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

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Mezinárodní výzkumy vědomostí a dovedností a jejich využití I

Doc. RNDr. Jana Straková, Ph.D.

Mgr. Petra Holečková

2020

Strategický projekt UTB ve Zlíně, reg. č. CZ.02.2.69/0.0/0.0/16_015/0002204

Informace o autorech:

Straková, Jana, doc. RNDr. PhD.
Afiliace: Ústav výzkumu a rozvoje vzdělávání, PedF UK
Kontakt: jana.strakova@pedf.cuni.cz

Holečková, Petra, Mgr.
Afiliace: Ústav výzkumu a rozvoje vzdělávání, PedF UK
Kontakt: petra.holeckova@pedf.cuni.cz

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

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

OBSAH

OBSAH 3
ÚVOD 4
1 MEZINÁRODNÍ VÝZKUMY VĚDOMOSTÍ A DOVEDNOSTÍ 5
1.1 HISTORIE VZNIKU MEZINÁRODNÍCH VÝZKUMŮ 5
1.2 NEJDŮLEŽITĚJŠÍ MEZINÁRODNÍ VÝZKUMY V ŽÁKOVSKÉ POPULACI, ZAPOJENÍ ČR 8
1.3 METODIKA MEZINÁRODNÍCH VÝZKUMŮ 11
SEZNAM POUŽITÉ LITERATURY 19
SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK 20

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

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

ÚVOD

Tento materiál slouží jako podklad pro workshop týkající se mezinárodních výzkumů výsledků vzdělávání v ČR a na ně navazujících longitudinálních studií.

Studijní materiál poskytuje stručnou informaci o historii mezinárodních výzkumů vědomostí a dovedností v žákovské i dospělé populaci, o výzkumech realizovaných v ČR a jejich metodice. Uvádí odkazy na výzkumné nástroje, datové soubory, technickou dokumentaci, publikace výsledků a analytické nástroje umožňující provádění jednoduchých analýz se zohledněním specifik dat daných metodikou sběru.

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

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

1 MEZINÁRODNÍ VÝZKUMY VĚDOMOSTÍ A DOVEDNOSTÍ

Tato kapitola přináší základní informaci o mezinárodních výzkumech vědomostí a dovedností. První část poskytuje vhled do jejich historie a vývoje. Druhá část provádí přehled nejdůležitějších výzkumů v žákovské populaci a informuje o účasti ČR v těchto výzkumech. Uvádí odkazy na rozmanité zdroje, kde je možno se o výzkumech poučit, stáhnout si datové soubory, výzkumné nástroje a publikace s výsledky. Třetí část provádí přehled nejdůležitějších metodologických specifik výzkumů s informací o jednoduchých nástrojích umožňujících správně provádět základní analýzy těchto dat. Bližší vhled do problematiky poskytne publikace Strakové (2016) a další průběžně citovaná, případně doporučená literatura.

1.1 Historie vzniku mezinárodních výzkumů

S realizací prvních mezinárodních výzkumů výsledků vzdělávání začali v roce 1958 vzdělavatelé, psychologové, sociologové a psychometrici, kteří se v Institutu pro vzdělávání organizace UNESCO zabývali problematikou hodnocení práce školy a výsledků žáků. Shodovali se 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 odkrýt důležité zákonitosti, které v rámci jednotlivých systémů není možno odhalit. Jejich cílem bylo identifikovat faktory, které mají konsistentní a významný dopad na výsledky vzdělávání. První pilotní šetření, do kterého se zapojilo 12 vzdělávacích systémů, bylo realizováno v roce 1960. V roce 1967 byla právně ustanovena Mezinárodní asociace pro hodnocení výsledků vzdělávání (International Association for the Evaluation of Educational Achievement – IEA), která dodnes realizuje výzkumy v rozmanitých oblastech školního vzdělávání. V ČR jsou nejznámější její výzkumy čtenářské gramotnosti (PIRLS) a matematického a přírodovědného vzdělávání (TIMSS) ve 4. ročnících ZŠ.

O jednotlivých výzkumech pojednává zevrubněji druhá část této kapitoly. Podrobné informace o historii a současných aktivitách této organizace můžeme získat na jejích webových stránkách[1]. V češtině jsou informace dostupné v publikaci Strakové (2016).

