



Hodnocení vzdělávacího přínosu škol pomocí přidané hodnoty: srovnání čtyř evropských zemí¹

EVA POTUŽNÍKOVÁ

Abstrakt: Cílem práce je na základě porovnání situace ve čtyřech evropských vzdělávacích systémech (Anglie, Francie, Norsko, Polsko) poukázat na možnosti, ale i rizika spojená s evaluací škol na základě přidané hodnoty.

Metody. Přístupy jednotlivých zemí jsou porovnány podle předem určeného souboru kritérií. Zdrojem dat je technická dokumentace o modelech přidané hodnoty, webové portály pro zveřejňování výsledků a texty o vývoji modelů a využívání dat o přidané hodnotě.

Výsledky. Preferovaným zdrojem dat pro výpočet přidané hodnoty jsou standardizované testy či zkoušky. Dostupnost dat ovlivňuje volbu vzdělávacího stupně, hodnocených vzdělávacích oblastí (předmětů) i zařazení kontextových proměnných. Při zveřejňování přidané hodnoty země brání otevřenému srovnávání škol, netvoří žebříčky a přiznávají statistickou nejistotu spojenou s měřením. Informace o přidané hodnotě slouží jako zpětná vazba školám.

Závěry. Indikátory přidané hodnoty jsou spravedlivým ukazatelem vzdělávacího přínosu škol, s jejich zaváděním jsou však spojena rizika jako vynechání důležitých proměnných kvůli nedostupnosti dat, nedostatek odborných kapacit pro implementaci a rozvoj metody a malá využitelnost pro zlepšování kvality vzdělávání.

Klíčová slova: měření přidané hodnoty, standardizované testování, kvalita škol, srovnávání, Anglie, Francie, Norsko, Polsko

Jedním ze znaků současné vzdělávací politiky je obrat pozornosti k výsledkům vzdělávání. Řada zemí definovala očekávané výstupy a zavedla testy, jimiž ověřuje jejich dosažení. Od devadesátých let 20. století pak roste využívání testů nejen k posouzení individuálního výkonu,

ale také k hodnocení práce škol, učitelů i celých vzdělávacích systémů (Eurydice, 2009) s cílem zvýšit transparentnost a zároveň zlepšit kvalitu vzdělávání (Mons, 2009).

Hodnocení škol na základě dosažených výsledků vychází z předpokladu, že

¹ Části textu (deskriptivní informace o modelech používaných v Anglii, Francii a Norsku) byly publikovány ve zprávě České školní inspekce (ČŠI, 2019). Pro potřeby předloženého textu byly tyto části zkráceny a doplněny o nová zjištění.



standardizované testy dobře měří kvalitu vzdělávání ve školách (Mons, 2009). Výzkumy efektivity vzdělávání však dlouhodobě ukazují, že za výsledky žáků jsou zodpovědné spíše mimoškolní faktory (Coleman et al., 1966; Reynolds et al., 2014). Posuzování kvality vzdělávání podle testových výsledků je proto mnohými považováno za nespravedlivé, protože činí školy zodpovědné za něco, co mohou ovlivnit jen v omezené míře. Vysoký tlak na výkon spolu s vnímanou neférovostí hodnocení může mít přitom negativní dopad na motivaci k učení a vyučování (Jones, 2007; Mons, 2009) i na celkovou míru důvěry v systému (Penninckx et al., 2017).

V posledních dvou desetiletích se začaly prosazovat přístupy, které místo dosažených výsledků sledují tzv. přidanou hodnotu škol (OECD, 2008). V situaci, kdy jsou mezi školami značné rozdíly ve složení žáků, je toto hodnocení potenciálně férovější (Everson, 2017; Harris, 2011) a zároveň je vnímáno jako užitečnější pro zlepšování kvality vzdělávání (Penninckx et al., 2017). Cílem této práce je porovnat přístupy k zjišťování přidané hodnoty ve vybraných evropských zemích a na základě jejich zkušeností lépe porozumět možnostem, ale i rizikům spojeným s touto metodou, o níž uvažují i dokumenty české vzdělávací politiky. Nejprve představím koncept přidané hodnoty ve vzdělávání a její zjišťování, poté popíšu přístupy používané ve vybraných

zemích a poukážu na podobnosti a rozdíly, na závěr pojednám o přínosech a limitech měření přidané hodnoty v kontextu vzdělávací politiky.

1. KONCEPT PŘIDANÉ HODNOTY VE VZDĚLÁVÁNÍ

V ekonomii termín „přidaná hodnota“ vyjadřuje rozdíl mezi hodnotou výrobku a cenou surovin použitých k jeho zhotovení. Ve vzdělávání se tento termín používá v přeneseném smyslu pro posouzení vzdělávacího přínosu školy² na základě učebního pokroku, který zde žáci učiní. S myšlenkou zjišťovat přidanou hodnotu na základě učebního pokroku přišel začátkem sedmdesátých let 20. století Eric Hanushek (1971), který navrhl obecný model pro analýzu vzdělávacích výsledků v závislosti na různých typech „vstupů“. Tento model zohledňuje také vliv rodiny, vrstevníků a další vstupy, které škola nemůže ovlivnit, aby poskytoval objektivnější informace o vzdělávacím přínosu školy. Hanushek (tamtéž) upřesňuje, že pokud model na straně vstupů zahrnuje počáteční úroveň znalostí a dovedností, může pominout řadu dalších vstupů, které mají potenciálně vliv na dosažený výsledek, aniž by to způsobilo vážné zkresení: „Model s počátečním výkonem měří ‚přidanou hodnotu‘ různých vstupů a zkresení nastane pouze tehdy, když chybějící [proměnná] koreluje s rychlostí učení“ (s. 281).

² Přidanou hodnotu lze zjišťovat pro jakékoli jednotky vzdělávacího systému, k nimž lze přiřadit výsledky žáků; tato práce se zaměřuje na přidanou hodnotou škol.



Přidanou hodnotu školy lze definovat jako rozdíl mezi výsledkem, jehož žáci skutečně dosáhnou, a výsledkem, jaký by se u nich dal očekávat, kdyby se – při stejných vstupních znalostech a dovednostech – vzdělávali v průměrné škole (Everson, 2017; Harris, 2011). Přidaná hodnota se určuje pomocí statistických modelů, z nichž nejběžnější jsou lineárně regresní modely, které predikují očekávaný výsledek na základě předchozího výsledku, a hierarchické (víceúrovňové) lineární modely, které navíc zohledňují vzájemnou podobnost žáků z téže školy (Levy et al., 2019; OECD, 2008). Protože nejsilnějším prediktorem dosaženého výsledku jsou předchozí znalosti a dovednosti (Lenkeit, 2013), je určování přidané hodnoty na základě předchozího výsledku přesnější než její zjišťování pomocí tzv. kontextových modelů, které zohledňují sociodemografické charakteristiky žáků, ale ne jejich vstupní znalosti a dovednosti (Kelly & Downey, 2010; OECD, 2008; Tekwe et al., 2004).

