence11.pdf.

- Nevo, D. (1995). *School-Based Evaluation*. Great Yarmouth: Elsevier Science.
- Nilsen, T.; Gustafsson, J.-E. (2014). School emphasis on academic success: exploring changes in science performance in Norway between 2007 and 2011 employing two-level SEM. *Educational Research and Evaluation*, 20(4), s.308-327. doi:<http://dx.doi.org/10.1080/13803611.2014.941371>.
- Nussbaum, M.C. (2011). *Creating Capabilities. Human Development Approach*. Cambridge (MA): The Belknap Press of Harvard University Press.
- OECD (1997). *Prepared for life*. Paris: OECD.
- OECD (2000). *Measuring student knowledge and skills: The PISA 2000 assessment of reading, mathematical, and scientific literacy*. Paris: OECD.
- OECD (2000). *Literacy in the Information Age*. Final Report from the Adult Literacy Survey. Paris: OECD.
- OECD (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris: OECD.
- OECD (2002). *Strategy paper for Education Comitee and CERI governing board. DeSeCo. Theoretical and Conceptual Foundation*. Paris: OECD.
- OECD (2003). *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*. Paris: OECD.
- OECD(2004a). *Learning for Tomorrow's World: First Results from PISA 2003*. Paris: OECD.
- OECD (2004b). *What makes school systems perform? Seeing school systems thgough the prism of PISA*. Paris: OECD.
- OECD (2007). *Science Competencies for Tomorrow's World. First Results from PISA 2006*. Paris: OECD.
- OECD (2010a). *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I)*. Paris: OECD.
- OECD (2010b). *Pathways to Success. How Knowledge and Skills at Age 15 Shape Future Lives in Canada*. Paris: OECD.
- OECD (2012). *Literacy, Numeracy and Problem Solving in Technology-Rich Environments. Framework for the OECD Survey of Adult Skills*. Paris: OECD.
- OECD (2013a). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. Paris: OECD.
- OECD (2013b). *What Students Know and Can Do: Student Performance in Mathematics, Reading and Science*. Paris: OECD.
- OECD (2013c). *2013 OECD Skills Outlook 2013*. First Results form the Surevy of Adult Skills. Paris: OECD. <http://dx.doi.org/10.1787/9789264190511-en>.
- OECD (2013d). *PISA 2012 Results: What Makes Schools Successful (Volume IV): Resources, Policies and Practices*. Paris: OECD Publishing.
- OECD (2013e). *PISA 2015 Draft Collaborative Problem Solving Framework*. <https://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf>.
- OECD (2014a). *PISA 2012 Technical Report*. <http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>.
- OECD (2014b). *PISA 2015. Technical Standards*. <http://www.oecd.org/pisa/pisaproducts/PISA-2015-Technical-Standards.pdf>.
- OECD (2014c). *PISA 2015 Draft Questionnaire Framework*. Paris: OECD.
- OECD (2014d). *Response to Points Raised in Heinz-Dieter Meyer 'OPEN LETTER'*. Paris: OECD.
- Olson, J. F.; Martin, M.O.; Mullis, I.V.S. (Eds.). (2008). *TIMSS 2007 Technical Report*. Boston: TIMSS & PIRLS International Study Center. <http://timss.bc.edu/timss2007/techreport.html>.