Důležitou charakteristikou výzkumů IEA je jejich orientace na školní kurikulum. Je zjišťováno kurikulum zamýšlené – prostřednictvím analýz kurikula zúčastněných zemí a analýz učebnic, dále kurikulum realizované – prostřednictvím žákovských, učitelských a ředitelských dotazníků a kurikulum dosažené – prostřednictvím žákovských testů. Výzkumy IEA jsou zacíleny na poslední ročníky jednotlivých stupňů vzdělávání – 4. a 8. ročníky povinného vzdělávání a pro vybrané výzkumy rovněž 3. ročníky středoškolského vzdělávání. Testovány jsou vždy celé třídy, aby bylo možno propojit data od žáků a učitelů.

Dalším a nyní nejvýznamnějším organizátorem mezinárodních výzkumů v žákovské populaci je Organizace pro hospodářskou spolupráci a rozvoj (OECD). OECD vydává každoročně publikaci Education at a Glance (EaG), do které vstupují nejen statistické indikátory, ale rovněž další indikátory charakterizující rozmanité aspekty vzdělávacích systémů. 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.

[1] https://www.iea.nl

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, kteří byli v rámci projektu osloveni. Projekt shromáždil několik desítek rozmanitých pohledů a přístupů k uchopení problematiky klíčových kompetencí, které byly syntetizovány do následujících tří oblastí (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 interaktivně nástroje

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

Fungovat v sociálně heterogenních skupinách

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

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 velmi podobných oblastem hodnoceným ve výzkumech IEA, tedy čtenářská gramotnost, matematická gramotnost a přírodovědná gramotnost. Testované vědomosti a dovednosti nejsou odvozovány z kurikul jednotlivých zemí, ale z expertního pohledu mezinárodních týmů odborníků, kteří se na konceptualizaci jednotlivých oblastí podílejí. Hodnocení (známé pod zkratkou PISA), které podrobněji popíšeme ve druhé části této kapitoly, probíhá na konci povinného vzdělávání a neúčastní

se jej celé třídy, ale patnáctiletí žáci vybraní náhodně ze všech tříd a ročníků vzdělávajících žáky tohoto věku.

Mezinárodních výzkumů vědomostí a dovedností žáků se v současné době účastní více než 80 zemí z celého světa a jejich výsledky jsou považovány za důležitou zpětnou vazbu pro tvůrce vzdělávacích politik. Výzkumy jsou realizovány periodicky, což umožňuje sledovat nejen okamžitý výkon jednotlivých zemí, ale rovněž dlouhodobější trendy. Výzkumy poskytují řadu užitečných informací o vědomostech a dovednostech žáků v jednotlivých vzdělávacích systémech, o jejich vývoji a o vzdělávacích kontextech. Přesto se však nepodařilo plně naplnit ambici zakladatelů asociace IEA identifikovat univerzální charakteristiky vzdělávacích systémů nejdůležitější pro jejich kvalitní a efektivní fungování. Ukázalo se, že vzdělávací systémy jsou příliš komplexní a jsou příliš ukotveny v kulturách jednotlivých společností a že přenositelnost úspěšných vzorců je velmi omezená.

Vzhledem k vážnosti, která je výsledkům výzkumů přikládána, je velmi důležité, aby jejich výsledky nebyly napadnutelné. Z toho důvodu je věnována velká péče vymezení testovaných oblastí, metodice sběru dat a jejich analytickému zpracování. Kolem mezinárodních výzkumů se sdružují nejlepší odborníci na hodnocení výsledků vzdělávání. Obsahové a metodologické aspekty výzkumů tedy mohou sloužit jako zdroj cenné inspirace a poučení.

Výzkumy se také neustále vyvíjejí, aby splňovaly nároky současné rychle se měnící společnosti. Dochází k rozšiřování testovaných oblastí o další kognitivní i nekognitivní výstupy. Jsou zdokonalovány statistické metody ke zpracování dat. Zároveň se uplatňují moderní metody administrace výzkumů, typicky elektronická administrace.

1.2 Nejdůležitější mezinárodní výzkumy v žákovské populaci, zapojení ČR

Česká republika se zapojila do mezinárodních aktivit asociace IEA v roce 1993 a v roce 1995 zrealizovala ve 4. a 8. ročníku ZŠ výzkum čtenářské gramotnosti, který byl v ostatních zemích proveden v roce 1992. Prvním mezinárodním výzkumem, do kterého se řádně zapojila, byl výzkum Third International Mathematics and Science Study (TIMSS). Výzkum TIMSS v roce 1995 byl prvním z výzkumů, ve kterých se testovala matematika a přírodovědné předměty společně, a zahájil čtyřletý cyklus nyní známý pod názvem Trends

in International Mathematics and Science Study. Ve výzkumu v roce 1995 byli testováni 10letí a 14letí žáci a žáci v posledním ročníku středních škol.