Při interpretaci přidané hodnoty je třeba mít na paměti, že vypočtená hodnota je pouze odhadem vzdělávacího přínosu, jehož přesnost závisí na použitém modelu, kvalitě dat, ale také na podmínkách vzdělávacího systému, např. na rozložení žáků ve školách. Velká výzkumná pozornost byla věnována hodnocení různých modelů, ale zatím bez jednoznačného závěru. Vzhledem ke „klastrové“ povaze dat ze vzdělávacího systému jsou hierarchické modely obvykle hodnoceny lépe než klasické jednoúrovňové regresní modely (Levy et al., 2019). Protože některé sku-

piny žáků mohou i přes úsilí školy učinit menší pokrok v důsledku nižší podpory rodiny (Harris, 2011), doporučuje řada autorů přidat do modelu sociodemografické proměnné, které mohou mít vliv na výsledky žáků (Koedel, Mihaly & Rockoff, 2015; Leckie & Goldstein, 2017, 2019; Reynolds et al., 2014; Tekwe et al., 2004), především socioekonomický status (Everson, 2017). Za určitých podmínek však nemusí být jejich přidání potřeba, např. když model obsahuje větší počet předchozích výsledků z různých let nebo předmětů (Ballou, Sanders & Wright, 2004; Koedel et al., 2015). Z praktických důvodů je účelné použít co nejjednodušší model, pokud je „dostatečně dobrý“, tj. negeneruje vážně zkreslené odhady (Kelly & Downey, 2010; OECD, 2008; Perry, 2016; Tekwe et al., 2004). Vlivem malého počtu žáků, na nichž jsou počítány hodnoty pro jednotlivé školy, jsou získané odhady zatíženy velkou náhodnou chybou (Harris, 2011). Při používání dat o přidané hodnotě k evaluaci škol se proto doporučuje pracovat s intervaly spolehlivosti (Perry, 2016), stanovit minimální počet žáků, pro něž lze odhad pokládat za věrohodný (OECD, 2008), nebo navýšit počet žáků spojením dat z více věkových kohort (Harris, 2011; OECD, 2008).

Zjišťování či měření přidané hodnoty (*value-added measurement*) se rozvíjí především v USA, kde je díky zákonné povinnosti každoročního testování množství dat o výsledcích žáků, jež řada států využívá k posuzování přidané hodnoty učitelů (Everson, 2017). Tato praxe je velmi kontroverzní, i kritici však připouštějí,



že odhady přidané hodnoty jsou lepším měřítkem kvality práce učitelů než dřívější posuzovací kritéria (Harris, 2011). Evropské země mají oproti USA méně dat z testování, přesto i zde je patrný zájem o měření přidané hodnoty. V roce 2005 zahájila OECD na návrh Norska projekt, jehož cílem bylo prozkoumat možnosti této metody a podpořit její zavádění. Do projektu se zapojila Austrálie a dvanáct evropských zemí včetně České republiky (Středisko vzdělávací politiky, 2008). Na projekt u nás navázali Ryška (2009) a Malach, Malčík & Krpec (2011), kteří provedli i delší empirické analýzy, kvůli nedostatku dat však nebylo téma dále podrobněji zkoumáno.

Měření přidané hodnoty nyní nabývá na aktuálnosti po zavedení jednotných přijímacích zkoušek na střední školy, jejichž výsledky by potenciálně bylo možné propojit s maturitami, a také v souvislosti s obnovením zájmu o plošné testování v 5. a 9. ročníku ZŠ (MŠMT, 2020). V dokumentech české vzdělávací politiky je měření přidané hodnoty zmiňováno jako vhodná metoda pro hodnocení kvality vzdělávání (MŠMT, 2014, 2015), aniž by byly dostatečně prodiskutovány její možnosti a meze. Cílem této práce je přispět do diskuse analýzou zkušeností z evropských zemí. Práce hledá odpovědi na tyto výzkumné otázky:

(1) Jaká data a modely se v Evropě používají k měření přidané hodnoty?

(2) K jakým účelům se využívají informace o přidané hodnotě a jaký je jejich přínos?

2. METODA PRÁCE A ZDROJE DAT

Předložená práce je srovnávací studií, jejímž předmětem je používání metody měření přidané hodnoty k hodnocení kvality škol v různých zemích. Chce přinést závěry relevantní pro vzdělávací politiku, a proto byla analýza předem omezena na země, které měření přidané hodnoty realizují na administrativních (nikoli výzkumných) datech. Dále byla analýza omezena na Evropu, kde testování probíhá v podobném rozsahu, jaký se plánuje v ČR. V prvním kroku byla vybrána Anglie, která má s měřením přidané hodnoty dlouhodobou zkušenost (OECD, 2008), poté byly přidány Francie, Norsko a Polsko. Snahou bylo získat soubor sociokulturně rozmanitých zemí, u nichž lze očekávat různé interakce mezi globálně rozšířeným nástrojem a místním kontextem (Dvořák, 2015). Závěry založené na porovnání těchto zemí mohou mít obecnější platnost.

Práce je založena na analýze dokumentů. Hlavním zdrojem dat jsou vládní dokumenty, oficiální analytické zprávy a webové portály s výsledky škol, které byly doplněny odbornými texty o vývoji modelů a využívání indikátorů přidané hodnoty. Sběr a analýza dat probíhaly iterativně. Nejprve jsem z webových stránek ministerstev a institucí zodpovědných za hodnocení shromáždila základní informace o modelech přidané hodnoty, které jsem utřídila podle předem stanovených srovnávacích kritérií. Poté jsem na internetu a v bibliografických databázích UK dohledávala informace k nedosta-



tečně obsazeným kritériím. Použila jsem i texty v národních jazycích vyhledané podle názvů indikátorů získaných z anglicky psaných textů. Pro snazší orientaci v norských a polských textech jsem využila strojový překlad do angličtiny pomocí Google Translate.

Na základě literatury věnované systémům národního testování (Eurydice, 2009; Mons, 2009) a specifikaci modelů přidané hodnoty (Everson, 2017; Koedel et al., 2015; Tekwe et al., 2004) jsem zvolila tato srovnávací kritéria:

- stupeň vzdělávání, na kterém se měří přidaná hodnota;
- vzdělávací oblasti (předměty), pro něž se sleduje přidaná hodnota;
- typ vzdělávacích výsledků vstupujících do výpočtu přidané hodnoty;
- statistický model použitý pro výpočet;
- sdělování informací o přidané hodnotě;
- využívání údajů o přidané hodnotě.

3. VÝSLEDKY

Prezentace výsledků je inspirována Beredayovým (1964) modelem srovnávání ve čtyřech krocích: popis, interpretace, juxtapozice a srovnání. Kostru pro juxtapozici (postavení zemí vedle sebe) a jejich následné srovnání tvoří zmíněná srovnávací kritéria (Vlček, 2015). Popis sleduje vývojovou linii, která napomáhá porozumění situaci v dané zemi a interpretaci získaných dat. Do závěrečného srovnání vstupují nejnovější přístupy k měření přidané hodnoty, ale částečně je zohledněn i vývoj modelů.

3.1 Anglie

Od roku 1992 byly v Anglii zveřejňovány žebříčky s výsledky sekundárních škol ve zkouškách GCSE (*General Certificate of Secondary Education*), které měly zvýšit transparentnost vzdělávacího systému, pomoci rodičům při volbě školy a motivovat školy ke zlepšování v konkurenčním prostředí. Měřítkem úspěšnosti školy byl podíl žáků, kteří získali alespoň pět známek A*–C. Školy, které nedosáhly stanoveného podílu žáků, byly hodnoceny jako selhávající a čelily závažným důsledkům včetně zvýšené externí kontroly, sloučení s jinou školou nebo uzavření (Ježková et al., 2010). V roce 1996 začaly být zveřejňovány analogické žebříčky primárních škol založené na výsledcích žáků v národních testech z angličtiny, matematiky a přírodovědy³ na konci 6. ročníku. Po kritice odborníků poukazujících na neférovost žebříčků a deformování výuky, jež se na mnohých školách jednostranně zaměřovala na hraniční žáky, kteří měli ještě šanci získat známku C (Leckie & Goldstein, 2017), bylo měřítko 5 A*–C v roce 2002 doplněno o přidanou hodnotu, která mohla být spočítána díky propojení individuálních výsledků s výsledky v 6. ročníku přes žákovský identifikační kód (Kelly & Downey, 2010).

