

# Diskuse

---

---

## Jak lze výsledky žáků z externích zkoušek použít pro podporu jejich rozvoje?

Bolesław Niemierko

Počátkem 21. století začali být žáci polských základních, nižších sekundárních (gymnasií) a vyšších sekundárních (lyceí) škol přezkušováni externími testovacími centry ze základních kognitivních dovedností a znalostí předmětu. Výsledky testů jsou předloženy žákům a vedení příslušných škol v podobě škály dle hrubého skóru, procenta úspěšnosti a staninových škál, které jim umožní usilovat o vzdělání na vyšší úrovni a o místa v každoročně zveřejňovaných „tabulkách školní ligy“. Tento postup je užitečný při procesu výběru žáků a při evaluaci výsledků vzdělávání, ale je málo přínosný pro každodenní činnost učitelů, ve které je klíčové flexibilní řízení žákova učení. Učitelé musí pracovat s žáky, jací *jsou*, nikoli jací *by měli být* na daném stupni školy. Musí podporovat rozvoj všech žáků bez ohledu na jejich schopnost dosáhnout předepsaných vzdělávacích standardů.

Autor příspěvku doporučuje aplikovat přístup *přidané hodnoty* jako obecný rámec pro průběžné hodnocení a *ročníkové ekvivalenty* (grade equivalents – GE)<sup>1</sup> jako měřítko pro sledování růstu žáků a jejich postupu po jednotlivých stupních vzdělávacího systému. Příspěvek popisuje možnosti a meze tohoto nového postupu a také nejdůležitější výsledky pilotní studie.

---

<sup>1</sup>Pozn. překl.: překlad anglického termínu *grade equivalents* jako *ročníkové ekvivalenty*, resp. *standardizované ročníkové ekvivalenty*, zde vychází z pojetí dle Educational Measurement Group of Pearson ([http://www.pearsonedmeasurement.com/research/faq\\_2d.htm](http://www.pearsonedmeasurement.com/research/faq_2d.htm)).

## Systém externích zkoušek v Polsku

Externí zkoušky jsou součástí probíhající reformy vzdělávacího systému v Polsku a jsou stále ještě novým jevem. Byly koncipovány zejména jako prostředek k zajištění srovnatelnosti (1) vzdělávacích výsledků mezi školami na stejném stupni a (2) požadavků k přijetí na školy vyššího stupně.

Bylo zřízeno osm zkušebních komisí a Centrální testovací výbor ve Varšavě, které byly pověřeny přípravou, prováděním a certifikací celostátního zkoušení. Existují tři úrovně zkoušek:

1. Po ukončení šestileté základní školy – zkouška v celkové délce jedné hodiny ze čtení, psaní, matematického uvažování a studijních dovedností.
2. Po ukončení tříleté nižší sekundární školy (gymnasia) – oddělený test z humanitních předmětů a další test z matematiky a přírodních věd.
3. Testy po ukončení tříleté vyšší sekundární školy (lycea, matura) – testy: (1) z polského jazyka a literatury, (2) z cizího (především anglického) jazyka, (3) z jednoho volitelného předmětu (od roku 2010 bude povinná matematika).

## Jak vnímá systém zkoušek veřejnost?

První názory veřejnosti na uvedený systém byly spíše negativní, především na úrovni základní školy. Zejména rodiče a novináři značně zveličovali nebezpečí *testové úzkosti* a *intelektuální bezmocnosti* u třináctiletých a rigidnost a krutost procesu měření. Avšak poté, co vešel ve známost obsah testů a podoba testových položek, objevila se zcela opačná a výraznější hrozba: považování výsledků testů za jediný a neoddiskutovatelný doklad kvality výuky a kvality řízení v příslušných školách.

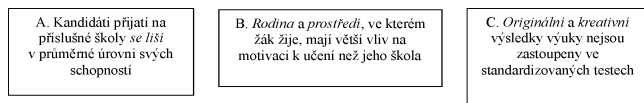
Stereotypy „dobrých“ a „špatných“ škol, takto označených na základě externích zkoušek, mají pevné kořeny a několik racionálních premis. Můžeme je podhalit pomocí metody zvané *rozbalování* (Haertel, 1991) – viz obr. 1.

