

Zjišťování a vyhodnocování výsledků vzdělávání žáků

Posuzovací archy

Jana Straková, Petr Suchomel



Zjišťování a vyhodnocování výsledků vzdělávání žáků

Posuzovací archy

Jana Straková, Petr Suchomel

Zjišťování a vyhodnocování výsledků vzdělávání žáků. Posuzovací archy.
Jana Straková, Petr Suchomel

Odborná recenze: Mgr. et Mgr. Jan Mareš, Ph.D., Mgr. Radek Schindler

Jazyková korektura: Mgr. Radomír Novák

Praha: Národní ústav pro vzdělávání, 2012.

ISBN: 978-80-87063-60-6

ZJIŠŤOVÁNÍ A VYHODNOCOVÁNÍ VÝSLEDKŮ VZDĚLÁVÁNÍ ŽÁKŮ. Posuzovací archy.

Obsah

1. Úvodem	5
2. Cíle nástroje	6
3. Testy	7
3.1 Kritéria pro výběr testu	
3.2 Poznámky k současné nabídce českých evaluačních agentur	
3.3 Posuzování kvality výstupů	
4. Autentické evaluační nástroje	13
5. Slovníček pojmů	15
6. Seznam literatury.....	17
7. Přílohy	18
Příloha 1 Posuzovací arch pro výběr didaktického testu – verze s komentáři	
Příloha 2 Posuzovací arch pro výběr autentické evaluační úlohy – verze s komentáři	
Příloha 3 Posuzovací arch pro výběr didaktického testu – verze bez komentářů	
Příloha 4 Posuzovací arch pro výběr autentické evaluační úlohy – verze bez komentářů	

1. ÚVODEM

V České republice působí několik organizací, které nabízejí školám nástroje k hodnocení výsledků vzdělávání. Je určité užitečné, když se učitelé a ředitelé naučí těchto nástrojů využívat a doplní jimi hodnocení, které pravidelně provádějí ve svých školách. Není třeba, aby se učitelé učili konstrukci testů ani konstrukci komplexnějších evaluačních nástrojů a veškeré evaluační nástroje si vyráběli sami. Je však třeba, aby byli v oblasti hodnocení výsledků vzdělávání poučeni natolik, aby kladli na organizace, které se evaluací zabývají, dostatečně vysoké nároky. Jen budou-li učitelé a ředitelé náročnými zákazníky, bude možno docílit, aby evaluační nástroje, které jsou školám nabízeny, byly opravdu kvalitní a pro školy maximálně užitečné. K tomu je třeba, aby věděli, jaké vlastnosti má mít kvalitní evaluační nástroj, podle čeho lze volit vhodný nástroj pro daný účel, jaké užitečné informace je možno z evaluace získat a jak s nimi dále nakládat.

Tento text si klade za cíl pomoci pedagogům, kteří v této oblasti nemají žádné specializované vědomosti, naučit se kriticky hodnotit nabízené evaluační nástroje. Text se snaží o maximální srozumitelnost a užitečnost. Při tvorbě nástroje vycházeli autoři ze svých dlouholetých zkušeností z rozhovorů s pedagogy o testech a testování. V rámci vlastní tvorby nástroje uskutečnili několik hloubkových rozhovorů zjišťujících potřeby pedagogů a problémy, se kterými se při výběru testů a využívání jejich výsledků nejčastěji potýkají. V nich se autoři opětovně přesvědčili o tom, že poučenost pedagogů v této oblasti je relativně nízká. Proto text nepojednává problematiku hodnocení výsledků vzdělávání v celé její šíři, ale omezuje se na několik základních kritérií a soustřeďuje se na jejich podrobný popis a zdůvodnění jejich důležitosti.

Posuzovací archy uváděné v tomto textu jsou odlišné od ostatních nástrojů vytvořených v rámci projektu *Cesta ke kvalitě*. Nejedná se o standardizovaný nástroj přinášející konkrétní informaci o některém aspektu práce školy, ale pouze o pomůcku, která má usnadnit výběr nástrojů k hodnocení výsledků vzdělávání. Z tohoto důvodu je i struktura tohoto textu odlišná od struktury textů provádějících ostatní nástroje.

2. CÍLE NÁSTROJE

Hlavním cílem tohoto textu je poskytnout základní informace o charakteristikách kvalitních testů a testových výstupů, okrajově se zmiňujeme i o jiných evaluačních nástrojích. V kapitolách 3 a 4 uvádíme některé základní pojmy týkající se testů, respektive dalších evaluačních nástrojů, jejich charakteristik a výstupů. Kritéria popisovaná v úvodním textu přehledně uvádíme v příloze v podobě posuzovacích archů. Tyto posuzovací archy je možno přímo použít jako vodítko při výběru evaluačního nástroje a posuzování jeho kvality. Archy uvádíme ve dvou variantách – s vysvětlujícími komentáři a bez komentářů, pouze v podobě otázek.

Text i posuzovací archy jsou určeny primárně učitelům a ředitelům základních a středních škol. I když hledání odpovědí na všechny otázky kladené v rámci pozorovacího archu se může jevit pedagogům úmorné, přesto doporučujeme, aby se pokusili alespoň v počátcích využívání evaluačních nástrojů na všechny otázky poctivě odpovědět. Pomůže jim to naučit se o testech a evaluačních nástrojích strukturovaně přemýšlet.

3. TESTY

Při volbě nástroje si musíme nejprve odpovědět na otázku, co chceme zjišťovat, a následně hledat nástroj, který nejlépe odpovídá našim potřebám. Ani kvalitní nástroj nám dobře neposlouží, bude-li použit k nesprávnému účelu. Stejně tak je často zbytečné administrovat komplexní sadu evaluačních nástrojů jen proto, že je evaluační agentura nabízí vcelku, protože to je to pro ni jednodušší a finančně výhodnější než reagovat na individuální požadavky jednotlivých škol.

Dvěma nejčastějšími důvody pro vyhledání služeb organizací zaměřených na evaluaci jsou jednak snaha zjistit či ověřit, zda žáci umějí vše, co by umět měli (požadovanou úroveň vědomostí učitelé zpravidla vztahují k výstupům vzdělávacích programů), jednak potřeba porovnat výsledky žáků školy s výsledky žáků jiných škol. Každé z těchto potřeb může kvalitně vyhovět jiný typ testu – v prvním případě **test ověřovací**, ve druhém **test srovnávací**. Základním rozdílem mezi nimi je, že zatímco ověřovací test má obsah a obtížnost nastaveny tak, aby odrážel a pokrýval požadavky standardu (např. již zmíněné výstupy vzdělávacích programů), srovnávací test je sestaven z takových testových položek, v nichž jsou (nebo lze předpokládat, že budou) rozdíly mezi testovanými žáky co největší. Použití nevhodného typu testu nemůže přinést spolehlivá (a v některých případech nutně přináší zkreslená) zjištění. Tím zcela prvním, co by měl každý, kdo o využití externě připravených testů uvažuje, udělat, je rozhodnutí, který z uvedených cílů sleduje, a ověření, zda testy, jejichž využití zvažuje, jsou pro jeho cíl konstruovány.

V následujícím odstavci ukazujeme, podle jakých kritérií je možno testy vybírat a na co je třeba dávat pozor při výběru testu a při interpretaci výsledků. Řada náležitostí, které zde uvádíme, není v českém prostředí běžná. V zahraničí však tyto požadavky představují minimální standard. Jen budou-li pedagogové vyžadovat řádnou tvorbu testu a jeho řádnou dokumentaci, získají nástroje, které jim poskytnou užitečnou a spolehlivou informaci o jejich práci.

3.1 Kritéria pro výběr testu

Každý test by měl být opatřen dokumentací, která by měla vedle informace o tom, co a jak je testem ověřováno a jak mají být použity výsledky, obsahovat též některé údaje, podle kterých je možno posoudit kvalitu testu jako měřicího nástroje.

Absence dokumentace by měla představovat pro odběratele vážné varování.

Validita testu

Základním hlediskem při tvorbě a hodnocení testu je jeho validita. Validita nám říká, zda můžeme výsledky daného testu interpretovat (a následně výsledky použít) jednoznačným způsobem. Je založena na teoretických a empirických poznatcích, které se týkají měřené oblasti. Považujeme za důležité zdůraznit, že se nevztahuje k testu samotnému, ale ke způsobu interpretace výsledků. Například test z matematiky určený pro žáka 9. ročníku může být koncipován tak, aby na základě jeho výsledků bylo možno usoudit, jak žák zvládl učivo matematiky základní školy. To neznamená, že tentýž test je automaticky možno považovat za validní ve vztahu k rozhodování o dalším žakově studiu. Jedná se o dvě rozdílné interpretace: v jednom případě usuzujeme na zvládnutí učiva, ve druhém případě na vhodnost určité vzdělávací dráhy.