V šetření žáků středních škol se poprvé objevil koncept matematické a přírodovědné gramotnosti. 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 a do kterého se ČR dosud nezapojila.

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. Realizátoři usoudili, že jeho přidaná hodnota neopravňuje navýšení nákladů na realizaci.

Šetření TIMSS[2] 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, zpravidla jsou však testováni pouze žáci 4. ročníků.

[2] https://www.iea.nl/studies/iea/timss

Od roku 2001 organizuje IEA návazné studie čtenářské gramotnosti v pětiletých intervalech pod názvem Progress in International Reading Literacy Study (PIRLS)[3]: 2001, 2006, 2011, 2016. Tyto studie probíhají ve 4. ročnících. Jejich pojetí se od roku 1992 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.

Od roku 2000 realizuje OECD pravidelně ve tříletých intervalech (dosud tedy v letech 2000, 2003, 2006, 2009, 2012, 2015 a 2018) výzkum Programme for the International Student Assessment (PISA)[4]. Projekt PISA hodnotí čtenářskou, matematickou a přírodovědnou gramotnost u patnáctiletých žáků, přičemž v každém sběru klade na jednu z oblastí větší důraz. Tento výzkum je nejdůležitějším aktuálně realizovaným mezinárodním výzkumem a ČR se zúčastnila všech jeho sběrů. Vybrané informace pojednává Straková (2010) nebo Straková, Greger (2013).

Asociace IEA organizuje rovněž výzkumy v dalších oblastech vzdělávání. ČR se dvakrát zúčastnila výzkumu občanské výchovy Civic Education Study (CIVED) v roce 1998 a International Civic and Citizenship Education Study (ICCS)[5] v roce 2009. Tento výzkum žáků 8. ročníků 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. Výzkum přinesl důležité informace o znalostech a postojích českých žáků. Následujícího běhu v roce 2016 se ČR bohužel nezúčastnila a nezapojila se ani do výzkumu, který proběhne v roce 2022.

Dalším výzkumem asociace IEA je výzkum zaměřený na informační technologie International Computer and Information Literacy Study (ICILS)[6]. Tento výzkum zaměřený na žáky 8. ročníků hodnotil prostřednictvím autentických počítačových úloh porozumění počíta-

[3] https://www.iea.nl/studies/iea/pirls
[4] https://www.oecd.org/pisa/
[5] https://www.iea.nl/studies/iea/iccs
[6] https://www.iea.nl/studies/iea/icils

čům a dovednost s nimi pracovat. ČR se zúčastnila v roce 2014, ale do dalšího běhu tohoto výzkumu se rovněž nepřihlásila.

Zkoumání méně standardních oblastí pobíhá i v rámci výzkumu PISA. S každou hlavní hodnocenou oblastí se zpravidla pojí ještě nějaké přídatné zkoumání: se čtenářskou gramotností byly zkoumány metakognitivní dovednosti, s matematickou gramotností bylo zkoumáno řešení problémů a finanční gramotnost, s přírodovědnou gramotností postoje žáků k vědě a technice a k životnímu prostředí.

Nově bylo do výzkumu PISA zařazeno řešení problémů ve spolupráci a globální kompetence[7]. Kompetence pro řešení problémů ve spolupráci byla definována jako schopnost jednotlivce účinně se zapojit do procesu, ve kterém se dva nebo více aktérů pokusí vyřešit problém sdílením porozumění a úsilí potřebného k řešení a shromažďováním svých znalostí, dovedností a úsilí k dosažení tohoto řešení[8].

1.3 Metodika mezinárodních výzkumů

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. 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 nutně zatíženy výběrovou chybou. Měření vědomostí a dovedností je dále nutně zatíženo chybou měření. Použité metody a postupy se snaží výše uvedené chyby minimalizovat, ale

[7] https://www.oecd.org/pisa/pisaproducts/PISA2015-Released-FT-Cognitive-Items.pdf
[8] https://www.oecd.org/pisa/PISA-2018-Global-Competence-Questionnaire.pdf

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.