- Palečková, J.; Mandíková, D. (1997). *Třetí mezinárodní výzkum matematického a přírodovědného vzdělání. Výsledky českých žáků v praktických úlohách*. Praha: VÚP.
- Palečková a kol. (2007). *Hlavní zjištění výzkumu PISA 2006. Poradí si žáci s přírodními vědami?* Praha: ÚIV.
- Palečková, J.; Tomášek, V. (2013). *Hlavní zjištění PISA 2012. Matematická gramotnost patnáctiletých žáků*. Praha: ČŠI.
- Pelgrum, W. J.; Anderson, R.E. (2001). *ICT and the Emerging Paradigm*. Twente: University of Twente OCTO.
- Plomp, T.; Anderson, R.E.; Law, N.; Quale, A. (Eds.). (2003). *Cross-national information and communication technology policies and practices in education*. Greenwich, CT: Information Age Publishing.
- Potužníková, E.; Straková, J. (2006). Rozdíly ve vědomostech a dovednostech českých chlapců a děvčat na základě zjištění mezinárodních výzkumů. *Sociologický časopis / Czech Sociological Review* 42 (4), 701–718.
- Potužníková, E.; Lokajíčková, V.; V., Janík, T. (2014). Mezinárodní výzkumy školního vzdělání v České republice: zs.jištění a výzvy. *Pedagogická orientace*, 24(2), 185–221.
- Prezel, M. (2012). Empirický výzkum vzdělávání: vystačí naše stávající přístupy zítra? *Pedagogická orientace*, 2012, 22, 4, 475–489.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Reimers, F., M. (2013). *Assesing Global Education: an Opportunity for the OECD*. www.oecd.org/pisa/pisaproducts/Global-Competency.pdf.
- Rendl, M.; Vondrová, N. (2014). Kritická místa v matematice u českých žáků na základě výsledků šetření TIMSS 2007. *Pedagogická orientace*, 24(1), 22–57. DOI: <http://dx.doi.org/10.5817/PedOr2014-1-22>
- Ritzen, J. (2013). International Large-Scale Assessments as Change Agents. 13-24 in von Davier, M., Gonzales, E., Kirch, I., Yamamoto, K. (Eds.) *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research*. Dordrecht: Springer.
- Robitaille, D., F. (1995). Curriculum frameworks for Mathematics and Science. *TIMSS Monograph No.3*. Vancouver: Pacific Educational Press.
- Roth, K. a kol. (2006). *Teaching Science in Five Countries: Results from the TIMSS 1999 Video Study*. Washington: U.S. Department of Education, National Center for Education Statistics.
- Rychen D.S.; Salganik L.H. (Eds.). (2001). *Defining and Selecting Key Competencies*. Göttingen: Hogrefe & Huber Publishers.
- Řeháková, B. (2000). Nebojte se logistické regrese. *Sociologický časopis/Czech Sociological Review*, 36(4), 475–492.
- Santiago, P.; Gilmore, A.; Nusche, D.; Sammons, P. (2012). *OECD Reviews of Evaluation and Assessment in Education: Czech Republic*. OECD Publishing. Paris: OECD. Available online at: <http://www.oecd.org/dataoecd/33/47/49479976.pdf>.
- Schlotter, M.; Schwerdt, G.; Woessmann, L. (2014). Econometric Methods for Causal Evaluation of Education Policies and Practices: A Non-Technical Guide. 95-128 in Strietholt, R., Bos, W. Gustafsson, J.E., Rosén, M. (Eds.) *Educational Policy Evaluation through International Comparative Assessments*. Göttingen: Waxmann.
- Schmidt, W. H.; McKnight, C. C.; Valverde, G. A.; Houang, R. T., & Wiley, D. E. (Eds.) (1997a). *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in Mathematics – Volume 1*. Dordrecht: Kluwer Academic Publishers.

- Schmidt, W.H.; Raizen, S.; Britton, E.D.; Bianchi, L.J.; Wolfe, R. (Eds.) (1997b). *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Science - Volume 2*. Dordrecht: Kluwer Academic Publishers.
- Simonová, N.; Soukup, P. (2010) Působení primárních a sekundárních faktorů sociálního původu při přechodu na vysokou školu v ČR: výsledky výzkumu PISA-L. 298-321 in Matějů, P., Straková, J., Veselý, A. (Eds.) *Nerovnosti ve vzdělávání. Od měření k řešení*. Praha: SLON.
- Soukup, P. (2006). Proč užívat hierarchické lineární modely? *Sociologický časopis/Czech Sociological Review*, 42(5), 987-1012.
- Soukup, P. (Ed.) (2010). *Národní zpráva z Mezinárodní studie občanské výchovy*. Praha: ÚIV.
- Soukup, P. (2016). Možnosti praktické práce s daty z mezinárodních vzdělávacích studií: problémy a jejich praktická řešení. *Orbis Scholae* 1/2016.
- Starý, K.; Laufková, V. (2015). Kde žáci v polovině základního vzdělávání selhávají při porozumění čtenému polemickému textu. *Pedagogika* 4/2015, 392-412.
- Straková, J.; Tomášek, V.; Palečková, J. (1998). *Třetí mezinárodní výzkum matematického a přírodovědného vzdělání - Souhrnné výsledky žáků posledních ročníků středních škol*. Praha: VÚP.
- Straková, J. (2008). *Rozvíjení a hodnocení klíčových kompetencí*. Disertační práce. MU v Brně.
- Straková, J. (2009). Vzdělávací politika a mezinárodní výzkumy výsledků vzdělávání v ČR. *Orbis Scholae*, 2009(3), 103-118.
- Straková, J. (2010a). Přidaná hodnota studia na víceletých gymnáziích ve světle dostupných datových zdrojů. *Sociologický časopis / Czech Sociological Review* 46 (2): 187-210.
- Straková, J. (2010b). Pedagogické činnosti českých učitelů v mezinárodním srovnání. *Pedagogika* 60 (3-4), 276-291.
- Straková, J. (2011). Vzdělanostní nerovnosti a vzdělávací politika ve světě a v ČR. 55-75 in Kasíková, H., Straková, J. (Eds.) *Diverzita a diferenciacie v základním vzdělání*. Praha: Carolinum.
- Straková, J. (2013). Jak dál s kurikulární reformou. *Pedagogická orientace*, 23(5), 734-743. DOI: <http://dx.doi.org/10.5817/PedOr2013-5-734>.
- Straková, J.; Tomášek, V. (2013). Měření vědomostí a dovedností romských žáků v rámci šetření PISA. *Pedagogika*, 1/2013, 41-53.
- Straková, J.; Spilková, V.; Simonová, J.; Fiendaenderová, H.; Hanzák, T. (2013). Názory učitelů základních škol na potřebu změn ve školním vzdělávání. *Orbis Scholae*, 7 (1), 79-100.
- Straková, J.; Greger, D. (2013). Faktory ovlivňující přechod žáků 5. ročníků na osmileté gymnázium. *Orbis Scholae* 7 (3), 73-85.
- Straková, J.; Simonová, J. (2013). Vliv navštěvované střední školy na občanské postoje středoškoláků v ČR. *Orbis Scholae* 7 (3), 27-47.
- Straková, J.; Veselý, A. (2013) *Předpoklady úspěchu v práci a v životě: výsledky mezinárodního výzkumu dospělých OECD PIAAC*. Praha: Dům zahraniční spolupráce.
- Strietholt, R.; Gustafsson, J.E.; Rosén, M.; Bos, W. (2014) Outcomes and Causal Inference in International Comparative Assessments. 9-18 in Strietholt, R., Bos, W. Gustafsson, J.E., Rosén, M. (Eds.) *Educational Policy Evaluation through International Comparative Assessments*. Göttingen: Waxmann.
- Strietholt, R. (2014) What Does Justice Mean in Education? 51-58 in Strietholt, R., Bos, W. Gustafsson, J.E., Rosén, M. (Eds.) *Educational Policy Evaluation through International Comparative Assessments*. Göttingen: Waxmann.
- Štech, S. (2011) *Pisa - nástroj vzdělávací politiky nebo výzkumná metoda?* *Orbis Scholae* 011/1, 123-133.
- Štech, S. (2015) Proč se kritizuje PISA? *Pedagogická orientace*. 25, 4, 605-612.