Validitu testu stanovujeme rozmanitými způsoby v závislosti na charakteru testu a způsobu užití jeho výsledků. Například validitu testu určeného k výběru uchazečů pro nějaké studium můžeme stanovit prostřednictvím porovnání výsledků v testu s následnou úspěšností v tomto studiu. V jiných případech rozhodujeme o validitě například na základě expertního posouzení testu nebo na základě shody výsledků testu s výsledky jiného měření.

Zcela mimo možnosti téměř jakéhokoli testu je dát obhajitelnou informaci o tom, jak by měl být žák klasifikován (zejména proto, že klasifikace vždy odráží i jiné než vědomostní složky a řídí se specifickými kritérii jednotlivých učitelů). Přesto jsou snahy učitelů získat test, na jehož základě by mohli klasifikaci provést, poměrně časté; validita drtivé většiny testů je ovšem z hlediska klasifikace pramalá (neznamená to ovšem, že není možné použít výsledky vhodných testů ke zjištění, zda klasifikace konkrétních žáků více či méně odpovídá tomu, jak jsou žáci s obdobnými vědomostmi klasifikováni v jiných školách).

Každý test by měl obsahovat informaci o tom, jaká z možných interpretací je validní, tedy jak mohou být jeho výsledky používány, vhodnost tohoto využití by měla být teoreticky a/nebo empiricky doložena. Například na způsobilost žáka s určitým testovým výsledkem pro akademické studium může být usuzováno teoreticky z nároků tohoto studia nebo například empiricky z toho, že žáci s určitým testovým výsledkem v tomto studiu dobře prospívají.

Reliabilita

Vysoká validita zaručuje, že můžeme výsledek testu využít předepsaným způsobem. Potřebujeme se ještě dozvědět, jak kvalitní výsledek získáváme, jak moc se na něj můžeme spolehnout. Tuto informaci o kvalitě testu nám poskytuje údaj o jeho spolehlivosti, reliabilitě. Čím vyšší je reliabilita, tím větší je pravděpodobnost, že bychom při opakovaném měření dosáhli stejného výsledku, tedy že výsledek není náhodný¹. Pokud jsou součástí testu úlohy, které jsou hodnoceny hodnotiteli, měla by být součástí dokumentace i informace o míře shody mezi hodnotiteli. Z dokumentace k testu bychom měli také získat údaj o chybě měření. Je-li například výsledkem testu přiřazení žáků k určitým klasifikačním stupňům nebo úrovním způsobilosti, měli bychom získat informaci o tom, jaký podíl žáků je při opakovaném měření přiřazen ke stejnému klasifikačnímu stupni nebo ke stejné úrovni způsobilosti.

Reliabilita testu je obvykle stanovována teprve po proběhlém testování (zejména u velkých, hromadných testování), její hodnotu může ale již před testováním dobře ukázat například reliabilita pilotáží, v nichž byly testové úlohy připravovány pro ostré použití. Čím větší váhu chceme výsledkům testů dát, tím více narůstá na významu znalost reliability, a to jak při interpretaci výsledků, tak už při rozhodování o tom, zda daný test pořídíme a použijeme.

Popis testu

Povinnou součástí všech kvalitních nástrojů včetně testů vědomostí a dovedností žáků, se kterými se setkáváme v zahraničí, je přesný popis toho, co zjišťují, na jaké vědomosti a dovednosti jsou zaměřeny. Tento popis bývá zpravidla provázen rovněž zdůvodněním, proč se test zaměřuje právě na tyto aspekty vzdělávání, proč právě tyto aspekty považují tvůrci testu za důležité – z hlediska porozumění testované disciplíně, dalšího vzdělání nebo potřeb dospělého života. Takové informace zcela zásadním způsobem pomáhají vybrat si správný test pro daný účel i vytvořit si názor na kvalitu testu.

Testy jsou zpravidla koncipovány tak, aby jejich řešení trvalo žákům jednu, maximálně dvě vyučovací hodiny. Za tuto dobu žáci vyřeší jen omezené množství testových úloh. Chceme-li srovnávat výsledky jednotlivých žáků, zadáváme všem žákům stejný test, a tedy můžeme ověřit jen velmi omezené množství vzdělávacích cílů. Z tohoto důvodu je mimořádně důležité, aby tyto cíle byly dobře voleny. Aby byly ověřovány základní vědomosti a dovednosti, nikoli okrajové, zbytečné vzdělávací cíle². Na to, co je pro danou oblast stěžejní, mohou existovat odlišné názory. Proto je třeba vyžadovat, aby bylo vždy přesně vymezeno, na co se daný test zaměřuje, abychom mohli volit takový test, který hodnotí to, co my považujeme za zásadní a na co se zaměřujeme ve své výuce.

Součástí dokumentace k testu by tedy měl být popis toho, jak jsou ověřované vzdělávací cíle operacionalizovány: jaké dílčí cíle jsou zjišťovány a prostřednictvím jakého typu úloh a jak jsou jednotlivé dílčí cíle v testu pokryty. Součástí informace o testu by měly být i pokyny k administraci, případně k vyhodnocení žákovských odpovědí. Tyto pokyny by měly být striktně dodržovány.

Mnoho testů nabízených českými organizacemi zabývajícími se evaluací není opatřeno ani základní informací o tom, co daný test zjišťuje, natož vysvětlením, jak to měří. To staví uživatele do situace, kdy musí tvůrcům testu plně důvěřovat, že zjišťují to, co on považuje za důležité a na co se zaměřuje ve své výuce, a že to měří dobře.

1. Reliabilita je zpravidla charakterizována koeficientem Cronbachovo alfa, který nabývá hodnot 0-1. Test je tím spolehlivější, čím více se hodnota blíží jedničce. Za přijatelné jsou zpravidla považovány hodnoty vyšší než 0,75, v některých zdrojích též 0,85. V případě, že test obsahuje několik součástí, které zjišťují rozdílné vědomosti a dovednosti, musejí být hodnoty koeficientu stanoveny pro každou část zvlášť.

2. Zjišťujeme-li vědomosti a dovednosti žáků prostřednictvím výběrových šetření (jako jsou například mezinárodní srovnávací výzkumy PISA a TIMSS), jejichž cílem není hodnotit jednotlivé žáky, zadáváme různým žákům různé testy, což nám umožňuje ověřit daleko širší rozsah vzdělávacích cílů. V případě hodnocení jednotlivých žáků je však nutno se zaměřit na cíle, které jsou pro danou vzdělávací oblast stěžejní.

Tato důvěra v českém prostředí ovšem nemusí být zcela na místě: setkáváme se i s testy, které víceméně nahodile poskládají úlohy ze školních učebnic, z nichž některé testují důležité vědomosti a dovednosti a některé vědomosti a dovednosti zcela dílčí a okrajové, přičemž objem učiva, který chtějí zahrnout, je vzhledem k délce testu nesmyslně široký. Výsledek takového testu má pro učitele jen velmi omezenou informační hodnotu a často je obtížné jej správně interpretovat. Proto je velice důležité, aby uživatelé testů vyžadovali od organizací zabývajících se evaluací, aby testy řádně dokumentovaly. Organizace, na které budou kladeny vyšší nároky, budou také produkovat kvalitnější testy, neboť budou muset o své práci více přemýšlet.

Testová skóre

Tvůrci testu mají vždy na výběr z několika základních způsobů, jimiž lze vyjádřit výsledek testování – od prostého počtu získaných bodů přes různé formy vyjádření procentní úspěšnosti až po třeba percentilové pořadí či například z-skóre³. Každý způsob prezentace výsledku má své výhody i nevýhody. Počet bodů získaných v testu je transparentním údajem, dobře srozumitelným pro žáky i jejich rodiče, z vlastního údaje o počtu bodů ale nikterak nelze (bez dalších údajů) zjistit, jak žák dopadl v porovnání s ostatními. Percentil poskytuje dobrou informaci o tom, jak je na tom žák v porovnání s ostatními, ale nic neříká o tom, jakou část ověřovaného obsahu testu zvládl. Z-skóre je pro většinu žáků a rodičů hůře srozumitelné, ale sděluje lépe než jiné ukazatele, nakolik je žákův výsledek reálně horší či lepší než průměr dosažený všemi testovanými žáky.

Snahou těch, kdo využívají externě připravených testů a testování, by mělo vždy být získat výsledek žáků (tříd, školy) ve více formách, protože každá z nich může poskytnout jiný obraz výsledků. Tvůrce testu by měl vždy předem popsat a zdůvodnit, jaké způsoby vyhodnocení použije. Test by měl dále obsahovat popis stupnic, na kterých jsou stanoveny výsledky, jsou-li testování na základě výsledků rozdělování do kategorií, měl by být zdůvodněn způsob rozdělování. Pokud je výsledek testu vyhodnocován vzhledem k normám, měl by být podrobně zdokumentován způsob vytváření norem. Vznikají-li dvě nebo více variant testu, mělo by být podrobně popsáno, jak byla zajištěna jejich ekvivalence.