V letech 2002–2005 se přidaná hodnota počítala tzv. mediánovou metodou jako školní průměr rozdílů mezi skutečným výsledkem každého žáka a mediánem výsledků všech žáků z celé země,

³ Test z přírodovědy byl v roce 2009 zrušen.

kteří dosáhli v předchozím testování stejné úrovně výkonu (Ray, 2006). Tento postup lze považovat za variantu lineární regrese (Leckie & Goldstein, 2019). Přestože teoreticky měl indikátor přidané hodnoty zajistit spravedlivé hodnocení vzdělávacího přínosu škol, analýza reálných dat prokázala vysoké korelace mezi přidanou hodnotou a absolutními výsledky na konci sledovaného období (Gorard, 2006, 2008; Kelly & Downey, 2010). Model nezohledňoval socioekonomické proměnné a *de facto* zvýhodňoval školy navštěvované žáky s příznivějším rodinným zázemím.

V roce 2006 byl ve spolupráci s odborníky z akademické sféry (např. H. Goldsteinem) zaveden nový postup známý pod zkratkou CVA (*contextual value-added*). Přidaná hodnota byla odhadována pomocí hierarchického modelu, který vedle předchozího výsledku zohledňoval socio-demografické proměnné,⁴ jejich interakce a školní průměr a rozptýl předchozího výsledku. Indikátor CVA vedl k přehodnocení obrázku o efektivitě škol, zvláště těch s vyšším podílem znevýhodněných žáků (Kelly & Downey, 2010). Jeho nevýhodou byl složitý postup výpočtu (Ray, 2006), malá využitelnost pro zlepšování výuky (Kelly & Downey, 2010), kritizováno bylo rovněž implicitní snižování nároků na znevýhodněné žáky (Bradbury, 2011). To byly také oficiální důvody pro opuštění

CVA (Leckie & Goldstein, 2017), ačkoli motivací mohla být i obava z explicitního tematizování neschopnosti vzdělávacího systému poskytnout kvalitní vzdělání znevýhodněným žákům (Reynolds et al., 2014).

V roce 2011 byl CVA nahrazen indikátorem očekávaného pokroku (*expected progress*) sledujícím podíl žáků, kteří se během sekundární školy posunou v angličtině a matematice alespoň o tři úrovně národního kurikula.⁵ Deklarovaným cílem bylo posílit vazbu měřítka pokroku na národní kurikulum a poskytnout včasnou podporu žákům, jimž hrozí špatný výsledek u zkoušek GCSE. V rámci akontability mohl indikátor očekávaného pokroku vyvážit nedostatečný podíl žáků s pěti známkami A*–C, pokud se neúspěšní žáci posunuli o tři úrovně. Očekávaná míra pokroku však nerespektovala reálné možnosti studijně slabých, často znevýhodněných žáků, úrovně byly také nastaveny příliš široce (Leckie & Goldstein, 2017)⁶ a indikátor byl po pěti letech zrušen.

Od roku 2016 slouží k měření přidané hodnoty *Progress 8* (DfE, 2020), který lze považovat za návrat k původní mediánové metodě. Základem výpočtu je průměrný výsledek žáka v osmi zkouškách GCSE – tzv. *Attainment 8*, který v roce 2016 nahradil měřítko 5 A*–C. *Progress 8*

⁴ Pohlaví, věk, speciální vzdělávací potřeby, etnický původ, mateřský jazyk, meziškolní mobilita, nárok na bezplatný školní oběd, index příjmové deprivace místa bydliště a zda je žák v péči státu. Tyto proměnné byly dostupné ve školních matrikách a přes žakovský identifikační kód mohly být spárovány s výsledky (Ray, 2006).

⁵ Tomu odpovídá např. posun z úrovně 4 na konci primární školy na známku C ve zkoušce GCSE.

⁶ Například v roce 2014 učinilo v matematice očekávaný pokrok 96 % žáků startujících v horním pásmu úrovně 5, ale jen 20 % žáků startujících v dolním pásmu úrovně 3.



se počítá jako školní průměr rozdílů mezi individuálním skórem *Attainment 8* a průměrným skórem *Attainment 8* všech žáků, kteří v předchozím období dosáhli stejného výsledku. Jiné charakteristiky žáků se nezohledňují. Podobné indikátory se používají také pro primární školy a školy pro žáky nad 16 let, kde se počítají pouze pro akademické a všeobecné kvalifikace (viz tab. 1). Na rozdíl od minulosti, kdy indikátory založené na pokroku žáků doplňovaly či vyvažovaly informace o absolutní úspěšnosti, jsou nyní hlavním měřítkem kvality škol (DfE, 2019a; 2019b; 2020; Perry, 2016).

3.2 Francie

Od roku 1981 zveřejňoval *Monde de l'éducation* žebříčky úspěšnosti středních škol (*lycée*) v maturitní zkoušce (*baccalauréat*). Ministerstvo školství si uvědomovalo, že toto populární měřítko nezohledňuje rozmanitost podmínek, v nichž školy pracují, a chtělo dát školám do ruky nástroje, které jim poskytnou komplexní zpětnou vazbu (Pons, 2012). Tato tzv. reflektivní forma akontability, která školy nevystavuje přímým sankcím, ale předpokládá vlastní snahu o zlepšování ve světle dostupných informací, je pro Francii příznačná (Pons, 2020) a proniká i do současných dokumentů o přidáné hodnotě (DEPP, 2020; Evain & Evrard, 2017). V letech 1984–1993 postupně vznikla sada asi třiceti indikátorů, pro které byla stano-

vena referenční úroveň (většinou celostátní průměr), s níž se školy mohly poměřovat (Pons, 2012). Vedle tradičního podílu žáků, kteří úspěšně složí *baccalauréat*, se mj. začala hodnotit schopnost školy předcházet předčasným odchodům ze vzdělávání pomocí podílu žáků, kteří dojdou k úspěšnému složení zkoušky od zahájení studia na škole. Od roku 1994 se pro tyto dva indikátory počítá také přidáná hodnota, která vyjadřuje rozdíl mezi skutečným a očekávaným podílem (Duclos & Murat, 2014; Evain & Evrard, 2017).

První verze indikátorů přidáné hodnoty (*indicateurs de valeur ajoutée des lycées*, IVAL), používaná v letech 1994–2007, nezohledňovala předchozí výsledky žáků. Do výpočtu očekávaného podílu byly zahrnuty sociodemografické proměnné, které podle dostupných údajů dobře predikovaly šanci na úspěch u zkoušky – socioprofesionální kategorie otce (v neúplných rodinách matky) a věk žáka⁷ (Duclos & Murat, 2014). Nezohlednění předchozího výsledku, ačkoli byla k dispozici data z testování na vstupu do *lycée* (Pons, 2020), bylo odborníky považováno za hlavní limitaci modelu (Felouzis, 2005), který systematicky generoval nízkou přidanou hodnotu pro školy s nižšími očekávanými výsledky. Kritizován byl také zvolený typ modelu (lineární regrese). Už před rokem 1994 ověřovali výzkumníci model zohledňující předchozí výsledek, závěry výzkumu však nebyly pro ministerstvo dost přesvědčivé (Pons, 2012).

⁷ Takzvaný *retard scolaire*, který určuje, zda má žák v daném ročníku školní docházky odpovídající věk, o rok vyšší věk nebo o dva či více let vyšší věk.