- Thompson, S., C.J. Johnstone, and M.L. Thurlow. 2002 *Universal design applied to large scale assessments*. <http://www.cehd.umn.edu/NCEO/onlinepubs/synthesis44>.
- Thompson, S.; Hillman, K. (2010). *Against the odds: influences on the post-school success of low performers*. Available at: http://works.bepress.com/kylie_hillman/8/
- Tomášek, V. a kol. (2012). *Národní zpráva TIMSS 2011*. Praha: ČŠI.
- Torney-Purta, J.; Lehmann, R.; Oswald, H., & Schulz, W. (2001). *Citizenship and Education in Twenty-Eight Countries*. Amsterdam: IEA.
- Tourney-Purta, J.; Amadeo, J. (2013). The Contributions of International Large-Scale Studies in Civic Education and Engagement. 87-114 in von Davier, M., Gonzalez, E., Kirsch, E., Yamamoto, K. (Eds). *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research*. Dordrecht: Springer.
- Urbánek, T.; Denglerová, D.; Širůček, J. (2011). *Psychometrika. Měření v psychologii*. Praha: Portál.
- Van Avermaet, P.; Sierens, S. (2012). From the periphery to the core. Dealing with diversity in education. In: Timmerman, C., N. Clycq, Segaert, B. (Eds.). *Cultural transfer and education in a multicultural context*. Gent: Academia Press.
- Veselý, A. (2012). Institucionalizace neodpovědnosti, efektivity, nebo konformity? Reformy organizací veřejných služeb v teorii akontability. *Sociologický časopis / Czech Sociological Review*, 48, 4, 757-784.
- Von Davier, M.; Sinharay, S.; Oranje, A.; Beaton, A. (2006). Statistical Procedures used in the National Assessment of Educational Progress (NAEP): Recent developments and future directions. In *Handbook of statistics*, eds. Rao C. R., and Sinharay S., (Vol. 26): *Psychometrics*. Amsterdam: Elsevier.
- Voňková, H.; Hrabák, J. (2015). Řešení nesrovnatelnosti hodnocení ICT znalostí a dovedností žáků skrze ukotvující viněty. *Pedagogika*, 3/2015, 274-291.
- Weinert, F., E. Concepts of Competence. (2001). In: Rychen D.S. & Salganik L.H. (Eds.) *Defining and Selecting Key Competencies*. Göttingen: Hogrefe & Huber Publishers.
- Wendt, H.; Stubbe, T.C.; Schwippert, K.; Bos, W. (Eds.) (2015). *10 Jahre Internationalvergleichende Schulleistungsforschung in der Grundschule. Vertiefende Analysen zu IGLU und TIMSS 2001 bis 2011*. Münster: Waxmann.
- Whelan, F. (2009) *Lessons Learned: How Good Policies Produce Better Schools*. Fenton Whelan.
- Woolfolk Hoy, A.; Hoy, W.A.; Kurz, N.M. (2008) Teacher's academic optimism: The development and test of a new construct. *Teaching and Teacher Education*, 24, 821-835.
- Wu, M. (2009) *A Critical Comparison of the Contents of PISA and TIMSS Mathematics Assessments*. https://edsurveys.rti.org/PISA/documents/WuA_Critical_Comparison_of_the_Contents_of_PISA_and_TIMSS_psg_WU_06.1.pdf.
- Zhao, Y. (2014) *Who's Afraid of the Big Bad Dragon: Why China has the Best (and Worst)*. San Francisco: Jossey-Bass. A Wiley Brand.