Každý číselný výsledek vyjadřující výkon žáka v testu by měl být doplněn informací o *směrodatné odchylce* testování. Bez její znalosti nelze nikterak posoudit, zda například rozdíl jednoho bodu ve skóre žáků odráží skutečně významný rozdíl v jejich vědomostech, nebo zda je jen náhodnou fluktuací plynoucí z nikdy ne zcela přesných testových otázek. Je bohužel poměrně časté, že žáci jsou poměřováni na základě desetinových rozdílů v jejich výsledcích, ačkoli u řady testů (vzhledem k jejich reliabilitě) je pásmo, v němž jsou vlivem statistické nepřesnosti výsledky prakticky rovnocenné (odpovídají stejnému výkonu žáků), široké několik celých bodů.

Srovnávací a ověřovací testy⁴

U každého testu musí být zřejmé, zda se jedná o test, jehož cílem je rozdělit žáky podle úrovně jejich vědomostí a dovedností (srovnávací test), nebo pouze o ověření, zda si žáci osvojili jistou předem stanovenou úroveň vědomostí a dovedností (ověřovací test). Srovnávací test je vhodný typicky pro účely přijímacího řízení. Ověřovací testy se naopak velmi dobře hodí jako zpětná vazba pro učitele: poskytnou informaci o tom, jaký podíl žáků si osvojil vědomosti a dovednosti, které jsou považovány za důležité, respektive na které se jeho výuka zaměřuje.

Zatímco u srovnávacích testů vzniká norma (hranice rozdělující jednotlivé testované do skupin, např. na horší třetinu, průměr a lepší třetinu) až během testování, u ověřovacích testů je zcela nezbytné, aby vztah mezi výsledkem v testu a zařazením do jedné z úrovní byl stanoven, popsán a odůvodněn předem – není tedy např. možné u testu, jehož cílem je posouzení toho, kteří žáci dosáhli požadované úrovně, poukazovat na to, že hranice pro kladný výsledek v testu bude stanovena až po testování podle toho, jak testování žáci dopadnou (jakkoli k tomu obava, že požadovaného výsledku dosáhne příliš malý počet žáků, může svádět). Takový přístup vždy svědčí o tom, že vazba obsahu a obtížnosti testu k relevantním standardům a normám není dostatečná.

3. Standardizovaná skóre s průměrem 0 a směrodatnou odchylkou 1.

4. Norm-referenced, resp. criterion-referenced tests.

Vývoj/přidaná hodnota

Výsledky testování můžeme použít jako jednorázovou informaci, můžeme je však také použít pro srovnání dlouhodobější. Například můžeme chtít pravidelně testovat žáky určitého ročníku, abychom zjistili, zda jsme schopni vybavovat naše žáky stále stejnými vědomostmi a dovednostmi. V tom případě potřebujeme vědět, zda nám to testy, které zamýšlíme použít, skutečně umožní.

Můžeme také chtít dlouhodobě sledovat stejnou skupinu žáků a zjišťovat, jak se v průběhu studia zlepšují. Toto měření bývá zpravidla označováno jako měření přidané hodnoty, neboť zohledňuje vstupní úroveň žákových vědomostí. Poskytuje informaci o přírůstku ve vědomostech a dovednostech žáka díky péči té které školy/toho kterého učitele. Někdy je přidaná hodnota stanovována nejen prostřednictvím přírůstku vědomostí, ale jsou v ní zohledňovány další faktory, například rodinné zázemí žáka, které též ovlivňuje jeho učební pokrok. Pokud nás zajímá měření přidané hodnoty, potřebujeme vědět, zda dodavatel testu nabízí nějaká měření, která nám umožní porovnávat výsledky žáků v jednotlivých letech, a jakou metodiku výpočtu používá.

Je rozdíl, zda je předmětem měření absolutní posun žáka odpovídající sledování, jak postupně naplňuje objektivně stanovený standard vědomostí – všechny použité testy pak musejí takovému standardu obsahově odpovídat, nebo zda je cílem měření zjištění, jak se mění pozice žáka mezi všemi ostatními žáky (v tomto druhém případě je například nutné zajistit a prokázat, že se všech testování účastní složením i velikostí ekvivalentní skupina žáků). Objektivní měření a přiměřená interpretace výsledků právě u sledování vývoje úspěšnosti žáka/žáků v čase je netriviální záležitost, přesto bývá nabízeno ve značně zjednodušené podobě jako součást řady komerčně nabízených testových produktů, jejichž vhodnost k tomuto účelu je přinejmenším sporná.

3.2 Poznámky k současné nabídce českých evaluačních agentur

Zde uvádíme několik aspektů spojených s nabídkou českých organizací zabývajících se hodnocením výsledků vzdělávání, které nám připadají z hlediska využití školami problematické a považujeme za důležité na ně upozornit. Naš výčet určitě není vyčerpávající.

Testy studijních předpokladů, obecných dovedností

Školám jsou často nabízeny všelijaké testy studijních předpokladů či tzv. obecných dovedností. V případě těchto testů je třeba se ještě více zajímat o to, co vlastně měří, než u testů z dílčích vzdělávacích oblastí, abychom byli schopni jejich výsledky správně interpretovat. Tyto testy zpravidla posuzují základní jazykové, respektive matematické dovednosti žáků a jejich logické uvažování. Základním problémem při interpretaci jejich výsledků je ovšem skutečnost, že často nejsou dostatečně teoreticky ani empiricky podloženy. V zahraničí (např. ve Spojených státech) jsou takové či obdobné testy používány pro přijímání studentů k vysokoškolskému studiu. Existují doklady o tom, že studenti s vyšší úrovní těchto dovedností jsou úspěšnější při vysokoškolském studiu. Využití testu studijních dovedností k tomuto účelu má tedy dobrý smysl.

U nás jsou však výsledky testů používány k hodnocení žáků, škol a učitelů. A zde již je třeba maximální obezřetnosti: nevíme, do jaké míry působí na rozvoj těchto testovaných dovedností škola a jak se přirozeně vyvíjejí s věkem. Máme málo dokladů o tom, do jaké míry je možno je systematicky rozvíjet cíleným cvičením, a zpravidla se na tyto dovednosti ve výuce také nezaměřujeme. Proto je pro nás obtížné interpretovat informaci o úrovni těchto dovedností, případně jejich nárůstu u našich žáků. Zůstává zde řada nezodpovězených otázek: Je tento nárůst způsobem tím, že je lépe vzděláváme? Nebo tím, že zestárli? Je dobře, když rozdíl mezi obecnými dovednostmi a vědomostmi a dovednostmi v jednotlivých vzdělávacích oblastech je co největší, neboť to svědčí o tom, že jsme byli schopni žáky v dané oblasti maximálně rozvinout bez ohledu na jejich přirozené schopnosti? Nebo je naopak dobře, když je rozdíl co nejmenší, neboť to svědčí o tom, že jsme maximálně schopni rozvíjet u žáků tyto obecné dovednosti i v rámci předmětové výuky? Je vysoká úroveň obecných dovedností výsledkem naší práce, nebo pouze genetickým vkladem našich žáků?

Tyto otázky si samozřejmě můžeme klást i v případě jednotlivých vzdělávacích oblastí; ani tam není zcela zřejmé, že zlepšení výsledků je důsledkem naší práce, ale souvislost předmětových vědomostí a dovedností s výukou je přece jen těsnější. Z toho důvodu doporučujeme pečlivě zvážit, zda pro nás testování obecných dovedností má nějakou informační hodnotu, a pokud se rozhodneme, že ano, a test si objednáme, být maximálně obezřetní při interpretaci jeho výsledků.

Ještě diskutabilnější v tomto ohledu jsou tzv. testy všeobecného přehledu. Zde je zpravidla zjišťováno, zda žák zná vybraná fakta, která jsou tvůrcem testu považována za důležitá a vypovídající z hlediska všeobecného vzdělání. Hodnocení toho, co je a co není důležité, však může být velmi individuální. Zde by měl uživatel v každém případě vyžadovat podrobné zdůvodnění zacílení testu a jeho operacionalizace a měl by být maximálně obezřetný při posuzování validity.

Složení škol ve vzorku

Školy zpravidla získávají od organizátorů testování informaci o výsledcích žáků své školy a informaci o „celostátním průměru“. Tento „celostátní průměr“ je významně ovlivněn tím, jaké školy jsou do toho kterého testování zapojeny. Součástí zprávy o výsledcích testování by měla být podrobná informace o obecných charakteristikách ostatních škol (regionu, velikosti místa působnosti školy, velikosti školy, zaměření, v ideálním případě i složení žáků školy, charakterizované např. nejvyšším vzděláním rodičů). Na základě těchto informací si mohou učitelé vytvořit představu o tom, s kým se srovnávají. Analýzy podle typu škol by měly být rovněž samozřejmou součástí školních zpráv. Vedle celkového porovnání v rámci všech výsledků by tedy žáci (třídy, školy) měli být srovnáni postupně v rámci všech podskupin, do nichž svými charakteristikami patří.