Změna modelu přidané hodnoty v roce 2008 byla podmíněna souhrou několika okolností, mj. standardizací písemné části zkoušky *diplôme national de brevet* (DNB), kterou se zakončuje předchozí stupeň (*collège*), reorganizací ministerského odboru zodpovědného za evaluaci a zapojením země do projektu OECD (Duclos & Murat, 2014; Pons, 2012). Nový model byl koncipován jako dvouúrovňový logistický model s náhodnými efekty, který na individuální úrovni zohledňuje socioprofesionální kategorii otce (popř. matky), věk a pohlaví žáka a na úrovni školy podíl žáků v jednotlivých kategoriích plus průměrnou známku ve zkoušce DNB. Předchozí výsledek nebyl zařazen na individuální úrovni kvůli potížím s párováním dat. Dosazením do rovnice modelu se pro každého žáka spočítala očekávaná pravděpodobnost úspěchu a ze získaných hodnot se vytvořil školní průměr, který se porovnal se skutečným podílem úspěšných žáků (Duclos & Murat, 2014).

Poslední aktualizace modelu proběhla v roce 2015, kdy byla kvalita dat již natolik spolehlivá, že bylo možné spárovat výsledky většiny žáků u zkoušek DNB a *baccalauréat*. Zároveň se zpřesnilo měření sociálního původu pomocí socioekonomického indexu namísto čtyř socio-profesionálních kategorií (Rocher, 2016). Pro výpočet přidané hodnoty se i nadále používá víceúrovňový logistický model, který zohledňuje na individuální úrovni věk, pohlaví, socioekonomický index, průměr známek z písemných zkoušek DNB a na úrovni školy agregované varianty (podíly nebo průměry) stejných pro-

měnných (DEPP, 2020; Evain & Evrard, 2017). Kvůli velkým rozdílům mezi typy škol se koeficienty modelu odhadují zvlášť pro různé skupiny oborů (*série*).

3.3 Norsko

Norský vzdělávací systém byl tradičně založen na decentralizaci a důvěře ve školy, prioritou bylo zajištění široké dostupnosti vzdělání (Hatch, 2013). Zájem o dosažené výsledky vyvolal až tzv. PISA-šok po zveřejnění výsledků PISA 2000. Vláda zahájila práci na reformě, jejíž součástí byla revize kurikula, definování vzdělávacích standardů a měření výsledků (Tveit, 2014). V roce 2004 vstoupil v platnost národní systém hodnocení kvality zahrnující plošné testy, průzkumy mezi učiteli, žáky a rodiči, inspekční hodnocení a webový portál s otevřeným přístupem k datům (Nusche et al., 2011). To školy vnímaly jako narušení důvěry ve svou práci a zahájily bojkot testů, který vedl k pádu vlády (Elstad, 2009; Tveit, 2014).

Po dočasném pozastavení bylo testování v roce 2007 obnoveno jako zdroj informací pro včasnou diagnostiku a podporu selhávajících žáků (Elstad, 2009; Tveit, 2014). Podoba testů se po čase ustálila na plošných testech ze čtení, matematiky a angličtiny v 5. a 8. ročníku a na diagnostických testech v 1.–3. ročníku. Testy nemají významný dopad na žáky ani na školy, od škol se však očekává využívání dat k autoevaluaci (Nusche et al., 2011) a přízpusobení výuky podle dosažených výsledků (Tveit, 2014). Naopak tradiční součástí norského školství jsou závěrečné



zkoušky na konci základní školy (10. ročník) a střední školy, v současnosti tvořené kombinací centrálních a školou zadávaných zkoušek (Nusche et al., 2011). Výsledky žáků u zkoušek nemají formální dopad na hodnocení škol (Hatch, 2013) a podobně jako testy mají školám sloužit pro zlepšování výuky.

Už od zavedení národního systému hodnocení kvality v roce 2004 byl kromě výsledků sledován také tzv. přínos škol (*Skolebidragsindikatorer*) jako doplňkový indikátor umožňující rozpoznat efektivní školy, jejichž metody práce mohou být inspirací pro ostatní (Elstad, 2009; Hægeland et al., 2005). V první fázi nebyla k dispozici data o předchozích výsledcích, pro odhad přínosu školy se proto používal kontextový model kontrolující široké spektrum žákovských charakteristik dostupných ve vzájemně propojitelných databázích (Hægeland et al., 2011).⁸ Měření přidané hodnoty na základě předchozích výsledků začalo Norsko ověřovat s podporou projektu OECD nejprve s využitím známek. Následující analýzy, které již použily data z národních testů v 5. a 8. ročníku, podpořily model zohledňující předchozí výsledek a menší počet kontextových proměnných – vzdělání rodičů, příjem domácnosti a imigrační status (Hægeland et al., 2011; Steffensen et al., 2017). Pilotní analýzy pracovaly

s víceúrovňovými modely, pro praxi však byla upřednostněna jednoúrovňová regrese,⁹ která určuje přidanou hodnotu jako školní průměr žákovských reziduí (SSB, 2019).

Přidaná hodnota se počítá také pro střední školy, a to pro průměr známek na konci ročníku,¹⁰ pro podíl žáků, kteří postoupí do dalšího ročníku, a pro podíl žáků, kteří se účastní vzdělávání bez ohledu na to, zda postoupí. Pro výpočet se používá regresní, resp. logistický model s malým počtem kontrolních proměnných – průměr známek z předchozího ročníku, jeho druhá mocnina a pohlaví (SSB, 2018). Socioekonomické proměnné měly při kontrole předchozích známek jen malou vysvětlovací schopnost (Hægeland et al., 2011) a do finálního modelu pro střední školy nebyly zařazeny. Koefficienty modelu se odhadují zvlášť pro každou kombinaci ročníku a typu studia (všeobecné nebo odborné).

3.4 Polsko

Změny, které v Polsku probíhaly po roce 1989, se zprvu příliš netýkaly vzdělávání, v zemi panovalo všeobecné přesvědčení o dobré kvalitě školství. Nepříznivý ekonomický vývoj, růst nezaměstnanosti a špatné výsledky Polska v mezinárodním výzkumu funkční

⁸ Pohlaví, věk, rodinná struktura, imigrační status, vzdělání rodičů, příjem a majetek domácnosti (na základě daňových odvodů), pobírání sociálních dávek.

⁹ Do modelu je zařazena také druhá a třetí mocnina předchozího výsledku, jedná se tedy o nelineární (polynomiální) regresi.

¹⁰ Pro zvýšení srovnatelnosti mezi školami jsou známky korigovány výsledkem v jednotných závěrečných zkouškách (SSB, 2018).



gramotnosti dospělých však připravily půdu pro školskou reformu, která zahrnovala prodloužení povinné školní docházky z osmi na devět let, nahrazení základní školy šestiletou primární školou a tříletým gymnáziem (*gimnazjum*),¹¹ reformu kurikula a hodnocení vzdělávacích výsledků (Wiśniewski & Zahorska, 2020). V roce 2002 byl zaveden jednotný test na konci primární školy a závěrečná zkouška na gymnáziu, o tři roky později státní maturita (Jakubowski, 2017). Testy měly zaručit srovnatelnou kvalitu vzdělávání a podpořit implementaci kurikulární reformy.

Test na konci primární školy hodnotil znalosti a dovednosti žáků v polštině, matematice a cizím jazyce a neměl závažný dopad na žáky (všichni přecházeli do spádového gymnázia) ani na školy (Jakubowski, 2008). Zkouška na konci gymnázia obsahovala jazykově-humanitní část a matematicko-přírodovědnou část a spolu se známkami na vysvědčení ovlivňovala přijímání žáků k dalšímu studiu, na školy formálně dopad neměla (Anczewska & Charzyńska, 2012), podobně jako maturita (*matura*). Veřejnost a zřizovatelé však vnímali výsledky gymnázií, která byla jako nová a kontroverzní součást systému vysoce sledována, jako ukazatele kvality výuky (Jakubowski, 2008; Wiśniewski & Zahorska, 2020), a proto vznikla při Cen-

trální zkušební komisi expertní skupina, která se s podporou projektu OECD zaměřila na vývoj indikátorů přidané hodnoty.