Seznam zkratek

ACER	Australian Council for Educational Research
ALL	Adult Literacy and Lifeskills Survey
BPC	Board of Participating Countries
CERI	Centre for Educational Research and Innovation
CITO	Netherlands National Institute for Educational Measurement
CIVED	Civic Education Study
DeSeCo	Defining and Selecting Key Competencies
EaG	Education at a Glance
EDPC	Education Policy Committee
ETS	Education Testing Service
IALS	International Adult Literacy Survey
ICCS	International Civic and Citizenship Education Study
ICILS	International Computer and Information Literacy Study
IEA	International Association for the Evaluation of Educational Achievement
IEA DPC	IEA Data Processing Center
ISCO	International Standard Classification of Occupations
NIER	National Institute for Educational Research in Japan
OECD	Organisation for Economic Co-operation and Development
PIAAC	Programme for the International Assessment of Adult Competencies
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for the International Student Assessment
SITES	Second Information on Technology in Education Study
TALIS	Teaching and Learning International Survey
TIMSS	Trends in International Mathematics and Science Study

Příloha 1: Přehled realizace mezinárodních výzkumů výsledků vzdělávání žáků v ČR

rok	výzkum		populace	školy	žáci
1995	TIMSS	matematika, přírodní vědy	3. a 4. ročník, 7. a 8. ročník, poslední ročník SŠ	500	16 700
1995	RLS	čtenářská gramotnost	3. a 8. ročník	130	5 520
1999	TIMSS	matematika, přírodní vědy	8. ročník	150	3 600
1999	CIVED	výchova k občanství	8. ročník, 3. ročník SŠ	300	6 900
2000	PISA	matematická, přírodovědná a čtenářská gramotnost	15letí, 3. ročník SŠ	253	9 400
2003	PISA	matematická, přírodovědná a čtenářská gramotnost, řešení problémů	15letí, 9. ročník	260	9 919
2006	PISA	matematická, přírodovědná a čtenářská gramotnost	15letí, 9. ročník	245	9 016
2007	TIMSS	matematika, přírodní vědy	4. ročník, 8. ročník	291	9 000
2009	PISA	matematická, přírodovědná a čtenářská gramotnost	15letí	290	7 500
2009	ICCS	výchova k občanství	8. ročník	144	4 600
2011	TIMSS a PIRLS	matematika, přírodní vědy a čtenářská gramotnost	4. ročník	177	4 500
2012	PISA	matematická, přírodovědná a čtenářská gramotnost, řešení problémů, finanční gramotnost	15letí	297	6 535
2013	ICILS	informační a počítačová gramotnost	8. ročník	170	3 200
2015	PISA	matematická, přírodovědná a čtenářská gramotnost, řešení problémů ve spolupráci	15letí	345	6 900
2015	TIMSS	matematika, přírodní vědy	4. ročník	159	5 200

Příloha 2: Ilustrativní úlohy

V průběhu realizace výzkumů bylo publikováno mnoho souborů úloh s návody na jejich hodnocení, klasifikací prostřednictvím konceptuálních rámců a informací o národní a mezinárodní úspěšnosti. Soubory úloh je možno vyhledat na webových stránkách České školní inspekce, kde jsou zpřístupněny v takové podobě, že je možno je přímo využít pro práci ve třídě. Tato příloha nechce dublovat informace z publikací úloh. Přináší několik úloh, které ilustrují specifika mezinárodních výzkumů a diskusí, které probíhají nad úlohami v mezinárodních pracovních skupinách rozhodujících o použití úloh.

Jako první je uvedena úloha z matematiky, která byla vytvořena v rámci výzkumu PISA, aby ověřila schopnost žáků využívat v diskusi matematické argumenty. Do výzkumu nakonec nebyla zařazena z výchovných důvodů. Zástupci zúčastněných zemí se obávali, že k žákům vysílá signál, že argumenty je možno podržet účelu a že to tak v politickém prostředí běžně chodí. Úloha však zároveň dobře ilustruje důležitý aspekt matematické gramotnosti, který je ve výzkumu PISA ověřován.

Zbrojení

(zdroj PISA 2000)

Rozpočet na obranu určité země činil v r. 1980 30 milionů dolarů. Celkový rozpočet v tom samém roce byl přitom 500 milionů dolarů. V následujícím roce činil rozpočet na obranu 35 milionů dolarů, zatímco celkový rozpočet 605 milionů dolarů. Inflace v těchto dvou letech tvořila 10 procent.

1. *Jsi pozván/a, abys přednášel/a pacifistické společnosti. Tvým záměrem je vysvětlit, že rozpočet na obranu země v tomto období klesl. Vysvětli, jak bys to udělal/a.*
2. *Jsi pozván/a, abys přednášel/a na vojenské akademii. Tvým záměrem je vysvětlit, že rozpočet na obranu země v tomto období vzrostl. Vysvětli, jak bys to udělal/a.*

Další úloha, která nebyla zařazena z výchovných důvodů, byla vyvinuta, aby hodnotila přírodovědnou gramotnost ve výzkumu PISA. Úloha není nikterak inovativní z hlediska ověřovaných dovedností, neboť ověřuje znalost přírodovědných fakt. Z výzkumu byla vyřazena kvůli použitému tématu. Zástupci zúčastněných zemí se obávali, že by motivovala děvčata k aplikaci „čokoládové diety“ a tak potenciálně ohrožovala jejich zdraví.

Čokoláda

(zdroj PISA 2000)

Přečti si shrnutí jednoho novinového článku z novin The Daily Mail ze 30. března 1998.

Jeden novinový článek přinesl příběh 22leté studentky Jessiky, která drží „čokoládovou dietu“. Tvrdí, že je stále zdravá a udržuje si stabilní váhu 50 kg, přičemž sní 90 tabulek čokolády týdně a zcela vyloučila veškeré jiné potraviny kromě jednoho „řádného jídla“ každý pátý den. Jeden

odborník na výživu to komentoval slovy: „Jsem překvapen, že někdo může s takovou dietou žít. Tuky jí dávají energii k životu, ale dieta není vyvážená. V čokoládě jsou nějaké minerály a výživné látky, ale nedostává dostatek vitamínů. Později by se mohla dostat do vážných zdravotních problémů.”

Otázka 1:

V knize obsahující nutriční hodnoty potravin jsou o čokoládě uvedeny následující údaje. Předpokládej, že všechny tyto údaje platí pro druh čokolády, kterou Jessica stále jí. Předpokládej také, že tabulky čokolády, kterou jí, váží každá 100 gramů.

Tabulka 1: Výživné hodnoty 100 g čokolády

Bílkoviny (g)	Tuky (g)	Sacharidy (g)	Minerály		Vitamíny			Energetická hodnota (kJ)
			Vápník (mg)	Železo (mg)	A	B (mg)	C	
5	32	51	50	4	-	0.20	-	2142

Podle uvedené tabulky je ve 100 g čokolády obsaženo 32 g tuku a 2142 kJ energie. Odborník na výživu řekl: „Tuky jí dávají energii k životu ...”

Když sní někdo 100 g čokolády, pochází veškerá energie (2142 kJ), kterou takto získá, z 32 g tuku?

Na základě údajů z tabulky svou odpověď vysvětli.

Otázka 2:

Odborníci na výživu řekli, že Jessica „... nedostává ani zdaleka dost vitamínů”. Jedním z vitamínů, který v čokoládě není, je vitamín C. Jessica by snad mohla nahradit nedostatek vitamínu C tím, že by jedla potraviny bohaté na vitamín C v rámci svého „jednoho řádného jídla jednou za pět dní”.

Zde je seznam druhů jídel.

1. Ryby.
2. Ovoce.
3. Rýže.
4. Zelenina.

Které dva druhy jídel z tohoto seznamu bys doporučil(a) Jessice, aby si mohla doplnit nedostatek vitamínu C?

- A 1 a 2.
 B 1 a 3.
 C 1 a 4.
 D 2 a 3.
 E 2 a 4.
 F 3 a 4.

Následující dvě úlohy dobře ilustrují charakteristické aspekty přírodovědné gramotnosti. První úloha, která byla vyvinuta v rámci výzkumu TIMSS 1995, zjišťuje porozumění přírodovědným poznatkům. Úloha nebyla ve výzkumu dlouhodobě využívána, neboť se ukázalo, že ověřuje nějaké jiné dovednosti, než jsou testovány ostatními úlohami v rámci výzkumu TIMSS. Tedy přesto, že se odborníci z mezinárodní skupiny pro hodnocení přírodovědné oblasti shodovali v tom, že úloha je velmi podařená, nezapadala do ostatních úloh (vykazovala odlišné psychometrické vlastnosti) a z toho důvodu musela být z testu vyloučena.

Úloha také dobře ilustruje rozdíly mezi úlohami výzkumu asociace IEA a výzkumu PISA. Zatímco úlohy PISA jsou koncipovány kolem nějakého kontextu ve větších celcích, což je činí čtenářsky náročnější a vzájemně na sobě závislé, úlohy TIMSS se vyskytují izolovaně a jsou uvozeny pouze krátkými otázkami.