3.3 Posuzování kvality výstupů

Obsažnost a informativnost zprávy

Již jsme se zmiňovali o tom, že při volbě testu bychom měli mít jasno v tom, jakou informaci z něj chceme získat. Podle obsažnosti zprávy a bohatosti získaných informací posuzujeme, zda je investice do testování dostatečně efektivní. Požadované informace mohou být velice rozmanité: můžeme chtít informace nejen o výsledcích celé třídy, ale i o výsledcích jednotlivých žáků, nejen celkové výsledky, ale též výsledky v jednotlivých učebních oblastech a typech dovedností nebo i v jednotlivých úlohách. Můžeme chtít porovnání se známkami, můžeme požadovat srovnání výsledků podle typu škol, umístění školy nebo rodinného zázemí žáků. Můžeme též požadovat informaci o souvislosti výsledků s jinými proměnnými, například s motivací pro testovaný předmět nebo s převládajícími výukovými metodami. Důležité je, aby analýza výsledků byla schopna poskytnout podněty, či ještě lépe návody pro žáka i pro učitele, co a jak je možné vzhledem k výsledkům testování změnit. Pravidlem by mělo být mít možnost ještě před zakoupením testového produktu seznámit se s ukázkovou výsledkovou zprávou.

Celkový výsledek

V části věnované kvalitě testu jsme zdůrazňovali, že součástí popisu testu by měla být přesná informace o tom, jakým způsobem jsou prezentovány výsledky. Zajímá nás například, zda budou porovnávány body, procenta, percentilová skóre či z-skóre. Také nás zajímá, zda mají některé úlohy ve výsledném skóre větší váhu než jiné (například obtížnější úlohy mohou mít vyšší ohodnocení) nebo zda jsou strhávány body za špatnou odpověď. Vždy je výhodné, jsou-li výsledky prezentovány více způsoby, neboť nám mohou poskytnout o výsledcích žáků bohatší informaci.

V některých případech vzniká výsledek nějakým komplikovanějším způsobem: například může vzniknout kombinací výsledků z více testů nebo může být třeba ve výsledku zohledněna nějaká další skutečnost – například rodinné zázemí žáka. V těchto případech je ještě důležitější, abychom získali přesnou informaci o výpočtu, abychom dobře porozuměli tomu, na základě jakých údajů posuzujeme vědomosti a dovednosti našich žáků.

V řadě případů jsou výsledky prezentovány ve srovnání s výsledky žáků z jiných škol, kteří se do testování zapojili. V tom případě je třeba požadovat dostatečně podrobnou informaci o tom, s kým jsou naši žáci porovnáváni, abychom si mohli vytvořit názor na to, zda je srovnání s ostatními zúčastněnými školami a žáky relevantní. Důležitý je nejen počet ostatních žáků, ale zejména jejich složení podle typu školy, rodinného zázemí, sídla školy.

I v případě, že je pro nás nejdůležitější zjistit, jak si naši žáci vedou ve srovnání s jinými žáky, měli bychom požadovat informaci o tom, co naši žáci umějí, tedy jaké vědomosti a dovednosti mají žáci, kteří dosáhli určitého výsledku. Tato informace je nezávislá na skupině, s kterou je srovnávání prováděno. V optimálním případě (a vždy v případě ověřovacích testů) by vedle popisu vědomostí a dovedností žáků měla být k dispozici též informace, zda odpovídají nějaké normě, nějakému standardu.

Například můžeme mít informaci o rozmezí výsledků, ve kterém se pohybuje běžný výkon žáků v určitém věku, a můžeme porovnávat výsledky našich žáků s touto normou (jaký podíl našich žáků nedosahuje běžného výkonu, jaký podíl vykazuje vědomosti a dovednosti typické pro daný věk a jaký podíl je přesahuje). Můžeme mít také k dispozici nějaký minimální standard, který ukazuje hranici, jíž by měli dosáhnout všichni žáci daného věku. V tom případě nás bude zajímat, jací žáci minima nedosáhli, abychom jim mohli poskytnout dodatečnou podporu.

Pokud není k dispozici popis vědomostí a dovedností vztahujících se k určitému výkonu ani srovnání s normou a výsledek je prezentován opravdu pouze ve srovnání s jinými žáky, je vhodné požadovat vedle celkového srovnání ještě srovnání v rámci dílčích skupin (podle typu školy, typu studia, sídla, socio-ekonomického statusu). Takové srovnání může alespoň částečně odstranit zkreslení, která jsou způsobena nereprezentativním složením srovnávací skupiny, a lépe nám pomůže porozumět tomu, jak jsou na tom naši žáci ve srovnání s žáky, kteří se vzdělávají v podobných podmínkách.

Dílčí výsledky

Máme-li z výsledku testování získat podněty pro zlepšení naší výuky, potřebujeme se dozvědět, co přesně naši žáci neumějí, s jakými úkoly mají potíže. Součástí výsledné zprávy by tedy v každém případě měla být vedle informace o celkovém výsledku také informace o výsledcích v dílčích oblastech testovaných vědomostí a dovedností.

Ideální je, jsou-li spolu s dílčími výsledky poskytnuty rovněž výsledky v jednotlivých úlohách. Ty umožní učitelům nejlépe posoudit, zda žáci umějí to, co od nich vzhledem ke své výuce očekávají, rozpoznat chybné postupy a nepochopení, kterými jsou jejich výkony poznamenány, a naplánovat postupy k jejich cílenému odstraňování.

Další informace

Učitele často zajímá, jaká je souvislost mezi výsledkem v testu a školní známkou u jejich žáků a u žáků z jiných škol. Tato informace pomáhá učitelům učinit si představu o přísnosti jejich klasifikace ve srovnání s klasifikací jiných učitelů. Zde bychom ovšem měli vyžadovat srovnání podle jednotlivých typů škol. Je velmi pravděpodobné, že například v gymnaziálních a výběrových třídách bude nastavena jiná „laťka“ než na běžných základních školách.

Další zajímavou informací může být například informace o motivaci našich žáků pro daný obor nebo o jejich sebepojetí ve vztahu k danému předmětu (jak vysoko hodnotí své schopnosti, jak moc si věří). Existuje řada empirických dokladů o tom, že motivace a sebepojetí jsou pro úspěch minimálně stejně důležité jako vědomosti. Proto je tyto aspekty právem možno považovat za důležité výstupy učitelovy práce.

Velmi důležitá je také informace o výsledcích znevýhodněných skupin žáků, například cizinců nebo žáků ze znevýhodněného rodinného prostředí. Tato informace umožní učitelům ověřit, zda se jim daří rozvíjet stejně efektivně všechny žáky.

4. AUTENTICKÉ EVALUAČNÍ NÁSTROJE

Považujeme za důležité upozornit na to, že testy nejsou zdaleka jedinými nástroji, které by mohly školy od externích agentur popřát. V současné době jsou instituce zabývající se evaluací vyhledávány zejména proto, že poskytují testy a umožňují komparaci výsledků žáků jednotlivých tříd a škol s výsledky žáků z ostatních škol, které jsou do evaluační aktivity zapojeny. Učitelé by nicméně mohli požadovat i další typy evaluačních nástrojů, které tuto komparaci v principu také umožňují. Typickým pomocníkem pro každodenní práci učitelů jsou jednotná zadání úloh s návody na jejich vyhodnocení.

Tyto nástroje mají tu výhodu, že umožňují postihnout širší spektrum žákovských vědomostí a dovedností, než postihují tradiční testy, a sledovat jejich vývoj. Rovněž mohou být voleny tak, aby daleko lépe odpovídaly výuce, tedy tomu, co se žáci aktuálně ve škole učí. Pro profesionální agentury zabývající se evaluací není zároveň problém nástroje standardizovat a poskytnout tak učitelům informaci o tom, jaké jsou hodnocené vědomosti a dovednosti jejich žáků ve srovnání s celostátním průměrem, respektive se specifickými skupinami žáků a škol.

Učitelé si samozřejmě v principu mohou tvořit podobná zadání sami. Na webových stránkách amerických a britských organizací je možno nalézt mnoho pěkných příkladů zadání úloh a zejména schémat pro jejich vyhodnocování⁵. Tvorba je však časově náročná, vyžaduje určitou zkušenost a učitelova samostatná práce neumožňuje srovnání s jinými školami, tedy mu neposkytuje dostatečnou zpětnou vazbu o jeho práci. Proto se domníváme, že by se evaluační agentury měly zaměřit i na tento typ služeb, respektive že by učitelé měli od nich tento typ služeb vyžadovat. Pro lepší představu o tom, jaký typ nástrojů máme na mysli, uvádíme v rámečku ukázkou návodu pro hodnocení dějepisného referátu. Uvedená úloha umožňuje při vyhodnocování jistou subjektivitu, ale dobře ilustruje šíři dovedností, které je možno s jejich pomocí hodnotit. Zadání i návod by bylo samozřejmě ještě možno konkretizovat a objektivitu hodnocení zvýšit.

Dalším příkladem úloh, které by mohly posloužit učitelům k objektivnímu posouzení vědomostí a dovedností jejich žáků, jsou čtenářské úlohy⁶. Pěknou ukázkou takové úlohy je možno nalézt na stránkách Ústavu pro informace ve vzdělávání: <http://www.uiv.cz/soubor/1266>.

Zpracování dějepisného tématu formou referátu (základní škola)

Znalosti
4 Problém vysvětlí jasně, logicky a konzistentně. Jasně vymezí téma práce. Shromáždí vhodná a relevantní fakta a jasně je strukturuje.
3 Problém vysvětlí důkladně a víceméně přesně. Vhodně vymezí téma práce. Shromážděná fakta jsou správná, ale je jich málo.
2 Částečně vysvětlí problém, ale pomine některé klíčové prvky. Zadání práce není vymezeno správně. Fakta jsou zmatená a neuspořádaná.
1 Téma není vysvětleno vůbec. Nedostatečně vymezené zadání. Základní informace je nedostatečná.
Kvalita prezentace
4 Téma práce odpovídá zadání a je prezentováno nápaditě. Práce je velmi kreativní, je věcná, smysluplná, výstižná. Prezentace je plynulá, má logickou strukturu. Na obrazových materiálech je patrná důkladnost, důraz na detaily jako překlepy, úprava.
3 Téma práce souvisí se zadáním, práce je zajímavá a smysluplná. Prezentace je v zásadě plynulá. Obrazové materiály pěkně vypadají. Mírně nedostatečný důraz na detaily jako překlepy, úprava apod.
2 Téma práce příliš nesouvisí se zadáním. Sdělení je nejasné a chybějí podstatné prvky. Prezentace není smysluplná. Chyby v detailech.
1 Téma nesouvisí se zadáním, postrádá srozumitelnost. Velké chyby v detailech.
Kvalita analýzy
4 Práce je jasná, smysluplná, věcná, cílená. Má závěr. Jsou použity výstižné argumenty. Závěry jsou přesvědčivé a založené na důkazech.
3 Práce je smysluplná. Jasně formuluje problémy a řeší je. Argumenty jsou do značné míry uspořádané a výstižné. Závěry jsou celkem přesvědčivé.

5. V angličtině jsou schémata pro hodnocení vědomostí a dovedností zpravidla označována jako „rubrics“.

6. V angličtině bývají označovány jako reading inventories.

2 Práce není dostatečně smysluplná, neřeší jasně vymezený problém. Argumenty jsou zmatené a nejasné. Řešení problému je nedostatečné.
1 Práce není smysluplná, neformuluje problém. Argumenty nejsou podloženy fakty. Chybí řešení.
Zdroje
4 V práci bylo využito neobvyklých zdrojů. Bylo konzultováno několik institucí. Literatura obsahuje množství primárních i sekundárních zdrojů.
3 Použití dostatečných zdrojů, konzultována jedna nebo dvě instituce. Literatura obsahuje dostatečné množství primárních i sekundárních zdrojů.
2 Chybějí dostatečné zdroje. Konzultována byla aspoň jedna instituce. Literatura obsahuje nedostatečné množství primárních a sekundárních zdrojů.
1 Žádné zdroje nebo nedostatečné, nebyly konzultovány žádné instituce.

Zdroj: Košťálová, H., & Straková, J. (Eds). 2008. *Hodnocení, důvěra, dialog, růst*. Praha: SKAV.

Na jaké otázky je třeba si odpovědět při výběru autentické evaluační úlohy?

Autentické evaluační úlohy jsou zpravidla časově náročné jak pro učitele, tak pro žáka. Proto je třeba je vybírat pečlivě tak, aby čas s nimi strávený byl pro žáky co nejužitečnější a aby byla co nejužitečnější informace, kterou z nich žák i učitel získají. Tyto úlohy mohou být zadávány primárně jako zpětná vazba pro žáky, v tom případě není zásadní maximální objektivita hodnocení, je však důležité, aby úloha byla pro žáky maximálně stimulující.

Pokud je hlavním smyslem evaluace srovnat výsledky žáků s nějakými národními standardy nebo s výsledky žáků jiných škol, je důležitá podrobnost a jednoznačnost návodu na vyhodnocování, která určuje objektivitu vyhodnocování žákovských odpovědí.

5. SLOVNÍČEK POJMŮ

Cronbachovo alfa

Způsob vyjádření reliability (viz reliabilita).

decil

Způsob vyjádření individuálního výsledku srovnávacího testování. Rozdělením seřazených výsledků všech testovaných na deset stejně početných skupin získáme deset pásem – decilů. Přiřazení hodnoty decilu ke konkrétnímu výsledku říká, ve kterém pásmu se daný výsledek nachází. Jedná se o obdobu percentilu, kde se výsledek porovnává se stem stejně početných skupin testovaných (tj. každá skupina odpovídá jednomu procentu ze všech testovaných – odtud název). Analogicky lze výsledek zařadit např. do kvartilu (pořadí testovaných rozdělené do čtyř skupin), pentilu (pět skupin) atd.

diskriminační koeficient úlohy

Hodnota významná zejména pro sestavování srovnávacích testů. Zpravidla se stanovuje jako rozdíl mezi úspěšností úlohy v pětině celkově nejlepších řešitelů testu a úspěšností téže úlohy v pětině celkově nejhorších řešitelů testu. Vysoký diskriminační koeficient úlohy říká, že úlohu řešili lépe ti, kdo v celém testu dopadli lépe, než ti, kdo v celém testu dopadli hůře – to je logicky žádoucí zjištění. Nízké, nebo dokonce záporné hodnoty diskriminačního koeficientu úlohu diskvalifikují z použití v kvalitním testu.

korelace

Vzájemný vztah mezi dvěma procesy nebo veličinami. Jsou-li veličiny korelovány, se změnou jedné dochází ke změně druhé a naopak. Míra korelace se vyjadřuje korelačním koeficientem – nabývá hodnot od -1 do $+1$. Nulová hodnota korelačního koeficientu znamená, že procesy (veličiny) spolu vůbec měřitelně nesouvisí. Čím vyšší je kladná hodnota korelačního koeficientu, tím silněji platí, že zvýšení hodnoty jedné veličiny povede ke zvýšení také druhé veličiny. Naopak čím více se záporný korelační koeficient blíží hodnotě -1 , tím silněji platí, že zvýšení jedné veličiny je provázeno snížením hodnoty druhé veličiny a naopak.

KR-20 (Kuder-Richardsonova reliabilita)

Způsob vyjádření reliability (viz reliabilita).

ověřovací test (CR-test, criterion-referenced test)

Test, který poměřuje konkrétní dovednost testovaného s předem určenou kvalifikační hranicí (může být jediná u testů s výsledkem prospěl/neprospěl, může jít o více kvalifikačních stupňů, jako např. v případě rozřazovacího testu úrovně začátečník, mírně pokročilý, pokročilý apod.). Úkolem testu tedy je zjistit, kdo určitým stanoveným kritériím vyhověl a kdo nikoli. Výsledek se vztahuje výhradně ke kvalifikační stupnici, nikoli k výsledkům ostatních testovaných. Pro jeho strukturu a pro charakteristiky úloh, které jej tvoří, platí specifická pravidla, odlišná od pravidel platících pro srovnávací test, proto jeho výsledek lze jen velmi omezeně použít pro formulaci jakýchkoli srovnávacích soudů.

percentil

Způsob vyjádření individuálního výsledku srovnávacího testování. Jeho hodnota udává, kolik procent ze všech ostatních testovaných (žáků, tříd, škol) dosáhlo horšího výsledku v testu. Hodnota percentilu 100 znamená, že výsledek je nejlepší ze všech testovaných (všichni = 100 % ostatních testovaných – dosáhli horšího výsledku). Hodnota percentilu neříká ovšem nic o absolutním výsledku v testu – např. v testu, v němž lze získat maximálně 20 bodů, může stému percentilu odpovídat například 5 získaných bodů (tedy úspěšnost 25 %), pokud všichni ostatní testovaní získali ještě méně bodů. Naopak získání 18 bodů v témže testu může znamenat nulý percentil, pokud všichni ostatní testovaní získali 19 bodů.

pilotáž

Proces, během něž jsou zjišťovány důležité parametry úloh nebo celých testů. Pro sestavení kvalitního testu libovolného typu je vždy nezbytné znát řadu kvalitativních i kvantitativních parametrů (úspěšnost úlohy, diskriminační koeficient úlohy apod.). Zadáním úlohy přiměřeně velkému vzorku testovaných (měl by strukturou odpovídat cílové skupině testovaných) lze důležité parametry určit s dostatečnou přesností ještě před ostrým použitím úloh (testu).

reliabilita

Statistická veličina udávající spolehlivost testu (z anglického reliability – spolehlivost, hodnověrnost). Vyjadřuje se různými parametry podle statistického modelu výpočtu – nejčastěji jako Cronbachovo alfa, případně hodnota KR-20. Vyšší hodnota reliability zaručuje, že výsledky testování jsou méně zatížené vlivem náhody a faktorů nesouvisejících přímo s měřenou vlastností respondentů. Například u Cronbachova alfa jsou hodnoty nad 0,9 považovány za známku vysoce kvalitního testu, hodnoty kolem 0,8 uspokojivě kvalitního testu, hodnoty pod 0,75 už neopravňují příliš se na výsledky testování spolehnout.

směrodatná odchylka

Statistická veličina vypovídající o tom, nakolik jsou hodnoty v souboru dat rozptýleny kolem průměrné hodnoty. Z velikosti směrodatné odchylky lze rovněž činit odhad, s jakou pravděpodobností bychom při opakovaném měření získali stejné hodnoty, resp. jak široké je pásmo kolem naměřené hodnoty, v němž se s určitou vybranou pravděpodobností reálně nachází měřená hodnota. Pro standardní statistická data (dodržující normální rozdělení) platí, že skutečná hodnota měřeného parametru leží s 67% pravděpodobností v intervalu naměřená hodnota \pm jedna hodnota směrodatné odchylky, se zhruba 95% pravděpodobností pak v intervalu naměřená hodnota \pm dvojnásobek standardní směrodatné odchylky.

srovnávací test (NR-test, norm-referenced test)

Test, jehož cílem je porovnat kvality testovaných, tedy zjistit, kdo z testovaných je v testem měřené charakteristice lepší a kdo horší – typickým použitím může být selekce určeného počtu např. uchazečů o studium v situaci, kdy nemůže být přijat každý, kdo splňuje kvalifikační předpoklady. Pro jeho strukturu a pro charakteristiky úloh, které jej tvoří, platí specifická pravidla, odlišná od pravidel platících pro ověřovací test, proto jeho výsledek lze jen omezeně použít pro formulaci jakýchkoli objektivizujících soudů o absolutní úrovni dovedností testovaného.

úspěšnost úlohy

Způsob vyjádření toho, jak testování řešili konkrétní úlohu. Je možné ji vyjádřit jako tzv. hrubou úspěšnost (podíl všech testovaných, kteří úlohu vyřešili správně), je ale také možné získávat přesnější vyjádření úspěšnosti – například vyloučením těch, kdo úlohu nestihli vůbec řešit (někteří z nich by ji vyřešili, jiní ne, jejich zahrnutí do všech testovaných vnáší do hodnoty úspěšnosti chybu), případně odečítáním části bodů za nesprávně zodpovězenou úlohu (tzv. čistá úspěšnost – podíl součtu bodů za danou úlohu všemi testovanými s penalizací těch, kteří chybovali, ku celkovému maximálně možnému počtu bodů získatelných v dané úloze).

validita

Vyjádření vhodnosti úlohy nebo testu pro zamýšlený účel. Obvykle není kvantifikována, lze ji ale doložit např. nezávislými oponentními posudky, protokoly z přípravy úloh či testu. Podle typu a účelu měření rozlišujeme více typů validity, např.:

- obsahová validita udává, nakolik měří daná úloha či test skutečně to, co deklaruje
- kriteriální validita udává, nakolik se shoduje obsah/náročnost úlohy/testu s předem daným standardem
- konstrukční validita udává, zda je test sestaven podle zásad platných pro daný typ měření
- prediktivní validita udává, zda vypovídají výsledky testu věrohodně o předpokladech testovaných dosáhnout v budoucnu požadovaných výsledků (např. v dalším stupni studia apod.)
- souběžná validita udává, nakolik odpovídají výsledky daného měření výsledkům jiných měření.

z-skóre

Způsob vyjádření výsledků testování. Z-skóre říká, o kolik hodnot směrodatné odchylky měření se konkrétní výsledek liší od průměru všech naměřených hodnot. Je sice obvykle méně srozumitelné pro laickou veřejnost, ale v jednoduchých číslech vyjadřuje, nakolik věrohodné jsou rozdíly mezi testovanými nebo odstupy testovaných od kriteriálních rozhraní.

6. SEZNAM LITERATURY

Standardy pro pedagogické a psychologické testování. (2001). Praha: Testcentrum.
Urbánek, T., Denglerová, D., & Širůček, J. (2011). *Psychometrika*. Praha: Portál.

7. PŘÍLOHY

Příloha 1 Posuzovací arch pro výběr didaktického testu – verze s komentáři

<p>1. Potřebuji pro daný evaluační účel srovnání výkonů svých žáků s výkony žáků z jiných škol? Například v situaci, kdy budeme chtít žákům poskytnout co možná nejkvalitnější zpětnou vazbu k jejich práci a stanovit jejich silné a slabé stránky, zvolíme velmi pravděpodobně nějakou komplexnější evaluační úlohu, která nám umožní důkladně posoudit žákovy vědomosti a dovednosti. Budeme-li chtít vyhodnotit práci v delším časovém období, můžeme zvolit ověřovací test, u něž budeme požadovat prokázanou shodu se standardy, které považujeme za relevantní (odpovídají cílům a způsobům naší výuky). V situaci, kdy budeme chtít vědět, jak žáci ovládají základní matematické operace ve srovnání s podobnými žáky v jiných školách, zvolíme test, který nám tuto komparaci umožní. Již při výběru evaluačního nástroje je ale třeba získat od jeho tvůrců informaci o tom, zda bude možné srovnání provést na dostatečně velkém vzorku žáků ze škol, které jsou svými charakteristikami blízké naší škole (srovnání se školami převážně jiných typů zpravidla mnoho užitečného nepřinese).</p>	
<p>2. Potřebuji ohodnotit vědomosti a dovednosti žáků týkající se širšího celku učiva? Autentické evaluační úlohy nám umožní získat hlubší vhled do dílčích vědomostí a zejména dovedností a poskytnout žákům kvalitní zpětnou vazbu, která jim pomůže se dále zlepšovat. Testy se hodí k (povrchnějšímu) hodnocení širšího učebního celku. Dobrou volbou jsou však jen takové testy, u nichž je předem dána stupnice převádějící výsledek v testu na jednotlivé dosažené úrovně, a to na základě odůvodněného a pro nás přijatelného postupu (např. zjištění, že převod výsledků na přibližné klasifikační stupně byl stanoven na základě gymnaziálních standardů, může být pro řadu negymnaziálních středních škol rozumnou překážkou pro využití takových testů).</p>	
<p>3. Je test vhodný pro účel, pro který ho potřebuji? Je doporučená interpretace výsledků shodná s mými potřebami? Vždy bychom měli volit test, který je určen přesně k tomu účelu, pro který ho potřebujeme. Neměli bychom například používat přijímací test k dalšímu studiu k hodnocení práce žáka na nižším stupni. Přijímací test by měl být validizován ve vztahu k úspěšnosti žáka v dalším studiu, při ověřování testu by mělo být prokázáno, že žáci, kteří mají v testu dobré výsledky, jsou úspěšní v dalším studiu. To však nutně neříká nic o tom, kde své vědomosti a dovednosti získali, tedy jak pracovali na nižším stupni. Čím přesněji si před pořízením testu vymezíme svá očekávání a své potřeby, tím lépe můžeme hledat test, který je schopný je co nejlépe naplnit.</p>	
<p>4. Je dostatečně dobře popsáno, jaké vzdělávací cíle test ověřuje? Je dostatečně dobře zacílen, neklade si příliš velké ambice, neověřuje příliš mnoho cílů? Test musí být opatřen podrobným popisem cílů, jejichž dosažení ověřuje, a zdůvodněním jejich důležitosti. Záběr testu by neměl být příliš široký, neboť žák v omezeném testovacím čase může soustředěně zodpovídat pouze malé množství otázek. Příliš ambiciózní záběr testu vede k velmi povrchnímu a nespolehlivému měření.</p>	
<p>5. Poskytne mi tento test validní informaci o výsledcích mých žáků? Zjišťuje to, na co se ve výuce zaměřuji a co považuji za důležité? Vždy si musím položit otázku, zda cíle, které test ověřuje a které autor testu považuje za důležité, jsou důležité také pro mne, zda skutečně stojí v centru mé výuky. Pouze pokud je zacílení testu shodné se zacílením mé výuky, je možno uvažovat o použití testu k ověření výsledků mé práce.</p>	
<p>6. Je testový nástroj dostatečně kvalitní? Je dostatečně hodnověrně doloženo, že testové úlohy spolehlivě ověřují stanovené vzdělávací cíle? Má dostatečně vysokou reliabilitu? Autor testu musí být schopen doložit, jak přesně daný test ověřuje stanovené vzdělávací cíle, a dokumentovat jeho kvalitu. Autor testu by měl být nejen schopen a ochoten předložit informace o tom, jak byl test vytvořen, ale také o tom, jaké má test parametry (před použitím testu jde o parametry pilotních testů = meziproductů při přípravě testu), jaké parametry měly jeho obdobné testy, jaké parametry očekává v ostrém použití testu a jak bude reagovat na případnou situaci, kdy by statistické parametry ostrého testování (reliabilita, průměrná úspěšnost, směrodatná odchylka, ale také třeba nedostatečný vzorek škol celkem nebo škol stejného typu) neumožňovaly věrohodné porovnání výsledků a tím jejich použití pro evaluaci.</p>	
<p>7. Je testový nástroj finančně dostupný? Jsou nároky na administraci přijatelné? Je přijatelná jeho časová náročnost? Čas, který věnujeme testování, bychom mohli využít k nějakým výukovým aktivitám. Proto bychom si měli vždy položit otázku, zda bude tento čas dobře využit, zda výsledky testování jsou pro nás a pro naše žáky natolik užitečné, že za tuto investici stojí.</p>	
<p>8. Je test určen pro žáky podobné těm, které vyučuji? Byl na takových žácích pilotován? Autor testu by měl vždy poskytnout podrobnou informaci o tom, pro jaké respondenty je test určen, na jakých respondentech byl ověřován, případně standardizován. Například test určený pro dobře situované žáky z majoritní společnosti nemusí být kulturně vhodný pro romské žáky z vyloučených lokalit a může o úrovni jejich vědomostí a dovedností přinášet zkreslené informace.</p>	
<p>9. Je zpráva, kterou obdržím, dostatečně informativní? Poskytne mi užitečné a dostatečně podrobné informace, na jejichž základě budu moci modifikovat svoji výuku? Poskytne užitečné informace mým žákům? Při výběru testu je užitečné si učinit představu o podobě a obsažnosti zpětné vazby, která bude zapojeným školám a žákům zaslána. Tato informace je zásadní pro posouzení, zda bude investice do testu efektivní. Následující body podrobně rozpracovávají, na co je třeba se zaměřit při posuzování zprávy.</p>	