Vágní zobecnění umístěné v horní části obr. 1, které zpochybňuje většina odborníků, je podpořeno dvěma očividně přesvědčivými tvrzeními umístěnými ve střední části, jež závisejí na premisách umístěných ve spodní části obrázku, které jsou nejen bezvýhradně pravdivé, ale také zásadní pro školské systémy a pedagogické měření. Rozbalováním původního názoru jsme došli k jeho pravdivým základům. Avšak pokud obrátíme šipky a směr odvozování, zjistíme, že pravdivé argumenty vedou naprosto nelogicky k nejistému závěru.



Obrázek 1: Rozbalování domněnky, že výsledky externích zkoušek odrážejí kvalitu vzdělávání.

Co způsobuje tento rozpor? Zjednodušení, redukce počtu výchozích premis. Když jsme zohlednili několik dalších premis, jak je uvedeno na obr. 2, odvozování „nahoru“ na obr. 1 by bylo obezřetnější a tvrzení v horní části obrázku by se logicky ukázalo jako místně podmíněné a ne obecně pravdivé.



Obrázek 2: Opomíjené premisy k závěru o vlivu externích zkoušek na kvalitu výuky

Premisy A, B, C na obr. 2, které jsou opomíjeny laickou veřejností při hodnocení školního vzdělávání, jsou předmětem nesčetných studií z oblasti pedagogiky, sociologie a pedagogického výzkumu, jež jsou veřejnosti jen málo známé. Metoda rozbalování odhaluje neúplnost argumentů postavených na předem zvolených postřezích a úvahách. Vzdělávání je příliš složité na to, aby bylo posuzováno jediným souborem měřítek, jakkoli jsou teoreticky správná a administrativně propracovaná. Žádný ukazatel žákova výkonu sám o sobě nepostačuje k zhodnocení jeho schopností a činnosti jeho školy.

## Sumativní a formativní hodnocení

V čem se externí zkouška primárně liší od interního přezkoušení žákových výkonů? Rozdíl je chápán jako alternativa sumativního a formativního hodnocení (Scriven, 1967). Zatímco sumativní hodnocení je navrženo a použito

váno k informování o dosahovaných výkonech na konci výukového období, formativní hodnocení se přímo podílí na zlepšování výsledků učení a mělo by doprovázet proces výuky. První z uvedených hodnocení je nezbytné zejména pro certifikaci žáků a zodpovědnost školy (Koretz a Hamilton, 2006), druhé je nezbytné pro podporu učení žáků prostřednictvím mechanismu *zpětné vazby*.

Zatímco hodnocení v rámci školy by mohlo být – a v mnoha zemích již mnoho let je – používáno k monitorování učení žáků a k certifikování jejich výkonů, rozvíjecí potenciál externích zkoušek a jiných typů standardizovaného testování je stále sporný. Bylo učiněno mnoho pokusů, jakým způsobem by mohly být výsledky externích zkoušek učitelům a žákům sděleny. Nebylo ale dosaženo v podstatě ničeho víc než občasných podnícení k většímu úsilí pomocí cvičných zkoušek.

Hlavním omezením v komunikaci zkoušejícího se zkoušeným a v úspěšné rozvíjecí funkci externích zkoušek je možná jazyk, jímž jsou prezentovány výsledky testů, zejména problém *škálování*. Dříve než v tomto příspěvku nabídnu novou měřicí škálu pro tento účel, zmíním se krátce o škálách, které jsou v současnosti používány zkušebními komisemi v Polsku.

## Procento úspěšnosti a standardizované škály

Když je hrubý skór, tedy počet bodů získaných v testu, posuzován samostatně, poskytuje jen málo informací o kvalitě dosahovaných výkonů. Jeho význam závisí na délce testu, způsobu hodnocení, typu položek (hádání!) a jejich obtížnosti a na administraci testu. Některé formální podmínky jsou vyrovnány, když se spočítá procento úspěšnosti, což je druh odvozeného skóru. Jednoduchost rozdělení hrubého skóru dle počtu dostupných bodů (tj. délkou škály) ho činí nesmírně populárním mezi žáky, rodiči, novináři a ostatními neodborníky.