V roce 2006 byla pro gymnázia vytvořena aplikace, do níž mohly školy nahrát data svých žáků a spočítat si přidanou hodnotu. Tento tzv. kalkulátor vycházel z jednoduchého regresního modelu zohledňujícího výsledek testu v 6. ročníku,¹² pohlaví a dyslexii, socioekonomické proměnné nebyly v datech z testování dostupné (Jakubowski, 2008). O možnostech využívání dat o přidané hodnotě byly školy informovány na seminářích a v publikacích vydávaných expertní skupinou. V roce 2008 začaly být zveřejňovány centrálně spočítané indikátory přidané hodnoty gymnázií, které byly založeny na datech ze tří kohort, počítány víceúrovňovým regresním modelem s náhodným efektem pro konstantu (reprezentujícím přidanou hodnotu školy) a uváděny s 95% intervaly spolehlivosti (Žóttak, 2013). Kalkulátor pro výpočet jednoletých indikátorů zůstal školám k dispozici pro vlastní potřebu. Pro střední školy byly první (jednoleté) indikátory vytvořeny v roce 2010 po zavedení povinné maturity z polštiny a matematiky (tamtéž). Později byly nahrazeny tříletými indikátory počítanými podobně jako u gymnázií, ale zvláště pro všeobecně vzdělávací (*liceum*) a odborné (*technikum*) školy.

¹¹ Gymnázia měla zvýšit kvalitu vzdělávání zvláště ve venkovských oblastech. Měla více prostředků na vybavení, kvalifikovanější učitele a náročnější vzdělávací standardy než původní základní školy. Veřejnosti však byla vnímána rozporuplně.

¹² Pro lepší zachycení nelineárního vztahu mezi výsledkem na vstupu a výsledkem na výstupu byl výsledek v 6. ročníku vyjádřen polynomem až pátého stupně (viz Jakubowski, 2008).



Po nástupu konzervativní vlády byl v roce 2016 schválen školský zákon, který zavedl nové národní kurikulum a přehodnotil i mnoho dalších prvků původní reformy (Wiśniewski & Zahorska, 2020). Povinná školní docházka byla zkrácena na osm let a převedena zpět na základní školu, gymnázia byla zrušena. Gymnaziální zkouška byla nahrazena zkouškou na konci 8. ročníku, test pro 6. ročník byl zrušen, státní maturita zůstává v platnosti (Eurydice, 2019). Další plány současné polské vlády v oblasti měření přidané hodnoty se autorce textu nepodařilo zjistit.

3.5 Srovnání zemí

Hlavní parametry aktuálních přístupů k měření přidané hodnoty ve srovnávaných zemích jsou uvedeny v tabulce 1. Následující text shrnuje podobnosti a rozdíly mezi zeměmi s ohledem na srovnávací kritéria.

Vzdělávací stupeň pro měření přidané hodnoty. Určujícím kritériem pro volbu vzdělávacího stupně, na němž se sleduje přidaná hodnota, je dostupnost národně srovnatelných dat o znalostech a dovednostech žáků na výstupu i na vstupu. To komplikuje měření přidané hodnoty na prvním stupni, pro který obvykle chybí vstupní údaje. Přestože výpočet přidané hodnoty je možné založit také např. na známkách, ve všech zemích je patrná tendence spoléhat na standardizované testy a zkoušky. Známky vstupují pouze do norských modelů pro střední školy, ovšem pro korekci výsledkem žáků

v jednotných zkouškách. Anglický model pro primární školy zajišťuje srovnatelnost vstupního hodnocení prostřednictvím centrálně definovaných úrovní národního kurikula.

Vzdělávací výsledky, pro něž se zjišťuje přidaná hodnota. Vzdělávací oblasti, v nichž se sleduje přidaná hodnota, jsou dány obsahem národních testů a zkoušek. Zatímco testy hodnotí znalosti žáků v mateřském jazyce, matematice, popř. cizím jazyce, zkoušky zahrnují širší spektrum předmětů a kromě jednotné části mohou obsahovat i školní část. Země z obdobných dat vytvářejí různé varianty indikátorů přidané hodnoty, které jako závisle proměnnou uvažují průměrný výsledek z testovaných předmětů či závěrečných zkoušek, výsledek v jednotlivých předmětech či skupinách předmětů, výsledek v jednotné části zkoušky či prosté splnění zkoušky. Ve dvou srovnávaných zemích (Francie, Norsko) je na vyšší sekundární úrovni zjišťována také přidaná hodnota pro udržení žáků ve vzdělávání. To lze považovat za inspirativní rozšíření metody mimo oblast kognitivních výsledků.

Statistický model a kontrolní proměnné. Pro výpočet přidané hodnoty používají země různé modely, které se postupně vyvíjely. V tomto vývoji pozorujeme dvě protichůdné tendence – ve Francii a v Polsku postupný přechod k hierarchickým modelům, které jsou preferovány i v odborné literatuře jako lépe odpovídající datům, v Anglii a Norsku příklon k jednoúrovňovým modelům, u nichž je vyzdvihována jednoduchost a transparentnost. Pro kontrolu počátečních znalostí

Tab. 1 Měření přidávané hodnoty škol ve čtyřech evropských vzdělávacích systémech

Země	Název indikátoru přidávané hodnoty	Stupeň vzdělávání	Závisle proměnná	Model	Kontrolní proměnné	Nakládání s náhodnou chybou	Deklarované důvody pro volbu modelu
Anglie	Progress scores	Primární	Výsledek v testu ze čtení, psaní a matematiky v 6. ročníku (3 indikátory)	Lineární regrese	GPA* ze čtení, psaní a matematiky na konci 2. ročníku na základě učitelova zařazení žáka na úroveň národního kurikula	Nezveřejňování hodnot pro $N < 5$ Intervaly spolehlivosti	Transparentnost Srozumitelnost Vysoké nároky na všechny
	Progress 8	Nižší sekundární	GPA v 8 zkouškách GCSE v 10. a 11. ročníku	Lineární regrese	GPA z angličtiny a matematiky v testu na konci 6. ročníku	Nezveřejňování hodnot pro $N < 5$ Intervaly spolehlivosti	
	Level 3 value added	Vyšší sekundární	Známka dosažená ve vyšších kvalifikačních zkouškách	Lineární regrese v rámci jedné kvalifikační zkoušky	GPA v 8 zkouškách GCSE v 10. a 11. ročníku	Nezveřejňování hodnot pro $N < 5$ Intervaly spolehlivosti	
Francie	Indicateurs de valeur ajoutée des lycées	Vyšší sekundární	Pravděpodobnost složení závěrečné zkoušky <i>baccalauréat</i> Pravděpodobnost postupu ke zkoušce Pravděpodobnost složení zkoušky s vyznamenáním (3 indikátory)	Víceúrovňová logistická regrese v rámci skupiny oborů	GPA z francouzštiny, matematiky, přírodovědného a společenského vzdělání na základě DNB ve zkoušce DNB na konci 9. ročníku Věk Pohlaví Socioekonomický index	Nezveřejňování hodnot pro $N < 20$ (všeobecné/technické obory), resp. $N < 10$ (odborné obory) Nezveřejňování hodnot, pokud více než 50% žáků nemá výsledek z DNB	Srovnatelnost mezi školami Zlepšování díky porovnání