Židle

(zdroj TIMSS 1995)

Co by zůstalo ze židle, kdybyste z ní odstranili všechny atomy?

- a) Židle by zůstala, ale měla by menší hmotnost.
- b) Židle by zůstala přesně stejná jako před tím.
- c) Ze židle by nezůstalo nic.
- d) Na podlaze by zůstala jen louže kapaliny.

Další přírodovědná úloha byla vytvořena v rámci výzkumu PISA 2000 a ověřuje dovednost vytvářet hypotézy a tyto hypotézy ověřovat. V tomto ohledu ji autorka této publikace považuje za inspirativní pro českou školu, ve které jsou tyto dovednosti nedostatečně kultivovány.

Mouchy

(zdroj PISA 2000)

Farmář choval dojnice na pokusné farmě. Ve chlévě u dobytka bylo tolik much, že to ohrožovalo zdraví zvířat. Farmář proto provedl postřik chléva i dobytka roztokem insekticidu A (prostředku na hubení hmyzu A). Téměř všechny mouchy tak vyhubil. Za nějakou dobu se však mouchy zase rozmnožily. Farmář znovu provedl postřik insekticidem. Výsledek byl obdobný jako při prvním postřiku. Uhynula většina much, ale ne všechny. Brzy se mouchy znovu rozmnožily a farmář je znovu postříkal insekticidem. Když se totéž opakovalo popáté, bylo jasné, že insekticid A účinkuje na mouchy čím dál tím méně.

Farmář si uvědomil, že napoprvé si připravil velké množství roztoku insekticidu a s tím pak prováděl všechny postřiky. Z toho usoudil, že se možná roztok insekticidu časem rozložil.

1. *Farmář se domníval, že se insekticid časem rozložil. Stručně vysvětlete, jak by se tento předpoklad dal ověřit.*
2. *Farmář se domníval, že se insekticid časem rozložil. Navrhněte dvě jiná vysvětlení, proč insekticid A „účinkoval na mouchy čím dál tím méně“.*

Poslední zde uvedená úloha byla vytvořena v rámci výzkumu PISA 2000 a nebyla dlouhodobě využívána, protože byla pro žáky příliš obtížnou. Velmi hezky však ilustruje ověřování jedné ze tří složek konceptuálního rámce, kterou tvoří interpretace předloženého textu.

Kdyby

(upraveno dle PISA 2000)

Níže uvedený materiál se objevil v jednom časopise jako reklama na motocykly Harley-Davidson. S pomocí této reklamy odpověz na následující otázky.

Otázka 1: Kdyby autor mohl žít svůj život znovu, tak by

- A žil déle a víc pracoval.
- B zkusil být opatrnější a vnímavější než dříve.
- C si užíval přítomnosti místo toho, aby ji nechal jen tak ubíhat.
- D se zajímal víc o ostatní než o sebe.

Otázka 2: Souhlasíš s tímto poselstvím básně o životě? Vysvětli vlastními slovy, proč ano, nebo proč ne.

Otázka 3: V informaci se objevují slova: „Nebo také můžete počkat do zítřka.“ Vypadá to, jako by tato slova byla v rozporu s básní. Proč jsou v reklamě uvedena?

Kdyby

Kdybych měl žít svůj život znovu, zkusil bych pro příště dělat víc chyb.

Uvolnil bych se.

Dělal bych více hloupostí, než jsem dělal na této cestě.

Vím, že bych bral vážně jenom pár věcí.

Více bych se chytal šancí.

Víc bych cestoval.

Vylezl bych na více hor, plaval bych ve více řekách a pozoroval bych více západů slunce.

Jedl bych více zmrzliny a méně fazolí.

Mé problémy by byly opravdovější a netrápil bych se vymyšlenými.

*Víš, jsem jeden z těch, kteří žijí opatrně a rozumně a klidně,
hodinu za hodinou, den za dnem.*

Ach, i já jsem měl své okamžiky. Kdybych to měl všechno prožít znovu, měl bych jich víc.

Vlastně bych nic jiného nezkoušel. Prostě bych žil.

*Kdybych měl prožít znovu svůj život,
časně z jara bych se zul a byl bych tak až pozdě do podzimu.*

Víc bych chodil za školu.

Víc bych jezdil na kolotoči.

Natrhal bych více sedmikrásek.



Od £ 4 995

Volejte VB 0345 883 1340, Francie 0803 83 1340, Německo/Rakousko 0049 61 05/284 284
Benelux 0031 5820409



a všechny ostatní státy 0044 1753 829 500. Nebo také můžete počkat do zítřka.
<http://www.harley-davidson.com>

Summary

International Large-Scale Assessment Surveys: methodology, benefits, risks, and opportunities

International Large-Scale Assessment Surveys (ILSAS) started in 1958 when a group of scholars, educational psychologists, sociologists, and psychometricians met at the UNESCO Institute for Education in Hamburg to discuss the problems of school and student evaluation. They argued that effective evaluation requires examination of both the inputs into education and its outcomes (such as knowledge, attitudes, and participation). They viewed the world as a natural educational laboratory, where different school systems experiment in different ways to obtain optimal results in the education of their young people. They assumed that if research could obtain evidence from across a wide range of systems, the variability would be sufficient to reveal important relationships which would otherwise escape detection within a single education system. They strongly rejected data-free assertions about the relative merits of various education systems, and aimed to identify factors that would have meaningful and consistent influences on educational outcomes.

In 1967, they founded the International Association for the Evaluation of Educational Achievement, which since then has been organizing international assessment surveys in numerous educational areas. Its best-known surveys are the Trends in International Mathematics and Science Study (TIMSS), which assesses knowledge, skills, and attitudes in mathematics and science in students in grades 4 and 8 every four years, and the Progress in Reading Literacy Study (PIRLS), which assesses the reading skills and habits of students in grade 4 every five years. Since 1967, the number of countries participating in the studies of the International Association for the Evaluation of Educational Achievement has increased from 12 to more than 70.

In 2000, the Organization for Economic Cooperation and Development (OECD) launched its own study of mathematics, science, and reading skills, called Programme for International Student Assessment (PISA). IEA studies are declared to be based on school knowledge. It means that they aim at measuring the extent to which students acquired skills and knowledge taught at school. PISA, on the other hand, declares that it does not measure what is taught at school, but rather whether students acquired knowledge, skills, and attitudes needed for success in their personal and social life and on the labour market. While the conceptual frameworks of IEA studies are designed as an intersection of the curricula of participating countries, the PISA frameworks are developed by renowned experts in particular areas on the basis of their perception of the importance of particular competencies in real-life situations. PISA originated in the belief of policy makers from OECD bodies that schools lag behind the rapid development of society. It wanted to bring the attention of educational leaders to skills, knowledge, and attitudes that are important in modern society. OECD PISA assesses students at the end of their compulsory schooling and takes place every three years. Although it was developed for OECD countries, it is also open to non-OECD education systems and the number of participating countries is similar to the number of participating countries in IEA studies.

In addition to assessment surveys in school populations, the OECD also piloted the assessment of higher education students in a study called Assessment of Learning Outcomes in Higher Education (AHELO) and launched a cycle of surveys of adult skills called the Programme

for the International Assessment of Adult Competencies (PIAAC). The survey of adult skills is planned to take place every 10 years.

ILSAS involve the best international experts in educational measurement, psychometrics, and statistics. The studies influence the developments in the area of educational measurement and also in educational effectiveness research. The methodology of the studies and information about their development can serve as an invaluable learning resource. At the same time, knowledge about the methodology is important for correct work with the data and for its full exploitation.

The innovations in the frameworks of domains assessed and also new domains that are developed for the new data collections well demonstrate the changes in the perception of the importance of various educational outcomes. The transfer to electronic administration allows new competencies to be assessed and new item formats to be used. ILSAS reflect the growing emphasis on non-cognitive outcomes and try to find ways to assess them.

The methodology of the studies is very complex. All the procedures, including sampling (two-level PPS sampling), the translation and adaptation of the instruments and administration and coding are highly sophisticated and standardized. Cognitive scores are estimated by IRT analysis combined with multiple imputations. The specific features of the ways in which the data was collected and scaled have to be taken into account in the data analysis. A user has to use weights properly, make provision for cluster sampling (use replicate weights), and work correctly with plausible values if he/she wants to avoid incorrect conclusions. Data cannot be properly analyzed in commonly used statistical packages without using specific modules and procedures. However, many utilities have been developed in recent years to make the analyses easier and make the data accessible to a wider community of researchers and policy makers. The IDB analyzer developed by IEA DPC generates syntaxes for simple analyses in SPSS for data from many international surveys of both IEA and OECD. NCES provided the International Data Explorer, which allows descriptive statistics to be obtained online. New statistical tools are being developed to allow sophisticated analyses of the data from ILSAS.

ILSAS themselves are experiencing further and fast development in many respects. To name a few: new methods have been explored to adjust the difficulty of the assessment to the needs of developing countries; the comparability of paper-and-pencil and electronic tests has been explored and justified; electronic administration has been utilized to introduce adaptive testing; the questionnaires are framed according to the findings of educational effectiveness research in a more systematic manner, which increases the analytical potential of the data that is collected; new analytical methods are being developed to overcome the disadvantage of the cross-sectional character of ILSAS, and the findings, especially of OECD surveys, are being presented in various new ways and in many different formats.

The intention to make the results understandable to a wider audience leads to simplifications that are sometimes somewhat audacious. Policy makers get used to the findings and use them regularly to justify education policy decisions. The impact of ILSAS on education policy has been subject to extensive criticism for a long time. Recently it has mainly been targeted at PISA. PISA is accused of contributing to the escalation of testing, to increased reliance on quantitative measures, and to a shift of attention to short-term fixes designed to help a country climb the rankings quickly. By emphasizing a narrow range of measurable aspects of education, PISA is said to take attention away from the less measurable or immeasurable educational objectives such as physical, moral, civic, and artistic development, thereby dangerously narrowing the collective imagination regarding what education is and ought to be about.

The critics argue that the OECD is naturally biased in favour of the economic role of public schools, but that preparing young men and women for gainful employment is not the only, and not even the main goal of public education, which has to prepare students for participation in democratic self-government, moral action, and a life of personal development, growth, and wellbeing. The OECD is accused of ignoring the imperfections of PISA and confusing policy makers with incorrectly simplified interpretations. This criticism is also very popular in the Czech educational community.

International surveys have provided Czech researchers and policy makers with information that would be impossible to gain otherwise as the Czech Republic does not carry out any national assessment that would allow trends to be monitored. The information about the development of educational outcomes could thus not be obtained otherwise. The international comparative studies are also a unique source of information about educational inequalities in the Czech Republic. The studies show that Czech students (and also adults) have deteriorated in mathematics and improved slightly in reading. Czech students have continuously above-average results in science and a very good level of IT literacy. In the period 2007-2009 the results of Czech students deteriorated, while since then they have been improving slightly. The downturn in 2007-2009 has not been satisfactorily explained so far. Czech students have more negative attitudes towards school in general and towards mathematics and reading than their peers from other countries. This is also true for their civic attitudes.

Educational inequalities in the Czech Republic are relatively high and continuously increasing. The Czech Republic exhibits a strong relationship between student achievement and socio-economic status and big (and growing) differences between schools in terms of student results and the composition of the student body. The transition to higher tracks at both lower and upper secondary levels is strongly influenced by family background; this also remains the case after ability has been controlled for. Teacher questionnaires repeatedly show relatively low levels of trust in students and their parents. Czech teachers also exhibit low levels of belief in their own skills to motivate and successfully educate students who do not have motivation from their families. The level of satisfaction with their profession is also relatively low among Czech teachers. These are all important findings that deserve attention from the educational community.

The concepts of the domains that were assessed were accepted differently by the Czech educational community. The concept of reading literacy, especially as defined in PIRLS, has been regarded as useful by both teachers and researchers. Teachers use released items in their classes and try actively to develop the skills defined in the framework. Researchers use the released items for detailed studies of student understanding. The mathematics framework has not aroused any interest, probably because the Czech Republic has a long tradition of mathematics education and the framework corresponds to the way mathematics is taught in Czech classrooms. Unfortunately, the concept of science has received a similar reception, which is a pity as some of its elements (formulating and testing hypotheses, designing experiments) are badly lacking in Czech science instruction.

ILSAs brought the attention of Czech society to educational inequalities. However, it has not resulted in any attempt to eliminate differentiation in the education system. The belief that students can be educated more effectively in homogeneous classes is deeply embedded in Czech society.

ILSAs were the source of much useful information for both researchers and policy makers. They were also enriched by the addition of two national longitudinal studies: the first

one followed PISA 2003 students in their transition to higher education or the labour market, the second followed PIRLS and TIMSS 2011 students in their passage through the education system. At the same time, ILSAS have not caused any harm to the Czech education system as they have not stimulated any policy measures at all, even with respect to findings where policy interventions would be highly desirable (e.g. educational inequalities). The Czech Republic should endeavour to utilize ILSAs more effectively. This means involving researchers in the national implementation of the surveys so that they could learn from the resources they provide. It is also necessary to plan the participation of the Czech Republic more conceptually so that ILSAS could be utilized for getting answers to topical questions of education policy by extended sampling and cleverly designed national questions and subsequent analyses. The Czech Republic should use the opportunity to learn about the state of the art in the area of educational measurement. Czech governmental agencies could profit from learning about methods of sample surveys, IRT scaling, adaptive testing, framework development, and documenting the development and implementation of instruments. The way national testing is organized in the Czech Republic reveals the need of Czech administrators for professional development in this area. Czech educational researchers, instead of criticizing PISA, should comment on the Czech testing enterprises and thus contribute to their gradual improvement.

Mezinárodní výzkumy výsledků vzdělávání.

Metodologie, přínosy, rizika a příležitosti

Jana Straková

Recenzovali:

Prof. PhDr. Arnošt Veselý, Ph. D.

PhDr. Ing. Petr Soukup

Jazyková korektura:

PhDr. Helena Justová

Vydala Univerzita Karlova v Praze – Pedagogická fakulta

Rok vydání: 2016

Počet stran: 202

Formát: B5

1. vydání

Vytiskla tiskárna Nakladatelství Karolinum

ISBN 978-80-7290-884-4