Odborníci však nemají procento úspěšnosti v oblibě kvůli jeho skrytým předpokladům: 100% úspěšnost na horní hranici, 0% úspěšnost na dolní hranici a jednotka 1 %, domněle rovnocenná v celé škále (linearita škálování). Doporučují normalizované standardní škály, ve kterých je rozložení testových výsledků převedeno do zvonovitého formátu (Gaussova křivka) a skóry nám sdělí, jak daleko je dosažený výkon od aritmetického průměru normativní skupiny. Na základě vzorců a postupů zveřejňovaných ve většině běžně dostupných knih o pedagogickém měření (Niemierko, 1999, s. 271 až 275) a nesčetným cvičením v kurzech testování pro zkoušející personál se

v Polsku stala oblíbenou devítibodová staninová škála (s průměrnou hodnotou 5 a směrodatnou odchylkou 2).

Staniny mají své výhody, např. srovnatelnost mezi testy a různými okruhy učiva, rozumnou délku škály (pro testy s omezenou reliabilitou), jednočíslicové symboly (1–9). Mají nicméně i své nevýhody, jako jsou sofistikovaný odhad normálního rozložení, náročné statistické postupy a příliš hrubý vzorek pro výběr (až 20 % populace). Poslední jmenovaná nevýhoda způsobuje, že pro závěrečnou zkoušku na středních školách v Polsku („matura“) se nyní uvažuje o *akademické škále* (škála CEEB = College Entrance Examination Board – Komise pro přijímací zkoušky na vysoké školy – pozn. překl., s průměrnou hodnotou 500 a směrodatnou odchylkou 100).

### Potřeba sociálního hodnocení

Hrubé skóry, procento úspěšnosti a staninové skóry neuspokojují potřeby formativního hodnocení v každodenní práci učitele, zejména, pokud jde o slabší žáky. Jak poznamenala Susan Brookhart (1993), „v hodnocení jsou dva standardy spravedlnosti: průměrný a nadprůměrný žák získá, „co si zaslouží“, ale slabší žák dostane „šanci“, pokud to může být nějak zdůvodnitelné“. Takto se objevuje druhý systém kurikulárních standardů (Niemierko, 2002, kap. XIV; Szyling, 2007), který způsobuje, že standardy jsou v rámci jednotlivých ročníků systematicky vyšší než adekvátní normy dané kurikulárně validními testy.

Při hodnocení v rámci školy se zohledňuje *úsilí* žáků stejně jako jejich *schopnost, prostředí*, ve kterém žijí, *vzdělávací a profesní plány* a další okolnosti. Tyto kontextové proměnné dělají ze školního hodnocení hodnocení „sociální“, které je z politického a psychologického hlediska hodnotnější než externí zkoušky. Takové hodnocení podporuje zlepšování slabších žáků a je evidentně flexibilnější než jakákoli výuka a evaluace postavená na standardech.

Jak by se mohly dosahované výsledky externích zkoušek stát *sociálními* v tom smyslu, aby povzbuzovaly každého žáka dle jeho možností k afektivnímu a kognitivnímu rozvoji? Vhodné řešení problému je chápání výsledku zkoušky jako vlastního výkonu (self-referenced) na rozdíl od absolutního výkonu (criterion-referenced) nebo relativního výkonu (norm-referenced) (Nitko, 1983, s. 341). To znamená, že dřívější (výchozí) úroveň žákových znalostí a dovedností musí být zohledněna při hodnocení stávající úrovně

žákem dosažených výsledků. Tento přístup nás přivádí k hodnocení s přidanou hodnotou.

## Hodnocení s přidanou hodnotou pro žáky a školy

*Přidaná vzdělávací hodnota* znamená nárůst v dosahovaných výkonech žáka nebo skupiny žáků v daném období ve vybrané oblasti (výkonu). Tento termín, převzatý z ekonomie, zavádí *rozvíjející přístup* ve vzdělávání, neboť pokrok žáka v po sobě následujících stupních vzdělávání je chápán jako *přidaná hodnota*.

Ačkoli „pokrok žáka“ vzbuzoval u učitelů a rodičů pozornost od samého počátku naší civilizace, přidaná hodnota, nejprve aplikovaná spíše na školy než na jednotlivé žáky, je jednou ze „změn v testování počínajících v druhé polovině 80. let“ (Koretz a Hamilton, 2006, s. 534). Hlavní důvody pro toto zdržení jsou dvojí:

1. Převládající *vzdělávací model* typu *třída–předmět–vyučovací hodina*. Učitelé se vyhýbají testování obsahů, které předcházejí nebo následují po té části učiva, již právě vyučují. Žáci mohou věnovat málo pozornosti obsahu testu, nebo dokonce zpochybňovat jeho relevanci, pokud není test zaměřen na jejich momentální dosaženou úroveň (Kolen, 2006).
2. *Nízká reliabilita* měření změny, které je vždy postaveno na dvou skórech zatížených odchylkami měření. Ronald Hambleton, přední představitel psychometrie, komentoval svoji nechuť zabývat se měřením růstu slovy: „Historie technických parametrů přírůstkových skóru je dlouhá a smutná.“ (1978, s. 284)

Bude moderní historie přidané hodnoty ve vzdělání veselá a se šťastným koncem? Krystyna Szmigel a Anna Rappe potvrdily v několika navazujících výzkumech příznivý postoj ředitelů polských středních škol vůči tomuto novému ukazateli kvality výuky. Jednu ze svých zpráv (2005) zakončují tímto vyjádřením: „Pokud jsou informace o přidané hodnotě moudře používány, mohou představovat nástroj vzdělávací politiky, jenž by mohl zabránit nekontrolované selekci žáků na úrovni populace, školy a třídy. Takové informace by umožnily vystoupit ze stínu určitým školám a regionům, jejichž umístění je dle výsledků externích zkoušek slabé.“

Snadnost interpretace a „spravedlnost“ v očích neprivilegovaných žáků, škol a regionů dělají z koeficientu přidané hodnoty slibný nástroj evaluace v oblasti vzdělávání.

## Ročníkové ekvivalenty jako měřítko přidané vzdělávací hodnoty

Navrhovaný přístup k individuálnímu vzdělávacímu růstu ho prezentuje v letech průměrného dosahovaného výkonu, tj. používání skóre *ročníkových ekvivalentů* (GE) (Lyman, 1978) s malou úpravou. Aby byl koeficient GE přímo přídavný, musí odkazovat na počet let a měsíců, které potřebuje průměrný žák k získání testového skóre, nikoli na číslo ročníku a počet měsíců v tomto ročníku. Potom číslice 4,5 vyjadřuje úroveň 4 roků a 5 měsíců výuky, tj. střed 5. ročníku, nikoli 4. ročníku; žákův pokrok o 1,2 roku bude znamenat  $GE\ 4,5 + 1,2 = 5,7$  roků.

Nejdůležitější omezení používání a významu GE jsou následující (Nitko, 1983):

1. Pouze *základní školní dovednosti* (čtení, psaní, matematické uvažování a přírodovědné porozumění) jsou rozvíjeny dostatečně soustavně, aby mohly být měřeny ve všech ročnících základní a střední školy.
2. Je potřeba vytvořit *řadu testů* vzrůstající obtížnosti, aby byly pokryty všechny uvažované ročníky, a *vertikální škály* (Kolen, 2000) GE.
3. Každá *interpolace*, a tím spíše *extrapolace*, učiněná při škálování je zjednodušováním, které přehlíží nedokonalou linearitu skóre, letní ztráty, efekt stropu (horní hranice výkonu) apod.
4. Přestože dvě měření daného rysu/výkonu většinou vysoce korelují (0,80 a více), je třeba brát v úvahu *efekt regrese*.
5. *Omezená reliabilita* obou měření (0,80–0,90) má za následek poměrně značné směrodatné odchylky GE a interval spolehlivosti 0,95 pro GE většinou přesáhne jeden rok školní výuky, čímž se stávají měsíční koeficienty pro jednotlivé žáky zbytečnými.
6. Běžně dochází k *vějířovému efektu*, neboť výkon lepších žáků roste rychleji než výkon těch slabších a rozptyl testových skóre se systematicky zvyšuje s vyššími ročníky. Tento efekt snižuje učební pokrok slabších žáků a zvyšuje pokrok lepších žáků, což redukuje sociální přínos použití GE.

## Pilotní studie GE

Pilotní studie používání ekvivalentních známek byla provedena na dvanácti nižších sekundárních školách v Polsku (Kutajcyk a Przychodzeń, 2008; Niemierko, 2009). Tabulka 1 ukazuje GE získané 535 žáky při povinných celostátních přijímacích zkouškách na nižší střední školy (GEVstup), jejich po-

krok v GE po třech letech studia na nižších sekundárních školách (GEPokrok) a GE absolventů těchto škol (GEVýstup).

Tab. 1: GE na nižších středních školách

GEVstup	≤3,4	3,5–4,4	4,5–5,4	5,5–6,4	6,5–7,4	7,5–8,4	≥8,5	Všechny: 5,9
Počet žáků	35	69	101	138	87	91	14	535
Průměrný GEPokrok	3,2	2,9	2,8	2,9	3,3	3,3	3,1	3
Průměrný GEVýstup	6,2	6,9	7,8	8,9	10,2	11,2	12,1	8,9

Jak vidíme v tab. 1, rozsah koeficientu GEPokrok v sedmi skupinách GEVstupu je spíše malý (2,8–3,3); vějířový efekt vyjádřený závislostí skóru přidané vzdělávací hodnoty (GEPokrok) na předchozím výkonu žáka (GEVstup) je nízký (korelace nepřesahuje 0,15). To znamená, že učitelé mohou očekávat podobný pokrok u všech svých žáků, nezávisle na jejich výkonu u přijímací zkoušky.

Ovšem údaje v tab. 1 přinášejí také pesimistický závěr. Vzhledem k tomu, že míra pokroku (GEPokrok) je mezi skupinami GEVstupu poměrně stabilní, původní rozdíly ve výkonu přetrvaly a rozdíl zhruba pěti let studia ve vstupních údajích skupiny mezi dvěma extrémami (3,4 a 8,5) se znovu objevuje ve výstupních údajích (GEVýstup). Tento obrovský rozdíl mezi nejnižší a nejvyšší skupinou žáků vznikl mnohem dříve, na základě jejich genetických předpokladů, prostředí, ve kterém žijí, a učebních stylů (viz obr. 2).

## Má škola vliv?

Údaje shromážděné v tab. 1 by mohly vzbudit dojem, že pokrok žáka měřený přístupem *přidané vzdělávací hodnoty* je nezávislý na škole a závisí pouze na jeho vlastní míře studia. Abychom mohli tento předpoklad ověřit, projdeme si údaje z pěti z celkového počtu dvanácti zkoumaných škol.

Tab. 2: GE ve vybraných školách

Škola	G03	G08	G04	G09	G12	Celkově
Průměrný GEVstup	5,1	5,6	5,9	6,5	7,8	6,2
Počet žáků	36	54	78	63	48	279
Průměrný GEPokrok	2,4	3,5	2,6	3	3,3	3,0
Průměrný GEVýstup	7,5	9,1	8,5	9,5	11,1	9,2

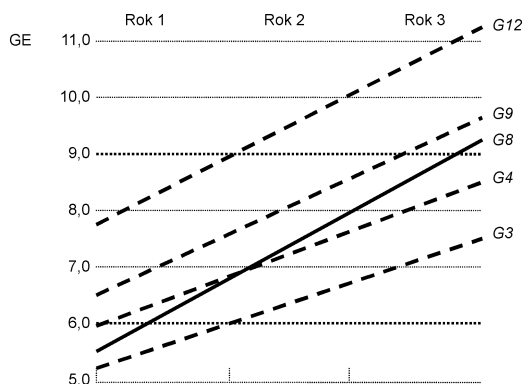


V tab. 2 vidíme:

1. školu s nízkými vstupními GE a malou přidanou hodnotou (*G03*);
2. školu s podprůměrnými vstupními GE a velkou přidanou hodnotou (*G08*);
3. školu s průměrnými vstupními GE a malou přidanou hodnotou (*G04*);
4. školu s vysokými vstupními GE a průměrnou přidanou hodnotou (*G09*);
5. školu s velmi vysokými vstupními GE a velkou přidanou hodnotou (*G12*).

(Poslední z těchto pěti, *G12*, je ve skutečnosti městská, neveřejná, vysoce výběrová škola, která byla k ostatním jedenácti venkovským školám přidána pro možnost srovnání).