<p>10. Jsou výsledky testu prezentovány srozumitelným způsobem? Je podrobně vysvětleno, jakým způsobem byly získány?</p> <p>Některé testovací společnosti například odečítají body za nesprávné odpovědi, jiné přikládají různým úlohám různou váhu v závislosti na jejich obtížnosti. Výsledky mohou být prezentovány pomocí hrubého skóre nebo procentuální úspěšnosti, mohou však také pracovat s umístěním jednotlivých žáků ve srovnání s ostatními. Při různých prezentacích se mohou výsledky našich žáků jevit v odlišném světle, z toho důvodu je důležitá podrobná informace o výpočtu výsledků. Vedle toho by mělo být pravidlem, že autoři testu poskytnou výsledky žáků v různé podobě, z nichž každá umožní pohled na výsledky z jiného užitečného úhlu.</p>	
<p>11. Budu mít dostatečně podrobnou informaci o tom, s kým jsou moji žáci porovnáváni? Je srovnání s ostatními zúčastněnými školami a žáky relevantní?</p> <p>V ČR jsou školám (i jednotlivým žákům) často nabízeny testy, které prezentují výsledky na základě pořadí ve srovnání s výsledky žáků z ostatních zúčastněných škol. Pro interpretaci výsledků je třeba mít dostatečné informace o tom, jaké další školy a žáci se testování zúčastnili. Je důležité vědět, s kolika ostatními žáky jsou výsledky našich žáků porovnávány, zda se jedná o žáky škol městských či venkovských, v jakých regionech se nacházejí, jaké je složení jejich žáků, jejich velikost, případně zaměření.</p>	
<p>12. Je uvedeno srovnání pro různé skupiny žáků?</p> <p>Informace o složení kontrolní skupiny je zásadní, ale někdy není pro interpretaci výsledků dostatečná. Je velmi užitečné mít samostatná srovnání pro dílčí skupiny podle typu školy, typu studia, sídla školy, socio-ekonomického statusu a podobně.</p>	
<p>13. Získám ze zprávy kvalitní informaci o tom, co moji žáci umějí a co se ještě potřebují naučit?</p> <p>I v případě, že je pro nás nejdůležitější zjistit, jak si naši žáci vedou ve srovnání s jinými žáky, měla by zpráva obsahovat také informaci o tom, jaké vědomosti a dovednosti mají žáci, kteří dosáhli určitého výsledku. Tato informace by měla být nezávislá na skupině, s kterou je srovnávání prováděno. Může se ukázat, že v situaci, kdy se srovnání z nějakého důvodu účastní samí výborní žáci, mohou být naši žáci ve srovnání podprůměrní, ale přesto jejich vědomosti a dovednosti mohou odpovídat našim představám a nárokům, které na ně klademe.</p>	
<p>14. Poskytuje zpráva informace o vědomostech a dovednostech v dílčích oblastech?</p> <p>Jak bylo uvedeno výše, kvalitní test by měla doprovázet informace o tom, jaké vědomosti a dovednosti ověřuje. Zpráva by pak měla prezentovat nejen celkové výsledky, ale i dílčí skóre a popis vědomostí a dovedností, které jim odpovídají, aby umožnila srovnání v jednotlivých okruzích vědomostí a dovedností. Tyto okruhy by ale měly být dostatečně popsány a vysvětleny – rozdíly mezi okruhy by měly být srozumitelné a zřetelné, vymezení okruhů by mělo být doplněno ilustrativními příklady.</p>	
<p>15. Poskytuje zpráva informace o výsledcích v jednotlivých úlohách? Jsou úlohy k dispozici?</p> <p>V případě testu, který má poskytovat zpětnou vazbu, je třeba, aby byly spolu se zprávou poskytnuty úlohy. Umožní pedagogům si učinit představu o tom, jaké typy úloh se jejich žákům dařily a jaké jim naopak činily obtíže. Informace o úspěšnosti žáků v jednotlivých úlohách poskytne pedagogům nejpřesnější informaci o tom, jestli žáci naplňují nároky, které jsou na ně kladeny. U jednotlivých úloh by mělo být uvedeno nejen jejich zařazení do tematických a dovednostních okruhů, samozřejmostí by mělo být uvedení průměrných úspěšností pro všechny žáky a pro jednotlivé skupiny žáků z různých škol (za např. celkově průměrnou úspěšností úlohy se mohou ukrývat významné rozdíly v tom, jak danou úlohu řešili žáci v různých typech škol, v různé velikosti a podobě.).</p>	
<p>16. Je uvedena souvislost mezi výsledky v testu a školní známkou?</p> <p>Tato informace pomáhá učitelům učinit si představu o přísnosti jejich klasifikace, a to ve srovnání s učiteli jiných tříd (účastní-li se šetření více tříd z téže školy) i ve srovnání s učiteli jiných podobných škol. Vzhledem k důležitosti, kterou má klasifikace v našich školách, bychom měli tuto informaci vyžadovat. Stejně jako u jiných předešlých kritérií i zde by ale mělo být samozřejmostí vyhodnocení korelace mezi výsledkem testu a známkou po jednotlivých skupinách škol (nepřekvapí, že například stejnému výkonu v testu může na různých středních školách odpovídat různá známka, jakkoli je diskutabilní, zda je to dobře).</p>	
<p>17. Přináší zpráva nějaké zajímavé kontextuální informace?</p> <p>Aby srovnání bylo relevantní, je třeba, aby srovnávalo srovnatelné. Proto je třeba, aby součástí šetření bylo i zjišťování kontextuálních proměnných, jako je typ školy, rodinné zázemí žáků a podobně. Výsledky ovšem mohou být s užitkem interpretovány i na pozadí dalších proměnných. Důležitá je například informace o oblíbenosti testovaných předmětů nebo sebekojetí žáků ve vztahu k těmto předmětům. Výzkumy ukazují, že motivace a sebedůvěra jsou stejně důležitými předpoklady úspěchu jako kognitivní dovednosti a vědomosti.</p>	
<p>18. Budu moci na tyto výsledky navázat v budoucnosti? Poslouží mi k nějakým dalším srovnáním?</p> <p>Výsledky testování můžeme použít jako jednorázovou informaci, můžeme je však také použít pro srovnání dlouhodobější. Například můžeme chtít pravidelně testovat žáky určitého ročníku, abychom zjistili, zda jsme schopni vybavovat naše žáky stále stejnými vědomostmi a dovednostmi. V tom případě potřebujeme vědět, zda je dodavatel testu schopen garantovat v jednotlivých bázích testování srovnatelnou obtížnost testů. Můžeme také chtít dlouhodobě sledovat stejnou skupinu žáků a zjišťovat, jak se v průběhu studia zlepšují. V tom případě potřebujeme vědět, zda dodavatel testu nabízí nějaká měření, která nám umožní porovnávat výsledky žáků v jednotlivých letech.</p>	

Příloha 2 Posuzovací arch pro výběr autentické evaluační úlohy – verze s komentáři