Norsko	Skolebidsrag-sindikatorer	Základní	GPA ze čtení, matematiky a angličtiny v testu v 8. ročníku	Jednoúrovňová nelineární regrese	GPA ze čtení, matematiky, angličtiny v testu v 5. ročníku Vzdělání rodičů Příjem domácnosti Imigrační status	Nezvěřejňování hodnot pro $N < 20$ Intervaly spolehlivosti Spojení dvou kohort	Transparentnost Srozumitelnost Využití pro zlepšování výuky
		Nižší sekundární	Výsledek z písemné zkoušky v 10. ročníku	Jednoúrovňová nelineární regrese	GPA ze čtení, matematiky a angličtiny v testu v 8. ročníku Vzdělání rodičů Příjem domácnosti Imigrační status	Nezvěřejňování hodnot pro $N < 20$ Intervaly spolehlivosti Spojení dvou kohort	
Polsko ^b	Edukacyjna wartość dodana	Vyšší sekundární	Průměr známek Dokonceení ročníku Účast na vzdělávání (3 indikátory)	Jednoúrovňová nelineární/logistická regrese v rámci ročníku a vzdělávacího programu	Průměr známek z předchozího ročníku Pohlaví	Nezvěřejňování hodnot pro $N < 20$ Intervaly spolehlivosti	Transparentnost Využití pro autoevaluaci (jednoleté) Robustnost (tříleté)
		Nižší sekundární	Výsledek v humanitní a přírodovědné části zkoušky na konci 9. ročníku (2 indikátory)	Jednoúrovňová nelineární regrese (jednoleté) / víceúrovňová regrese (tříleté)	GPA v testu na konci primární školy Pohlaví Dyslexie	Intervaly spolehlivosti Spojení tří kohort	
		Vyšší sekundární	Výsledek z písemných maturitních zkoušek z polštiny, matematiky, humanitních a přírodovědných předmětů (4 indikátory)	Víceúrovňová regrese v rámci typu školy	Výsledek v humanitní, resp. matematické části závěrečné zkoušky na konci 9. ročníku Pohlaví Dyslexie	Intervaly spolehlivosti Spojení tří kohort	

^a GPA = průměrné bodové skóre (grade point average)

^b Stav před reformou z roku 2016

a dovedností se většinou používá průměrný výsledek na konci předchozího období, pouze Polsko zohledňuje předchozí výsledky v příslušné skupině předmětů. Žádná ze srovnávaných zemí nezařazuje do modelu více předchozích výsledků. S výjimkou Anglie modely obsahují i další kontrolní proměnné včetně socioekonomického statusu. V Anglii je absence sociodemografických proměnných terčem odborné kritiky (Gorard, 2006; 2008; Leckie & Goldstein, 2017; 2019). V Polsku, kde nejsou data o sociálním zázemí žáků dostupná, je pomalejší pokrok slabších (často sociálně znevýhodněných) žáků modelován nelineární regresí. Tu využívá také Norsko. U středních škol nejsou rozdíly mezi obory zohledňovány pomocí kontrolních proměnných, ale tvorbou samostatných modelů pro různé typy škol.

Sdělování a využívání informací o přidané hodnotě. Přidaná hodnota škol je ve zkoumaných zemích zveřejňována prostřednictvím dashboardů (interaktivních nástrojů pro zobrazování dat), které jsou veřejně přístupné přes webové rozhraní a obsahují i další informace o školách.¹³ Školy nejsou řazeny do žebříčků, dashboard zobrazí pouze data pro jednu školu, kterou uživatel zadá či vybere z nabídky. Země činí i další opatření bránící přímému srovnávání škol a zdůrazňují nejistotu spojenou s vypočtenou hodnotou pomocí intervalů spolehlivosti. Žádná země nezveřejňuje údaje pro školy, kde by byl odhad přidané hodnoty založen na příliš malém počtu žáků.

Ve všech srovnávaných zemích je kvalita škol posuzována pomocí širší palety indikátorů, vedle přidané hodnoty jsou nadále sledovány dosažené výsledky, ale i další charakteristiky týkající se školních zdrojů, složení žáků, pokračování žáků ve vzdělávání apod. V žádné ze sledovaných zemí nemají údaje o přidané hodnotě zásadní dopad na školy. V Anglii, kde byl ještě při poslední aktualizaci indikátorů nízký výsledek spolu s nízkou přidanou hodnotou důvodem k zařazení školy do intervenčního programu, byly v roce 2019/2020 minimální požadavky na výkon škol zrušeny a jediným důvodem k intervencím je negativní hodnocení školní inspekci. V ostatních zemích byly údaje o přidané hodnotě od počátku koncipovány jako podklad pro autoevaluaci a vnitřní rozvoj škol.

DISKUSE

Ve všech srovnávaných zemích je patrná snaha o férové hodnocení škol a měření přidané hodnoty je považováno za vhodnou metodu. Při zavádění měření přidané hodnoty sehrál důležitou roli projekt OECD (2008), jenž přispěl ke sblížení představ o žádoucích přístupech k hodnocení (Grek, 2010). Funkční začlenění indikátorů přidané hodnoty do vzdělávacích systémů by však nebylo možné bez zapojení lokálních aktérů, kteří se podíleli na tvorbě a vývoji modelů (Dvořák, 2015). Jako velmi důležité se ukazuje nejen vytvoření pracoviště zodpovědného za návrh modelů, jejich ověření a později

¹³ Anglie: www.compare-school-performance.service.gov.uk; Francie: www.education.gouv.fr/recherche-ival; Norsko: www.skoleporten.udir.no (od července 2021 provoz ukončen, pilotuje se nové rozhraní *analysebrett* na webu udir.no); Polsko: ewd.edu.pl



každoroční zpracování dat, ale také dialog s akademickým výzkumem, který přispívá ke kritické reflexi stávajících přístupů. Výzkum v oblasti měření přidané hodnoty není završen a ani v mezinárodní komunitě nepanuje shoda na jednom dominantním přístupu (Everson, 2017; Levy et al., 2019), který by bylo možné převzít jako hotové řešení.

Zdrojem dat pro výpočet přidané hodnoty jsou zpravidla standardizované testy a zkoušky, jejichž kvalita má zásadní význam pro jakékoli výpovědi o přidané hodnotě. Jejich platnost může zpochybňovat jak úzký záběr standardizovaných testů (typicky matefský jazyk, matematika, případně cizí jazyk), tak problematické psychometrické vlastnosti či nesoulad testů s požadavky kurikulárních dokumentů. V kontextu České republiky, kde nedostatečné personální a finanční zajištění institucí zodpovědných za tvorbu testů vyvolává v části odborné veřejnosti pochybnosti o výpovědní hodnotě testů a jejich přínosu pro žáky, školy i řízení vzdělávacího systému, by bylo zapotřebí věnovat nejprve pozornost přípravě kvalitního hodnocení žákovských znalostí a dovedností, než začneme uvažovat o možnostech jeho dalšího využívání. K hodnocení výkonu vzdělávacího systému slouží v ČR často mezinárodní srovnávací šetření (PISA, TIMSS apod.), která jsou široce přijímána jako zdroj spolehlivých dat o znalostech a dovednostech žáků. Pokud by se v rámci národního navýšení vzorku podařilo zajistit, aby byly pro tytéž žáky k dispozici výsledky ve dvou různých ročnících, bylo by možné mezinárodní šetření využít k pilotnímu měření přidané hodnoty. Pro integraci metody do systému evaluace škol jsou nicméně – tak jako

v uvedených evropských zemích – potřeba kvalitní plošná data, která poskytnou údaje o všech školách.

Přestože měření přidané hodnoty lze technicky realizovat na všech vzdělávacích stupních, nejsnadnější je v povinném vzdělávání. U středních škol je třeba vyrovnat se mj. s odlišným složením žáků v různých typech škol, s možností volby maturitních předmětů a s předčasnými odchody ze vzdělávání. Pro Českou republiku by proto bylo vhodnější nezačínat s měřením přidané hodnoty na středních školách, být pro ně budou k dispozici data. Navíc spárování dat z přijímacích a maturitních testů, které mají jinak nastavenou obtížnost, by mohlo vést k podhodnocení přidané hodnoty škol, které přijímají žáky s vysokou úrovní vstupních znalostí a dovedností.

V žádném ze srovnávaných zemí se kvůli měření přidané hodnoty nesbírají nová data, zpracovávají se ta, která jsou v systému už dostupná. To s sebou nese riziko vychýlení odhadů v důsledku chybějících proměnných (Perry, 2016). Ve všech zkoumaných zemích je zřetelná tendence zohlednit také sociodemografické proměnné, a zvláště socioekonomický status, který koreluje nejen s dosaženými výsledky, ale i s pokrokem žáků v učení (Harris, 2011; Leckie & Goldstein, 2019). Nedostupnost dat o sociálním zázemí žáků je také v ČR vážným rizikem pro férovost hodnocení vzdělávacího přínosu škol.

Ve všech srovnávaných zemích je měření přidané hodnoty prezentováno jako nástroj podporující zlepšování výuky ve školách, ale reálné zkušenosti jsou méně optimistické. Například ve Francii ještě po dvaceti letech od zavedení prvních indikátorů přidané hod-



noty většina učitelů nerozuměla tomu, co měří, a neuměla je odlišit od jiných ukazatelů kvality vzdělávání (Bajou, Paulin-Moulard & Bossard, 2015). Obtíže s porozuměním přidané hodnotě byly pozorovány také v Anglii (Kelly & Downey, 2010). Obecně se ukazuje, že už samotné výsledky žáků v testech jsou pro učitele příliš vzdálené od každodenní práce ve třídách a málo užitečné pro její zlepšování (Pons, 2012; Tveit, 2014).

ZÁVĚR

Indikátory přidané hodnoty jsou vnímány jako spravedlivý ukazatel vzdělávacího přínosu škol, s jejich zaváděním jsou však spojena mnohá rizika. Vážně by mělo být bráno zejména riziko nedostatečných kapacit jak pro implementaci a rozvoj vlastního měření přidané hodnoty, tak pro tvorbu kvalitních testů coby zdroje vstupních dat. Reálnou hrozbou je rovněž vychýlení odhadů v důsledku chybějících proměnných

nebo malá využitelnost indikátorů ke zlepšování výuky ve třídách. Podcenění těchto rizik by mohlo nadlouho narušit důvěru ve smysluplnost a férovost měření přidané hodnoty, jak se to stalo například ve Francii (Pons, 2012).

Závěry práce jsou omezeny tím, že vycházejí z analýzy dokumentů, které přinášejí pouze povrchní vhled do procesů spojených s měřením přidané hodnoty. Stranou byly ponechány nezamýšlené dopady jako podryívání profesního sebepojetí učitelů (Mons, 2009) či rezistence škol (Fuller, 2019). Přitom rezistenci škol lze v českém prostředí s jeho skepsí vůči testování jistě očekávat. I pro měření přidané hodnoty platí, že efektivita jakéhokoli systému hodnocení škol závisí nejen na systému jako takovém, ale především na míře jeho akceptování školami, na „datové gramotnosti“ škol a v neposlední řadě na připravenosti vzdělávacího systému poskytnout školám metodickou podporu (Penninckx et al., 2017).

LITERATURA

- Anczewska, M., & Charzyńska, K. (2012). Educational assessment of pupils in Poland. *SA-eDUC Journal*, 9(1).
- Bajou, B., Paulin-Moulard, F., & Bossard, T. (2015). *Des facteurs de valeur ajoutée des lycées*. Ministère de l'éducation nationale de l'enseignement supérieur et de la recherche.
- Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, 29(1), 37–65.
- Bereday, G. (1964). *Comparative method in education*. Holt, Rinehart and Winston.
- Bradbury, A. (2011). Equity, ethnicity and the hidden dangers of 'contextual' measures of school performance. *Race, Ethnicity and Education*, 14(3), 277–291.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. U.S. Government Printing Office.
- ČŠI. (2019). *Zpráva o vybraných zahraničních přístupech k monitorování spravedlivosti vzdělávacího systému*. Dostupné z www.csicr.cz/cz/Dokumenty/Publikace-a-ostatni-vystupy/Zprava-o-vybraných-zahranicnich-pristupech-k-monit



- DfE. (2019a). *Primary school accountability in 2019: Technical guide. A technical guide for maintained secondary schools, academies and free schools*. Dostupné z www.gov.uk/government/publications/primary-school-accountability
- DfE. (2019b). *16 to 18 accountability measures: Technical guide*. Dostupné z www.gov.uk/government/publications/16-to-19-accountability-headline-measures-technical-guide
- DfE. (2020). *Secondary accountability measures. Guide for maintained secondary schools, academies and free schools*. Dostupné z www.gov.uk/government/publications/progres-8-school-performance-measure
- DEPP. (2020). *Les indicateurs de résultats des lycées publics et privés sous contrat. Baccalauréats général, technologique et professionnel 2019*. Dostupné z www.education.gouv.fr/methodologie-des-indicateurs-de-resultats-des-lycees-11948
- Duclos, M., & Murat, F. (2014). Comment évaluer la performance des lycées? Un point sur la méthodologie des IVAL (Indicateurs de valeur ajoutée des lycées). *Éducation et formations*, 85, 73–84.
- Dvořák, D. (2015). Proměny kurikulárního diskurzu: o aktérech, standardech a ledních medvědech. In D. Greger (Ed.), *Srovnávací pedagogika: Proměny a výzvy* (s. 101–117). Pedagogická fakulta UK.
- Elstad, E. (2009). Schools which are named, shamed and blamed by the media: School accountability in Norway. *Educational Assessment, Evaluation and Accountability*, 21(2), 173–189.
- Eurydice (2009). *National testing of pupils in Europe: Objectives, organisation and use of results*. Education, Audiovisual and Culture Executive Agency.
- Eurydice (2019). *National education systems: Poland* (online). Dostupné z https://eacea.ec.europa.eu/national-policies/eurydice/national-description_en
- Evain, F., & Evrard, L. (2017). Une meilleure mesure de la performance des lycées: Refonte de la méthodologie des IVAL (session 2015). *Éducation et formations*, 94, 91–116.
- Everson, K. C. (2017). Value-added modelling and educational accountability: Are we answering the real questions? *Review of Educational Research*, 87(1), 35–70.
- Felouzis, G. (2005). Performances et „valeur ajoutée“ des lycées: le marché scolaire fait des différences. *Revue française de sociologie*, 46(1), 3–36.
- Fuller, K. (2019). “That would be my red line”: an analysis of headteachers’ resistance of neoliberal education reforms. *Educational Review*, 71(1), 31–50.
- Gorard, S. (2006). Value-added is of little value. *Journal of Educational Policy*, 21(2), 233–241.
- Gorard, S. (2008). The value-added of primary schools: What is it really measuring? *Educational Review*, 60(2), 179–185.
- Grek, S. (2010). International organisations and the shared construction of policy ‘problems’: Problematisation and change in education governance in Europe. *European Educational Research Journal*, 9(3), 396–406.
- Hanushek, E. (1971). Teacher characteristics and gains in student achievement: Estimation using micro-data. *American Economic Review*, 61(2), 280–288.
- Harris, D. (2011). *Value-added measures in education: What every educator needs to know*. Harvard Education Press.



- Hatch, T. (2013). Beneath the surface of accountability: Answerability, responsibility and capacity-building in recent education reforms in Norway. *Journal of Educational Change*, 14(2), 113–138.
- Hægeland, T., Kirkebøen, L. J., Raaum, O., & Salvanes, K. G. (2005). *Skolebidragsindikatorer. Beregnet for avgangskarakterer fra grunnskolen for skoleårene 2002–2003 og 2003–2004*. Statistisk entralbyrå.
- Hægeland, T., Kirkebøen, L. J., Bratsberg, B., & Raaum, O. (2011). *Value added indikatorer. Et nyttig verktøy i kvalitetsvurdering av skoler?* Statistisk entralbyrå.
- Jakubowski, M. (2008). *Implementing value-added models of school assessment. EUI working paper RSCAS 2008/06*. European University Institute.
- Jakubowski, M. (2017). *Accountability in the Polish education system. Country case study prepared for the 2017/8 Global Education Monitoring Report*. UNESCO.
- Ježková, V., Dvořák, D., Chapman, Ch., et al. (2010). *Školní vzdělávání ve Velké Británii*. Karolinum.
- Jones, B. (2007). The unintended outcomes of high-stakes testing. *Journal of Applied School Psychology*, 23(2), 65–86.
- Kelly, A., & Downey, C. (2010). Value-added measures for schools in England: looking inside the 'black box' of complex metrics. *Educational Assessment, Evaluation & Accountability*, 22(3), 181–198.
- Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modelling: A review. *Economics of Education Review*, 47, 180–195.
- Leckie, G., & Goldstein, H. (2017). The evolution of school league tables in England 1992–2016: 'Contextual value-added', 'expected progress' and 'progress 8'. *British Educational Research Journal*, 43(2), 193–212.
- Leckie, G., & Goldstein, H. (2019). The importance of adjusting for pupil background in school value-added models: A study of Progress 8 and school accountability in England. *British Educational Research Journal*, 45(3), 518–537.
- Lenkeit, J. (2013). Effectiveness measures for cross-sectional studies: A comparison of value-added models and contextualized attainment models. *School Effectiveness and School Improvement*, 24, 39–63.
- Levy, J., Brunner, M., Keller, U., & Fischbach, A. (2019). Methodological issues in value-added modeling: An international review from 26 countries. *Educational Assessment, Evaluation, and Accountability*, 31, 257–287.
- Malach, J., Malčík, J., & Krpec, R. (2011). Přidaná hodnota vzdělávání a odpovědnost učitelů a škol. In B. Niemierko & M. K. Szmigel (Eds.), *Ewaluacja w edukacji: Koncepcje, metody, perspektywy* (s. 97–111). Polskie towarzystwo diagnostyki edukacyjnej.
- Mons, N. (2009). *Theoretical and real effects of standardized assessment*. Eurydice.
- MŠMT. (2014). *Strategie vzdělávací politiky České republiky do roku 2020*. Dostupné z www.msmt.cz/uploads/Strategie_2020_web.pdf
- MŠMT. (2015). *Dlouhodobý záměr vzdělávání a rozvoje vzdělávací soustavy České republiky na období 2015–2020*. Dostupné z www.msmt.cz/file/35167/



- MŠMT. (2020). *Strategie vzdělávací politiky České republiky do roku 2030+*. Dostupné z www.msmt.cz/vzdelavani/skolstvi-v-cr/strategie-2030
- Nusche, D., Earl, L., Maxwell, W., & Shewbridge, C. (2011). *OECD reviews of evaluation and assessment in education – Norway*. OECD Publishing.
- OECD. (2008). *Measuring improvements in learning outcomes*. Best practices to assess the value-added of schools. OECD Publishing.
- Penninckx, M., Quintelier, A., Vanhoof, J., De Mayer, S., & Van Petegem, P. (2017). Delphi study on standardized systems to monitor student learning outcomes in Flanders: Mechanisms for building trust and/or control? *Studia paedagogica*, 22(2), 9–31.
- Perry, T. (2016). English value-added measures: Examining the limitations of school performance measurement. *British Educational Research Journal*, 42(6), 1056–1080.
- Pons, X. (2012). De nouvelles connaissances pour une nouvelle gouvernance par les résultats? L'exemple de l'éducation en France et de trois outils d'évaluation. *Recherches sociologiques et anthropologiques*, 43(2), 15–34.
- Pons, X. (2020). National testing in education in France: Statisation, rationalisation and politicisation. *European Educational Research Journal*, 19(2), 109–124.
- Ray, A. (2006). *School value-added measures in England*. A paper for the OECD project on the development of value-added models in education systems. Department for Education and Skills.
- Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, Ch., & Stringfield, S. (2014). Educational effectiveness research (EER): A state-of-the-art review. *School Effectiveness and School Improvement*, 25(2), 197–230.
- Rocher, T. (2016). Construction d'un indice de position sociale des élèves. *Éducation et formations*, 90, 5–27.
- Ryška, R. (2009). *Evaluace a přidaná hodnota ve vzdělávání*. Pedagogická fakulta UK.
- SSB. (2018). *Metodebeskrivelse for skolebidragsindikatorer*. Dostupné z www.udir.no/tall-og-forskning/statistikk/statistikk-videregaende-skole/skolebidragsindikatorer-for-videregaende-skoler/
- SSB. (2019). *Datagrunnlag og metode for beregning av skolebidrags-indikatorer for grunnskoler*. Dostupné z www.udir.no/tall-og-forskning/statistikk/statistikk-grunnskole/skolebidragsindikatorer-for-grunnskoler/
- Steffensen, K., Ekren, R., Zachrisen, O. O., & Kirkebøen, L. J. (2017). *Er det forskjeller i skolers og kommuners bidrag til elevenes læring i grunnskolen? En kvantitativ studie*. Statistisk entralbyrå.
- Středisko vzdělávací politiky (2008). *Zjišťování přidané hodnoty. Celková zpráva projektu LS0602*. Pedagogická fakulta UK.
- Tekwe, C. D., Carter, R. L., Ma, C.-X., Algina, J., Lucas, M. E., Roth, J., ... & Resnick, M. B. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics*, 29(1), 11–36.
- Tveit, S. (2014). Educational assessment in Norway. *Assessment in Education: Principles, Policy & Practice*, 21(2), 221–237.



- Vlček, P. (2015). Srovnávací výzkum v pedagogice: některé úvahy o metodologii problémového přístupu. *Pedagogická orientace*, 25(3), 394–412.
- Wiśniewski, J., & Zahorska, M. (2020). Reforming education in Poland. In F. Reimers (Ed.), *Audacious education purposes* (s. 181–208). Springer.
- Žóltak, T. (2013). EWD jako způsob badania efektywności szkół. In M. Karwowski (Ed.), *Ścieżki rozwoju edukacyjnego młodzieży – szkoły pogymnazjalne* (s. 27–46). Wydawnictwo IFIS PAN.

Mgr. Eva Potužníková

Univerzita Karlova, Pedagogická fakulta, Ústav výzkumu a rozvoje vzdělávání;

e-mail: eva.potuznikova@pedf.cuni.cz

POTUŽNÍKOVÁ, E. Evaluation of Schools' Contribution to Pupils' Learning by Schools' Value-Added: Comparison of Four European Countries

Aims: *Based on a comparison of four European education systems (England, France, Norway, and Poland), the aim of the study is to identify the benefits and risks related to school evaluation using value-added assessment.*

Methods: *Value-added systems are compared according to a predefined set of criteria. Data sources include technical documentation of value-added models, web portals used to publish the results, and texts on the development of models and the use of value-added data.*

Results: *Pupils' results in standardised tests or exams are the preferred input data for value-added computation. Data availability determines both the choice of school level and the domains (subjects) to be assessed and the choice of contextual variables. The countries try to avoid direct comparison of schools, do not rank schools by value-added terms, and communicate statistical uncertainty related to the measurement. Information on the value-added serves as feedback to schools.*

Conclusions: *Value-added indicators are fair measures of schools' contribution to pupils' learning, but their introduction may be associated with risks, such as the omission of important variables because of data unavailability, lack of capacity to implement and develop the method, and limited utility in the school improvement process.*

Keywords: *value-added measurement, standardised assessment, school quality, comparison, England, France, Norway, Poland*