Výzkumy jsou řízeny z mezinárodních výzkumných center. V jednotlivých zemích je 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. 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. V ČR je za všechny mezinárodní výzkumy v žákovské populaci aktuálně zodpovědná Česká školní inspekce.

V první etapě přípravy mezinárodního výzkumu výsledků vzdělávání jsou vymezeny hodnocené oblasti prostřednictvím konceptuálních rámců. 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).

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 jed-

noho 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 v mezinárodních výzkumech jsou tvořené z poloviny úlohami, ve kterých žáci vybírají z několika nabízených odpovědí jedinou správnou. Polovinu testů tvoří úlohy, ve kterých žáci odpovídají jednoslovně či několika slovy nebo kde poskytují obsáhlejší odpověď nebo úplné řešení dané úlohy. Vlastní odpovědi žáků jsou následně vyhodnocovány hodnotiteli podle velmi podrobných návodů.

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.

Při počítačové administraci komplikuje situaci různá délka jednotlivých jazyků, neboť je důležité, aby se na obrazovku ve všech jazycích vešly vždy všechny potřebné informace. 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.

Velká péče je v mezinárodních výzkumech věnována tomu, aby úlohy byly voleny tak, aby žáky z žádných zemí nezvýhodňovaly, 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.

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. Výsledné národní překlady jsou verifikovány nezávislým překladatelem z mezinárodního koordinačního centra.

V mezinárodních výzkumech 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 nebo věku. Výběr je realizován jako systematický podle velikosti cílové populace na jednotlivých školách, v rámci těchto škol pak je náhodně vybrána jedna nebo dvě třídy v příslušném ročníku (výzkumy IEA) nebo určitý počet žáků příslušného věku (PISA). V rámci výběru jsou pro každou školu vybírány 2 školy náhradní, které jsou oslovovány 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. Země musí splnit stanovená kritéria pro návratnost. Například 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.

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 a podle jednotného scénáře, který je opět překládán do národních jazyků z anglické předlohy. Průběh administrace je pečlivě zaznamenáván do připravených záznamových archů.

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. Národní hodnotitelé jsou školeni pracovníky národního centra, kteří absolvovali mezinárodní školení, podle tréninkového manuálu a s využitím při-

pravených ukázek žákovských odpovědí. Žákovské odpovědi jsou hodnoceny nezávisle dvěma hodnotiteli.

Mezinárodní koordinační centrum zpravidla jednotlivá národní koordinační centra vybaví specializovaným software, 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 navíc pořídit hodnocení žákovských odpovědí a některých dotazníkových položek.

Datové soubory obsahují některé specifické proměnné, jejich využití je nezbytné pro řádnou práci s těmito soubory. První z těchto proměnných je žákovská váha. Ta zohledňuje skutečnost, že žáci jsou sice do výzkumů vybíráni náhodně, pravděpodobnost výběru každého z nich se ale liší. Váhy zajišťují, že každý respondent reprezentuje odpovídající počet žáků z cílové populace. Kompenzují nejčastěji skutečnost, že určité podskupiny cílové populace byly vybrány disproporčně (jsou zastoupeny více nebo naopak méně) a že se některé školy nebo žáci výzkumu nezúčastnili.

Kromě žákovské váhy jsou v datových souborech obsaženy váhy replikační. Ty korigují skutečnost, že výběr respondentů není prostý náhodný, ale skupinkový, ve kterém jsou žáci vybíráni prostřednictvím škol. Vybírat žáky ve školách je jednoduché a přirozené a mohou v nich snadno 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.

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 rozdíly. Výpočty standardních chyb jsou totiž v běžně dostupných statistických programech založeny na předpokladu prostého náhodné-

ho výběru, kdy z množiny všech možných respondentů náhodně vybíráme respondenty, které podrobíme zkoumání. V datových souborech jsou zařazeny proměnné, které při využití příslušných programů a metod analýzy dat umožňují provést korekci na skupinkový výběr, jejímž výsledkem jsou realistické standardní chyby.

Pro datové soubory z mezinárodních výzkumů vědomostí a dovedností je dále poněkud netypický způsob vyjadřování testových výsledků. 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. 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.

Pro modely IRT platí, že nemají žádné přirozené stupnice, tedy že 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).