Obr. 3 byl vytvořen na základě údajů z tab. 1, aby poukázal na rozdíly mezi školami v přidané vzdělávací hodnotě.



Obrázek 3: Přidaná vzdělávací hodnota v pěti tříletých nižších sekundárních školách

Trajektorie pokroku ve čtyřech školách – *G03*, *G04*, *G08* a *G12* – jsou téměř paralelní, s mírným *vějířovým efektem*, neboť rozdíl mezi nejvýše umístěnou (*G12*) a nejslabší školou (*G03*) vzrostl z 2,7 roku na 3,6 roku. Na tomto pozadí jedna škola, která byla na startovní čáře slabá (*G08*), dosáhla pokroku, který ji v cíli umísťuje nad národní úroveň (9,0). Dotazníky potvrdily, že učitelé ze školy *G08* byli nesmírně *orientovaní na úkol* v řízení učení žáků, což žáky přimělo k usilovné práci, nicméně zároveň je to negativně naladilo vůči učitelům a jejich metodám výuky (Niemierko, 2009, s. 281–286). Dost překvapivě se škola *G03*, nejslabší a poskytující nejmenší přidanou hodnotu, ale zjevně *orientovaná na vztahy* a přívětivá, ukázala

jako nejoblíbenější škola! Můžeme tedy konstatovat, že zázraky ve vzdělávání nejsou možné a že velkého pokroku školy je většinou, do značné míry, dosaženo na úkor žáků.

## Sledování rozvoje žáků pomocí GE

Žák, který je systematicky informován o svém pokroku v základních školních dovednostech, má větší možnost řídit svůj afektivní/motivační a kognitivní rozvoj. To se týká především pomalejších žáků, kteří nejsou schopni dosáhnout očekávaného výkonu v daném učivu a mohou se cítit odsouzeni k nejhorším známám při běžné praxi známkování.

GE jsou zřejmě škálou, kterou žáci a neoborníci přijímají nejlépe. Díky snadnému přiřazení k vlastní stupnici ročníků a převedení ročníků do roků studia je tato škála jedinečná ve smyslu *vnější validity* pedagogické evaluace. Avšak interpretace GE zacházejí někdy příliš daleko. Posouvání pomalejších žáků směrem dolů a rychlejších žáků směrem nahoru po stupnici ročníků by způsobilo v jejich rozvoji mnohem více škody než užitku. Úroveň jejich základních dovedností je zjevně důležitá, ale osvojení si učiva systematicky rok po roku je rovněž nezbytné.

Přínos adaptace skóreů GE na *pomocné* známky při známkování je *jednou ze složek* hodnocení dosahovaného výkonu a závisí na tom, jak úspěšně zabráníme jeho dezinterpretacím. Žáci, rodiče a veřejné mínění mohou přeceňovat informaci, kterou skóre přinášejí. Zneužití GE se lze vyhnout pouze intenzivním vysvětlováním – v tisku a prostřednictvím dalšího vzdělávání učitelů.

Abychom poskytli žákům a učitelům dostatek po sobě jdoucích skóreů GE – řekněme dva ročně – musíme přimět tvůrce (autory) a vydavatele testů, aby opatřovali baterie standardizovaných testů GE pro nejméně tři ročníky: pro nejbližší nižší, pro aktuální, pro nejbližší vyšší. Každý rok jsou v Polsku publikovány stovky testových zadání, z nichž většina je striktně paralelních k nástrojům externích zkoušek, ale žádné z nich není empiricky standardizované, resp. normované.

První krok směrem k pojmání výkonu žáka a školy jako *vlastního výkonu*, a tímto k podpoře rozvíjecího přístupu k hodnocení, by měla zřejmě realizovat centra externích zkoušek. Soukromí autoři a vydavatelé testů budou jistě následovat oficiální model.

## Shrnutí

- V polském vzdělávacím systému jsou tři po sobě jdoucí úrovně celo-státních *externích zkoušek*.
- Naše veřejné mínění má tendenci hodnotit školy a žáky pouze na základě jejich umístění na stupnici výsledků zkoušek.
- *Sumativní* funkce externího zkoušení převládá nad jeho *formativní* funkcí, která je z hlediska vzdělávání cennější.
- Při předkládání výsledků zkoušek žákům, rodičům a vedení školy jsou používány škály *procenta úspěšnosti* a *staninové* škály.
- Tyto škály neodrážejí žákovo *úsilí*, které představuje jednu z hlavních složek hodnocení ve třídě v rámci školy.
- V mnoha zemích lze pozorovat rostoucí zájem o přístup *přidané hodnoty* při hodnocení výkonu žáka a školy.
- Jedna škála dosaženého výkonu ve smyslu přidané hodnoty je tvořena *GE*, které ukazují průměrný počet let a měsíců výuky potřebných k získání úrovně dosaženého výkonu.
- Pilotní studie o *GE* prokázala jejich *přídavný* charakter, důležitý pro vyhodnocení a informování o pokroku.
- Každá škola může urychlit růst žáků v jejich základních školních dovednostech, avšak za cenu intenzivního *úsilí* ze strany učitelů i žáků.
- Poskytnutí ekvivalentních známek žákům jim může dát *zpětnou vazbu* o učebním procesu, je však třeba reflektovat nebezpečí přeceňování tohoto typu škálování.

## Literatura:

- BROOKHART, S. M. Teacher's grading practices: Meaning and values. *Journal of Educational Measurement*, 1993, č. 2.
- HAERTEL, E. H. Validity arguments for high-stakes testing: in search for evidence. *Educational Measurement: Issues and Practice*, 1999, č. 4.
- HAMBLETON, R. H. On the use of cut-off scores in educational settings. *Journal of Educational Measurement*, 1978, č. 3.
- KOLEN, M. J. Scaling and norming. In BRENNAN, L. R. (ed.). *Educational measurement. Fourth edition*. Westport: American Council on Education – Praeger, 2006.
- KORETZ, D. M., HAMILTON, L. S. Testing for accountability in K – 12. In BRENNAN, L. R. (ed.) *Educational measurement. Fourth edition*. Westport: American Council on Education – Praeger, 2006.
- KUTAJCYK, T., PRZYCHODZENI, B. *Czynniki skuteczności kształcenia ogólnego w gimnazjach wiejskich*. Gdańsk: OKE, 2008.

- LYMAN, H. B. *Test scores and what they mean*. 3. vyd. Englewood Cliffs, N. Y.: Prentice-Hall, 1978.
- NIEMIERKO, B. *Pomiar wyników kształcenia*. Varšava: WsiP, 1999.
- NIEMIERKO, B. *Ocenianie szkolne bez tajemnic*. Varšava: WsiP, 2002.
- NITKO, A. J. *Educational tests and measurements. An introduction*. New York: Harcourt, 1983.
- SCRIVEN, M. The methodology of evaluation. In STAKE, R. E. (ed.). *Perspectives of curriculum evaluation*. Chicago: Rand McNally, 1967.
- SZMIGEL, M. K., RAPPE, A. Komunikowanie wartości dodanej osiągnięć szkolnych uczniom, nauczycielom i dyrektorom szkół. In NIEMIERKO, B., SZYLING, G. (ed.) *Hołistyczne i analityczne metody diagnostyki edukacyjnej*. Gdańsk: Fundacja Rozwoju Uniwersytetu Gdańskiego, 2005.
- SZALENIEC, H. *Jak komunikować uczniom wyniki egzaminów*. Varšava: WsiP, 2004.
- SZYLING, G. *Drugi układ wymagań programowych w szkołach gimnazjalnych – znaczenie i funkcjonowanie*. Nепublikovaná disertace, Gdańska univerzita, 2007.

*Z anglického originálu „How students' scores from external examination may be used for supporting their development?“ přeložili Sonia Šamalíková a Tomáš Janík.*

NIEMIERKO, B. Jak lze výsledky žáků z externích zkoušek použít pro podporu jejich rozvoje? *Pedagogická orientace* 2009, roč. 19, č. 4, s. 92–103. ISSN 1211-4669.

**Autor:** Prof. dr. hab. Bolesław Niemierko, Wydział Nauk Humanistycznych i Społecznych, Akademia Marynarki Wojennej, ul. Śmidowicza 69, 81-103 Gdynia, Polska