<p>1. Vyžaduje vyřešení této úlohy vědomosti a dovednosti, které tvoří důležitou součást kurikula? Jedná-li se o větší časovou investici, je třeba ji zacílit na vědomosti a dovednosti, které jsou základní a měli by si je z výuky odnést všichni žáci. Musí se jednat o vědomosti a dovednosti, které jsou ve výuce rozvíjeny a na které je kladen velký důraz.</p>	
<p>2. Stojí úloha za čas a energii, která je nezbytná k jejímu řádnému vyřešení? I když je úloha zacílena na základní vědomosti a dovednosti, může být neefektivní: žáci při jejím řešení mohou ztrácet čas zbytečnostmi, které je nijak nerozvíjejí a jejichž informační hodnota je minimální.</p>	
<p>3. Je úloha pro žáky dostatečně náročná a motivující? Řešení úlohy by mělo být pro žáky intelektuálně stimulující.</p>	
<p>4. Pomůže zpětná vazba, kterou získají v průběhu práce na úloze, žákům vyhodnotit jejich práci a přizpůsobit ji danému úkolu? Otázky vycházejí z předpokladu, že by evaluace měla být v první řadě užitečná pro učení žáků. A také z předpokladu, že při řešení evaluační úlohy jsou žáci motivováni se učit. Tato otázka předjímá, že žákům bude v průběhu řešení úlohy poskytnuta podpora, která jim umožní rozpoznat slepé uličky a své řešení přehodnotit. Učitel se samozřejmě může rozhodnout, že v průběhu řešení podporu poskytovat nebude a poskytne zpětnou vazbu až na základě celkového výsledku.</p>	
<p>5. Pomohou žákům známky, které dostanou za danou úlohu, a s nimi spojené informace získat zpětnou vazbu a zlepšit jejich práci? Vzhledem k tomu, že se jedná o evaluační úlohu, je odpověď na tuto otázku zcela zásadní. Vyhodnocení úlohy musí žákům poskytnout důležité informace o tom, co již zvládají a kde by se měli ještě zlepšit a jak.</p>	
<p>6. Je návod na vyhodnocování žákovských odpovědí dostatečně podrobný a jednoznačný? Podrobnost a jednoznačnost návodu na vyhodnocování odpovědí je důležitá v případě, že bude úspěšnost žáků v úlohách srovnávána s normami nebo s výsledky žáků z jiných škol.</p>	
<p>7. Je úloha určena pro žáky podobné těm, které vyučují? Byla na takových žácích pilotována? Autor úlohy by měl vždy poskytnout podrobnou informaci o tom, pro jaké respondenty je úloha určena, na jakých respondentech byla ověřována. Důležitý je nejen věk žáků a typ vzdělávacího programu, ale i jejich rodinné zázemí. Pokud by byla úloha ověřována pouze na žácích určité skupiny, mohla by ve vztahu k jiným skupinám poskytnout zavádějící informaci.</p>	
<p>8. Budu mít dostatečně podrobnou informaci o tom, s jakým výkonem je výkon mých žáků porovnáván? Byla úloha standardizována na dostatečném počtu dostatečně různorodých žáků? Je srovnání s ostatními zúčastněnými školami relevantní? Pokud je pro nás důležité porovnání výsledků našich žáků v úloze s výsledky nějakých celorepublikových norem, musíme mít jistotu, že tyto normy jsou skutečně vytvořeny na vzorku žáků, který dobře reprezentuje žáky dané věkové kategorie v ČR. Stejně tak provádíme-li porovnání s žáky několika škol, musíme vědět, o jaké školy se jedná, zda jejich výběr je dostatečně reprezentativní.</p>	

Příloha 3 Posuzovací arch pro výběr didaktického testu – verze bez komentářů

1. Potřebuji pro daný evaluační účel srovnání výkonů svých žáků s výkony žáků z jiných škol?	
2. Potřebuji ohodnotit vědomosti a dovednosti žáků týkající se širšího celku učiva?	
3. Je test vhodný pro účel, pro který ho potřebuji? Je doporučená interpretace výsledků shodná s mými potřebami?	
4. Je dostatečně dobře popsáno, jaké vzdělávací cíle test ověřuje? Je dostatečně dobře zacílen, neklade si příliš velké ambice, neověřuje příliš mnoho cílů?	
5. Poskytne mi tento test validní informaci o výsledcích mých žáků? Zjišťuje to, na co se ve výuce zaměřuji a co považuji za důležité?	
6. Je testový nástroj dostatečně kvalitní? Je dostatečně hodnověrně doloženo, že testové úlohy spolehlivě ověřují stanovené vzdělávací cíle? Má dostatečně vysokou reliabilitu?	
7. Je testový nástroj finančně dostupný? Jsou nároky na administraci přijatelné? Je přijatelná jeho časová náročnost?	
8. Je test určen pro žáky podobné těm, které vyučuji? Byl na takových žácích pilotován?	
9. Je zpráva, kterou obdržím, dostatečně informativní? Poskytne mi užitečné a dostatečně podrobné informace, na jejichž základě budu moci modifikovat svoji výuku? Poskytne užitečné informace mým žákům?	
10. Jsou výsledky testu prezentovány srozumitelným způsobem? Je podrobně vysvětleno, jakým způsobem byly získány?	
11. Budu mít dostatečně podrobnou informaci o tom, s kým jsou moji žáci porovnáváni? Je srovnání s ostatními zúčastněnými školami a žáky relevantní?	
12. Je uvedeno srovnání pro různé skupiny žáků?	
13. Získám ze zprávy kvalitní informaci o tom, co moji žáci umějí a co se ještě potřebují naučit?	
14. Poskytuje zpráva informace o vědomostech a dovednostech v dílčích oblastech?	
15. Poskytuje zpráva informace o výsledcích v jednotlivých úlohách? Jsou úlohy k dispozici?	
16. Je uvedena souvislost mezi výsledky v testu a školní známkou?	
17. Přináší zpráva nějaké zajímavé kontextuální informace?	
18. Budu moci na tyto výsledky navázat v budoucnosti? Poslouží mi k nějakým dalším srovnáním?	

Příloha 4 Posuzovací arch pro výběr autentické evaluační úlohy – verze bez komentářů

1. Vyžaduje vyřešení této úlohy vědomosti a dovednosti, které tvoří důležitou součást kurikula?	
2. Stojí úloha za čas a energii, která je nezbytná k jejímu řádnému vyřešení?	
3. Je úloha pro žáky dostatečně náročná a motivující?	
4. Pomůže zpětná vazba, kterou získají v průběhu práce na úloze, žákům vyhodnotit jejich práci a přizpůsobit ji danému úkolu?	
5. Pomohou žákům známky, které dostanou za danou úlohu, a s nimi spojené informace získat zpětnou vazbu a zlepšit jejich práci?	
6. Je návod na vyhodnocování žákovských odpovědí dostatečně podrobný a jednoznačný?	
7. Je úloha určena pro žáky podobné těm, které vyučují? Byla na takových žácích pilotována?	
8. Budu mít dostatečně podrobnou informaci o tom, s jakým výkonem je výkon mých žáků porovnáván? Byla úloha standardizována na dostatečném počtu dostatečně různorodých žáků? Je srovnání s ostatními zúčastněnými školami relevantní?	

evaluační nástroje

1. Rámec pro vlastní hodnocení školy. Metodický průvodce.
2. Anketa pro rodiče. Anketa škole na míru.
3. Anketa pro žáky. Anketa škole na míru.
4. Anketa pro učitele. Anketa škole na míru.
5. Analýza dokumentace školy. Katalog kvantitativních indikátorů.
6. Dobrá škola. Metoda pro stanovení priorit školy.
7. Příprava na změnu. Metoda pro předjímání reakcí lidí.
8. Rámec profesních kvalit učitele. Hodnoticí a sebehodnoticí arch.
9. Profesní portfolio učitele. Soubor metod k hodnocení a sebehodnocení.
10. 360° zpětná vazba pro střední úroveň řízení. Soubor dotazníků a metodických doporučení.
11. Metody a formy výuky. Hospitační arch.
12. Učíme děti učit se. Hospitační arch.
13. Výuka v odborném výcviku. Hospitační arch.
14. Klima školy. Soubor dotazníků pro učitele, žáky a rodiče.
15. Klima školní třídy. Dotazník pro žáky.
16. Klima učitelského sboru. Dotazník pro učitele.
17. Společenství prvního stupně. Dotazník pro žáky formou počítačové hry.
18. Interakce učitele a žáků. Dotazník pro žáky.
19. Připravenost školy k inkluzivnímu vzdělávání. Dotazník pro učitele.
20. Předcházení problémům v chování žáků. Dotazník pro žáky.
21. Poradenská role školy. Posuzovací arch.
22. Mapování cílů kurikula. Posuzovací arch.
23. Strategie učení se cizímu jazyku. Dotazník pro žáky.
24. Školní výkonová motivace žáků. Dotazník pro žáky.
25. Postoje žáků ke škole. Dotazník pro žáky.
26. **Zjišťování a vyhodnocování výsledků vzdělávání žáků. Posuzovací archy.**
27. ICT v životě školy – Profil školy²¹. Metodický průvodce.
28. Internetová prezentace školy. Posuzovací arch.
29. Skupinová bilance absolventů. Metoda pro hodnocení průběhu školního vzdělávání.
30. Zpětná vazba absolventů a firem. Dotazník pro absolventy a metoda rozhovoru pro zástupce spolupracujících firem.