Každé měření vědomostí a dovedností je zatíženo chybou. Chyba může být ve všech případech redukována navýšením počtu testových položek, které respondent řeší. Již bylo 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 skórů, které jsou označovány jako plausibilní hodnoty. Tyto plausibilní hodnoty jsou pak užívány k prezentaci výsledků a ke standardním statistickým výpočtům. V datových souborech z mezinárodních výzkumů nalezneme pro každou testovanou oblast 5 nebo 10 plausibilních hodnot. Počítáme-li nějakou statistiku, měli bychom správně vypočítat tuto statistiku pro jednotlivé plausibilní hodnoty zvlášť a výsledky následně zprůměrovat. V praxi ovšem často počítáme pouze s jednou plausibilní hodnotou.

Při statistických výpočtech tedy musíme dbát na to, abychom používali váhy, abychom zohlednili způsob výběru respondentů pomocí replikačních vah a abychom řádně pracovali s plausibilními hodnotami. Nejjednodušší cesta je využít k tomuto účelu nástroje vytvořené přímo institucemi, které mezinárodní výzkumy realizují. Na stránkách výzkumu PISA na-

lezneme pomůcku Data explorer[9], která umožňuje počítat deskriptivní statistiky a porovnávat výsledky jednotlivých zemí v třídění podle kontextuálních proměnných. Na stránkách Mezinárodní asociace pro hodnocení výsledů vzdělávání IEA se nachází program IDB analyzer[10], který vytváří syntaxe pro statistický balík SPSS umožňující řádně počítat jednoduché deskriptivní i inferenční statistiky. Pro statistické výpočty dále potřebujeme popis proměnných obsažených v datových souborech. Veškeré potřebné informace jsou uvedeny v podrobné dokumentaci, kterou je možno si stáhnout ze stránek jednotlivých výzkumů.

[9] https://pisadataexplorer.oecd.org/ide/idepisa/
[10] https://www.iea.nl/data-tools/tools

SEZNAM POUŽITÉ LITERATURY

Palečková, J., Mandíková, D. (1997). Třetí mezinárodní výzkum matematického a přírodovědného vzdělání. Výsledky českých žáků v praktických úlohách. Praha: VÚP.

Rychen D.S., Salganik L.H. (Eds.). (2001). Defining and Selecting Key Competencies. Göttingen: Hogrefe & Huber Publishers.

Straková, J., Greger, D. (2013). Faktory ovlivňující přechod žáků 5. ročníků na osmileté gymnázium. Orbis Scholae 7 (3), 73 -85.

Straková, J. (2010). Přidaná hodnota studia na víceletých gymnáziích ve světle dostupných datových zdrojů. Sociologický časopis / Czech Sociological Review 46 (2): 187-210

Straková, J. (2016). Mezinárodní výzkumy výsledků vzdělávání. Metodologie, přínosy, rizika a příležitosti. Praha: Pedagogická fakulta UK.

Další doporučená literatura:

Anýžová, P., Večerník, J. (Eds.). (2019). Vzdělání, dovednosti a mobilita: zaměstnání a trh práce v České republice a evropských zemích. Praha: Univerzita Karlova, nakladatelství Karolinum.

Binkley, M., Sternberg, R., Jones, S. Nohara, D. (1999). An overarchning framework for understanding and asessing lifeskills. In: Frameworks: Working drafts. Briefing materials for National Study Managers meeting, Luxembourg.

Matějů, P, Hamplová, D., Hampl, P., Loužek, M., Weidnerová, S., Anýžová P., Smith M. L. (2017). Moc krásy: pomáhá krása a atraktivita k životnímu úspěchu? Praha: Univerzita Karlova, nakladatelství Karolinum.

Janíčko, M., Koutná M. (2016). Pracovní dráhy v době ekonomického oživení v České republice. Zpráva z výzkumu „SKILLS“. Praha: Národní vzdělávací fond, o.p.s., Národní observatoř zaměstnanosti a vzdělávání.

OECD. (2000). Literacy in the Information Age. Final Report from the Adult Literacy Survey. Paris: OECD.

SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK

ALL	Adult Literacy and Lifeskills Survey
CIVED	Civic Education Study
DeSeCo	Defining and Selecting Key Competencies
EaG	Education at a Glance
IALS	International Adult Literacy Survey
ICCS	International Civic and Citizenship Education Study
IEA	International Association for the Evaluation of Educational Achievement
OECD	Organisation for Economic Co-operation and Development
PIAAC	Programme for the International Assessment of Adult Competencies
PISA	Programme for the International Student Assessment
PIRLS	Progress in International Reading Literacy Study
TIMSS	Trends in International Mathematics and Science Study

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

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY