

Evaluace

Hned na začátku je třeba konstatovat, že tato oblast pedagogické evaluace je dosud ve stavu, jenž je charakterizován značnou nejasností, izolovaností dílčích přístupů a nerovnoměrností vývoje. To je nejenom naše konstatování, ale i zhodnocení publikované v nejnovějším materiálu OECD zabývajícím se evaluací kurikula (Pelgrum, Voogt, Plomp, 1995).

Tito autoři provedli analýzu více než 70 výzkumných prací o evaluaci kurikula a dospěli k závěrům, že

- . na jedné straně neexistuje shodná terminologie v oblasti evaluace kurikula, resp. vzdělávacích programů
- . na druhé straně evaluace kurikula má dosud jen deskriptivní charakter, tj. popisuje vlastnosti kurikula, avšak není použitelná k objasňování očekávaných výsledků a efektů kurikula.

V důsledku této poněkud nepřehledné situace musíme nejprve vyjasnit základní pojmy v této oblasti:

. Evaluace vzdělávacích programů (program evaluation, instructional design evaluation) je analytická a hodnotící činnost směřující obvykle k vyhodnocování jednotlivého kurzu ve školním či vysokoškolském vzdělávání nebo velmi často v profesním výcviku zaměstnanců (kurzy ve smyslu školení apod.). Princip je takovýto: Je naplánován např. výcvikový kurz k zlepšení odborné kvalifikace pracovníků podniku a tento kurz má určitý obsah, cíle a prostředky výuky. Po jeho realizaci se zjišťuje, zda plánovaný kurz splňuje očekávání, která do něho vložili jeho konstruktéři.

. Evaluace kurikula (curriculum evaluation, curriculum assessment) je analytická a hodnotící činnost komplexnější povahy a širšího rozsahu než evaluace jednotlivých vzdělávacích programů. Mohli bychom zde vyjít z koncepce E. Walterové (1994), podle níž kurikulum má několik různých dimenzí, které lze vyjádřit určitými otázkami (zde mírně modifikovanými):

Dimenze kurikula

PROČ vzdělávat? KOHO vzdělávat? CO vyučovat? KDY vzdělávat? JAK vzdělávat? Za jakých PODMÍNEK? S jakými EFEKTY?

Komponenty kurikula

FUNKCE a CÍLE vzdělávání, CHARAKTERISTIKY SUBJEKTŮ vzdělávání, OBSAH vzdělávání, ČAS vzdělávání, METODY a PROSTŘEDKY vzdělávání, ORGANIZACE vzdělávání, KONTROLA a HODNOCENÍ vzdělávání

Podle tohoto pojetí by evaluace kurikula měla být soustředěna na poslední dimenzi, tj. efekty očekávané po realizaci určitého kurikula. Ve skutečnosti je evaluace kurikula širší záležitostí, neboť pokrývá všechny uvedené dimenze, i když ve značně nerovnoměrné intenzitě. Tuto záležitost nyní vysvětlíme podrobněji.

Budeme přitom vycházet z analýzy evaluačních výzkumů kurikula, jak ji provedli pro účely OECD nizozemští odborníci Pelgrum, Voogt, Plomp (1995). Ti rozlišují čtyři skupiny charakteristik, které byly zkoumány v rámci evaluace kurikula. Jednotlivé skupiny zde uvedeme a k nim připojíme příklady z velkého počtu charakteristik (z nichž každá je v uvedené práci dokumentována nějakým publikovaným empirickým výzkumem):

(1) Obsah kurikulárních materiálů (celkem 43 charakteristik)

Jde o charakteristiky vztahující se jednak k obsahu plánovaného kurikula (např. evaluace toho, jaká témata pokrývá učivo v určitých vyučovacích předmětech), jednak k obsahu realizovaného kurikula (např. jaké typy úkolů s určitými obsahy žáci provádějí v reálných vyučovacích hodinách).

Příklady evaluačních charakteristik:

- . rozsah učiva
- . důraz na dovednosti
- . stupeň obtížnosti učiva
- . typy úkolů

(2) Časová dimenze kurikula (celkem 27 charakteristik)

Jde o charakteristiky vyjadřující množství času (výukových hodin, dnů, roků), které jsou v kurikulárních materiálech věnovány určitým komponentám obsahu (např. čas pro jednotlivé vyučovací předměty, skupiny předmětů aj.). V této části evaluace existuje již velký počet prací a kvantitativních indikátorů, které ukazují, jaké jsou časové charakteristiky kurikula ve školských systémech různých zemí. Nejnověji byl k této evaluaci vypracován materiál Evropské unie Key Data on Education in the European Union (1995). Dlouholetý komparativní výzkum v této oblasti byl prováděn v Německém ústavu pro mezinárodní pedagogický výzkum (DIPF) ve Frankfurtu n. Mohanem - viz nejnověji publikaci Die Zeitdimension in der Schule als Gegenstand des Bildungsvergleichs (Mitter, von Kopp, 1994, aj.). U nás k tomu publikovala stat V. Ježková (1993).

Příklady evaluačních charakteristik:

- . počet vyučovacích hodin na předměty
- . délka školního rok

- . délka povinné školní docházky
- . čas vymezený na domácí úkoly

(3) Organizace a profil výuky (celkem 44 charakteristik)

Do této skupiny jsou zařazeny především ty charakteristiky, které jsou zadány v kurikulárních programech, ale i některé charakteristiky reálného procesu výuky.

Příklady evaluačních charakteristik:

- . metody používané ve výuce
- . počet a druh používaných učebnic ve výuce
- . důraz kladený ve výuce na hodnocení žáků
- . začleňování skupinového učení do výuky

(4) Participanti kurikula (celkem 58 charakteristik)

Termínem participanti (participants) jsou v dané práci označováni subjekty, kteří jsou začleněni v procesech realizace kurikula (žáci, studenti, učitelé). Existuje velký počet dílčích charakteristik těchto participantů, které byly podrobovány evaluaci (např. vliv rozdílů mezi chlapci a dívkami na realizaci učiva, věkové charakteristiky žáků aj.).

Příklady evaluačních charakteristik:

- . velikost školy (v počtu žáků)
- . počet učitelů připadající na počet žáků
- . procento učitelů, kteří jsou začleněni do dalšího vzdělávání
- . počet žáků v soukromých školách

Tato klasifikace vyvinutá nizozemskými odborníky (Pelgrum, Voogt, Plomp, 1995) zahrnuje celkově 182 různých charakteristik, s jejichž využitím lze provádět evaluaci kurikula - přičemž všechny tyto charakteristiky jsou reálné v tom smyslu, že jejich působení bylo prokázáno určitými empirickými výzkumy. Avšak i přes velký počet těchto charakteristik nemůže být tento přístup k evaluaci kurikula považován za optimální, neboť v něm postrádáme dvě důležité složky evaluace:

Evaluace proveditelnosti kurikula (curriculum feasibility)

Analýza proveditelnosti (feasibility study) není zatím v pedagogických výzkumech příliš obvyklá, ačkoli je pro účely praxe velice důležitá. Tato analýza vyhodnocuje, za jakých podmínek a předpokladů je realizovatelné to, co se plánuje realizovat. Při evaluaci kurikula to znamená: Předtím, než je určité kurikulum (např. na národní úrovni) nebo nějaký vzdělávací program (např. projektovaný jen v jedné škole) zaveden do praxe, měla by být analyzována jeho proveditelnost, tj. zhodnocení toho, za jakých reálných podmínek je vůbec toto kurikulum realizovatelné v praxi. Tato evaluace proveditelnosti by měla poskytovat predikci možné úspěšnosti nebo neúspěšnosti, resp. možných potíží, plánovaného kurikula. Jde při- tom o takovéto otázky:

- . Jak adekvátně je kurikulum (obsah vzdělávání aj.) přizpůsobeno možnostem subjektů určitého věku (zejména možnostem osvojovat si plánované učivo)?
- . Mají subjekty nezbytné předchozí poznatky a dovednosti pro vykonávání činností předpokládaných kurikulem?
- . Jsou učitelé dostatečně připraveni na realizování nového kurikula? atd.

Evaluace proveditelnosti kurikula (či jiných vzdělávacích konstruktů) má značný význam tehdy, je-li uskutečněna před zaváděním těchto konstruktů do praxe. Avšak je známo, že právě u nás v minulosti byly zaváděny rozsáhlé změny kurikula (při různých školských reformách) bez evaluace proveditelnosti. Po jejich zavedení byly pak shledávány závažné problémy (např. nepřiměřeně velký rozsah učiva v jednotlivých předmětech), které se s velkými škodami projevovaly v praxi.

Proč se evaluace proveditelnosti prosazují nedostatečně v rozhodovacích procesech vzdělávací politiky? Patrně hlavním důvodem je to (jak uvádí Sosniak, 1989), že tato evaluace je časově a finančně nákladná. Předpokládá využití expertů - např. v případě proveditelnosti kurikula tým pedagogů, předmětových didaktiků, psychologů aj. - a pilotážní ověření v praxi, a proto je "nepopulární".

Evaluace efektivnosti kurikula

Mnoho úsilí je vyvíjeno na vyhodnocování toho, jaké produkty (výstupy) vytváří určité kurikulum. Tyto produkty jsou objektivně zjišťovány a měřeny jakožto vzdělávací výsledky. To představuje jádro a nejvíce rozvinutou součást pedagogické evaluace vůbec - proto se jí zabýváme podrobněji v kap. 8.

Existují zde však dvě problematické okolnosti:

- . Jednak ta, že se v mnoha evaluačních pracích předpokládá, že vzdělávací výsledky jsou přímým důsledkem charakteristik kurikula. To ovšem neodpovídá realitě, neboť ve vzdělávacích procesech působí vedle kurikula ještě determinanty z jiných zdrojů. Druhá potíž vzniká tím, že část výzkumných pracovníků, zvláště v USA, považuje za "kurikulum" jakýkoliv vzdělávací program, třeba krátkodobý kurz či část učiva apod. Potom

jakékoliv změny v učivu některých předmětů jsou označovány za "nové kurikulum": Typickým příkladem je výzkumná studie Nichollse, Nelsona, Gleavesové (1995) o "evaluaci dvou kontrastních kurikulů"; otištěná v americkém časopise Journal of Educational Psychology. V pojetí těchto autorů bylo jedním typem kurikula učivo o dynosaurech, zaměřené na fakta (např. s důrazem na pojmenování různých druhů dinosaurů), druhým typem kurikula bylo totéž učivo, avšak zaměřené na problémy (např. hypotézy o tom, proč dynosauři vyhynuli). Tato dvě rozdílná "kurikula" byla hodnocena žáky 1. - 8. ročníku americké základní školy, a to prostřednictvím řízených rozhovorů s žáky.

Autoři dospěli k závěru, že žáci nižších ročníků nerozlišují typy kurikulů, kdežto žáci vyšších ročníků (7: 8.) jsou schopni rozlišovat a vyhodnocovat "kontrastní kurikula" (v daném případě hodnotili pozitivněji kurikulum zaměřené problémově). Zároveň se v této práci poukazuje na důležitost toho, aby při evaluacích kurikula byla věnována pozornost také "hlasu žáků".

I když američtí autoři píšou v tomto výzkumu o rozdílných či kontrastních kurikulích, ve skutečnosti se jedná o typy výuky, resp. způsoby prezentace obsahu učiva ve výuce - v našich termínech bychom řekli výuka klasická (zaměřená na poznatky) vs. výuka heuristická (zaměřená na objevování problémů a jejich řešení). Uvádíme zde tento případ evaluace kurikula jako ilustraci konfúze pojmů, jež se v evaluačních výzkumech vyskytuje bohužel dosti často.

Nyní se budeme zabývat některými konkrétními aplikacemi, z nichž bude zřejmý způsob evaluační procedury a sledovaný praktický účel.

APLIKACE

A (4): VÝBĚR UČIVA PRO VZDĚLÁVACÍ PROGRAM

Jako příklad aplikace uvádíme starší, dnes již málo dostupnou a zapomenutou práci J. Langra a J. Váni (1944) Výzkum učebné látky z fyziky na škole národní. Tato práce publikovaná před půlstoletím překvapuje i dnes důvtipností a aplikovatelností své výzkumné metody.

Cílem tohoto výzkumu (provedeného již v letech 1937-39, ale v důsledku válečných poměrů publikovaného až později) bylo řešit problém "výběru učebné látky a stanovení učebného minima, které by odpovídalo současné praktické potřebě i schopnostem žactva" (s.3). Tento problém byl zkoumán na učivu fyziky pro tehdejší měšťanskou školu, a to na učivu ztvárněném v učebnicích. Postupovalo se takto:

. Nejprve byl analyzován současný stav učiva, tedy učivo v té podobě, jak bylo prezentováno v tehdejších učebnicích.

. V další etapě byl zjištěný a vyhodnocený stav učiva korigován prostřednictvím návrhů na výběr nového obsahu učiva fyziky.

Evaluační analýza učiva spočívala ve zjišťování jeho obsahových komponentů nazývaných "učebné prvky"; jež byly rozčleněny do čtyř kategorií poznatků. Rozborem jednotlivých učebnic byl získán inventář - celkem 605 učebných prvků fyzikálního učiva. V další etapě byl tento inventář vyhodnocován, a to podle tří kritérií:

(a) důležitost učebného prvku, a to s rozlišením dvou druhů významnosti učiva:

. teoretická důležitost učebného prvku ("jakou má obecně vzdělávací hodnotu a jak je nutný pro porozumění ostatní důležité látce" - s. 45)

. praktická důležitost učebného prvku ("Z hlediska praktického je z učebné látky důležité to, co je třeba znát v praktickém životě" - s. 45)

(b) pochopitelnost učebného prvku, tj. určení toho, do jaké míry je určitý obsahový komponent přístupný pro porozumění žáků daného věku

(c) zajímavost učebného prvku, tj. stupeň zájmu, který může v žácích vyvolávat určitý obsahový komponent učiva.

Evaluaci učiva podle uvedených kritérií prováděla skupina kompetentních posuzovatelů, složená jednak ze zkušených učitelů, jednak z odborníků působících ve sféře praxe, v níž se fyzikální znalosti a dovednosti uplatňují.

Konečně závěrečná část tohoto rozsáhlého výzkumu byla zaměřena na nový výběr "minima učebné látky" (v dnešních termínech: základní učivo) pro učebnice. I zde byla procedura výběru velmi promyšlená:

Na základě předchozích etap evaluace byl nejprve stanoven rozsah učiva, a to vzhledem k normovanému vyučovacímu času (podle osnov) a k stupni důležitosti učebných prvků. Jako přiměřený rozsah stanovili autoři soubor zhruba 300 prvků a v důsledku toho navrhli zredukovat rozsah učiva v stávajících učebnicích asi o polovinu. Pro jednotlivé obory fyziky pak stanovili příslušné rozsahy učiva, ostatní učivo zařadili mezi "látku rozšiřovací": Celá práce je završena tabulkami, v nichž je každý jednotlivý učební prvek popsán deseti parametry (z toho pět parametrů určujícími jeho důležitost), které rozhodují o jeho zařazení do základního učiva.

Je podivuhodné, jak přesným a důkladným způsobem přistupovali J. Langr a J. Váňa před více než půlstoletím k problematice výběru učiva na základě evaluační procedury. Jejich přístup -"Spolehlivou základnu pro výběr učiva lze získat objektivním výzkumem " (s. 119) - má aktuální platnost dodnes.

A (5): EVALUACE ODLIŠNÝCH VZDĚLÁVACÍCH PROGRAMŮ PRO STEJNÝ DRUH ŠKOLY

V pedagogické evaluaci se zřídka provádí to, aby byly podrobovány srovnávacímu hodnocení odlišné vzdělávací koncepce, programy apod., avšak fungující na téže druhu škol. Je charakteristické pro dřívější školské reformy (nejen u nás), že se automaticky předpokládalo, že každá "nová koncepce" vzdělávání je vždy lepší, "efektivnější" apod. než koncepce předchozí-aniz se stará a nová koncepce podrobovaly skutečné evaluaci.

Ojedinelý evaluační výzkum v této oblasti provedli u nás Hejnal a Hejnalová (1986). Je to analýza vědomostí žáků, kteří přicházeli na Střední odborné učiliště ČKD Praha 9, a to jednak z bývalé ZDŠ - základní devítileté školy ("stará koncepce"), jednak z nově zavedené ZŠ - základní školy (" nová koncepce"). Hodnocení bylo prováděno opakovaně čtyři roky po sobě, prostřednictvím stejného vstupního didaktického testu, pokrývajícího učivo z fyziky za 8. ročník. Zkoumaný soubor tvořilo celkem 179 žáků, z toho 100 absolventů ZDŠ a 79 absolventů ZŠ.

Vlastní evaluace spočívala v porovnávání výsledků obou skupin žáků ve vstupním didaktickém testu. Autoři však ve své práci publikovali (Hejnal a Hejnalová, 1986) nejen data o výsledcích testu, ale i ,data reflektující vztah mezi testovými výsledky a prospěchem žáků (známka ve fyzice v 8. ročníku), jak ilustruje následující tabulka (do tabulky jsme doplnili hodnoty variačního koeficientu Vz):

soubor žáků absolventi ZOŠ absolventi ZŠ

h

2,46 2,32

Z

4,40 5, 57

SZ

0,76 0,92

_r fl,19 J, 04

vz

17,2 °/ 25,7°/

U

i5,0 °~

35,8 °1

h = průměrná známka z fyziky v 8. ročníku

z = průměrná "známka" v didaktickém testu

sz = směrodatná odchylka

r = korelace mezi h - z

U = úspěšnost v testu (procentový podíl správně vyřešených úloh) Vz = variační koeficient (sz : z)

Tato data ukazují:

. Absolventi ZŠ (nová 'koncepce) projevili sice lepší fyzikální vědomosti (pokud byly pokryty testem) než absolventi dřívější ZDŠ (stará koncepce), avšak celkově byla úroveň vzdělávacích výsledků v obou skupinách žáků nízká (průměrná úspěšnost u žáků ZŠ jen 35,8 %, u žáků ZDŠ dokonce jen 15,0 %).

. Na nízkou úroveň fyzikálních vědomostí lze usuzovat i z toho, že i v souboru žáků ZŠ ani jedna úloha testu nebyla vyřešena většinou (tj. alespoň 55 %) žáků.

. Hodnocení prospěchu z fyziky známkou je v obou skupinách vždy lepší než výsledek v testu a celkově je korelace mezi hodnocením prospěchu ve škole a výsledkem v testu velmi nízká, u žáků ZŠ prakticky nulová. To ukazuje, že rozpor mezi prospěchem (hodnoceným školní známkou) a skutečnými vědomostmi ve fyzice je značný a v nové koncepci (u žáků ZŠ) se ještě zvýšil oproti staré koncepci (ZDŠ).

. O stejné skutečnosti vypovídá i variační koeficient (Vz): Nerovnoměrnost v rozložení úspěšnosti při testování se u žáků ZŠ zvýšila (Vz = 25,7 %) ve srovnání se žáky ZDŠ, tj. existovaly zřejmě větší difference mezi žáky s lepšími a horšími vědomostmi než u žáků ZDŠ (Vz = 17,2 %).

Obdobný charakter jako předchozí výzkum měla evaluace experimentálně kontrolované změny ve vzdělávacím programu, kterou provedli Kolek, Byčkovský, Lev, Štelcichová (1984). Šlo přitom o učivo matematiky na strojní fakultě ČVUT (tento výzkum je stručně popsán in Průcha, 1989 a).

V běžném jazykovém úzu významy obou termínů splývají. Označují celkově to, co se každodenně odehrává ve školních třídách mezi učiteli a žáky. V odborném vyjadřování není jednota: Někteří odborníci používají termíny vyučování a výuka jako synonyma, jiní odborníci - asi většina - je rozlišují. Oba termíny můžeme vymezit následovně:

. Vyučování označuje činnosti, které vykonává učitel, instruktor apod., při realizaci vzdělávacího procesu v interakci se vzdělávajícími se subjekty. Odpovídající angl. termíny jsou teaching, instruction.

. Výuka je širší pojem, který zahrnuje jednak vyučování (činnosti učitele), jednak učení (činnosti žáka) a také obsah těchto činností, jejich cíle, podmínky realizace, výsledky aj. Odpovídající angl. termín je educational (instructional) process nebo (mnohovýznamové) education.

Toto pojetí zastává také Pedagogický slovník (Průcha, Walterová, Ma-reš, 1995) a je rovněž ve shodě s obecně didaktickou koncepcí J. Maňáka (1994). Budeme je uplatňovat i v následujícím výkladu.

Je možno kvalitu (úroveň) výuky nějak exaktně hodnotit? Čím se odlišuje výuka s vysokým stupněm kvality od výuky s nízkým stupněm kvality? Jaký vliv na úroveň výuky má samotná úroveň vyučovacích činností učitele?

To jsou otázky, které stojí již dávno v centru pozornosti při evaluaci vzdělávacích procesů. V praxi se tato evaluace prováděla především jakožto kontrola: Školní inspektoři či jiní pracovníci školského managementu hodnotili školy a učitele, a to hlavně na základě pozorování vyučovacích hodin. Při tomto hodnocení se opírali spíše o intuitivní kritéria, tj. o své vlastní pedagogické zkušenosti a své představy o žádoucí úrovni výuky.

Je jasné, že tento způsob hodnocení nemohl poskytovat přesná a spolehlivá data k evaluaci výuky. Od 60. let se v mnoha evropských zemích a nyní i u nás) změnila role školní inspekce - z prioritní role kontrolní na roli monitorovací a poradní. Zároveň se postupně vypracovávaly a zdokonalovaly výzkumné nástroje, jimiž lze průběh výuky (vyučování, jednak spolehlivě zaznamenávat, jednak jeho úroveň vyhodnocovat, tj. přesně změřit. Mnohé z těchto nástrojů byly (nebo mohly by být) aplikovány i pro praktické účely evaluace, jak níže objasníme. **JAKÝMI ZPŮSOBY SE ÚROVEŇ VÝUKY DÁ HODNOTIT?**

Nejprve je nutno konstatovat, že existuje velký počet (několik stovek) jednotlivých technik a postupů k evaluaci výuky, které jsou publikovány v zahraničních odborných časopisech a knihách. Jsou specializovány podle vyučovacích předmětů, ročníků školy či věku žáků aj. a není snadné se v nich orientovat. Podle základního přístupu, který uplatňují, je možno je rozdělit na dvě velké skupiny:

(1) objektivní (observační techniky evaluace výuky)

(2) subjektivní (participační techniky evaluace výuky)

Obě tyto skupiny zde stručně objasníme a procedury v nich používané popíšeme jako případy Aplikací.

. **Objektivní techniky** evaluace výuky jsou založeny na pozorování (proto je můžeme nazývat observační techniky). Princip je následující: Pozorovatel (výzkumný nebo jiný pracovník provádějící evaluaci) je přítomen při výuce ve třídě a pomocí nějakého registračního zařízení (záznamový arch, magnetofon, videokamera aj.) zaznamenává určité parametry průběhu výuky - např. počet a druh učitelových otázek k žákům nebo délku časových úseků, v nichž žáci pracují na nějakých úkolech apod. Zaznamenaný průběh výuky se pak podle stanovených kritérií analyzuje a vyhodnocuje, tj. vyvozují se závěry o kvalitě pozorované výuky.

Tento způsob evaluace bývá nazýván objektivní, protože zachovává stejný princip jako při pozorování experimentů v přírodních vědách, kdy výzkumný pracovník nesmí zasahovat do průběhu pozorovaných procesů a musí se opírat jen o vědecky registrovatelný (měřitelný) jejich záznam.

Observační techniky evaluace výuky byly rozvinuty do vysoké dokonalosti v četných variantách a byly a jsou často používány. Také u nás byly poměrně často aplikovány, jak českými autory (viz podrobný výklad a rozsáhlou literaturu in Mareš a Křivohlavý, 1995), tak slovenskými (Gavora a kol., 1988). Většina těchto měření se soustředovala na analýzu komunikace mezi učiteli a žáky (tj. komunikace jako jedné ze složek výuky), pouze výjimečně také na analýzu různých činností učitelů a žáků (Ferko, 1986; Průcha, 1989 c, aj.), resp. na hodnocení kvality učitelových činností ve výuce (Chráška, 1995).

. **Subjektivní techniky** evaluace výuky jsou založeny na jiném principu - nikoliv na pozorování, nýbrž na výpovědích účastníků (participantů) výuky o určitých jejích vlastnostech. Bývají to nejčastěji autoevaluace učitelů (self reporting), tj. výpovědi učitelů (v rozhovorech, v deníkových záznamech aj.) o vlastních vyučovacích činnostech, nebo výpovědi žáků (či jiných subjektů učeny o tom, jak sami vnímají a prožívají výuku. Subjektivní evaluace výuky prováděné žáky se staly součástí širšího hodnocení edukačního prostředí (viz v kap. 7).

Specifickou oblastí evaluace výuky je hodnocení kvality vysokoškolské výuky a práce vysokoškolských učitelů, prováděné studenty. Používá se zde hlavně anonymních dotazníků s položkami škály (rating) hodnotícími jednotlivé kurzy, jednotlivé vyučující apod. (podrobněji o tom Mareš, 1988). V 80. letech se tento typ evaluace prosazoval na některých vysokých školách v USA a jinde, avšak později se od něho ustupovalo, hlavně kvůli potížím vznikajícím z navozování napětí mezi hodnotícími studenty a hodnocenými učiteli. Např. ve Finsku, kde je evaluace vysokých škol velmi rozvinuta a zajišťována na mezinárodní úrovni, se tento typ evaluace odmítá

kvůli "kontradiktorním efektům" (Grčhn, et al., 1993). U nás není tento způsob evaluace uplatňován, neboť se setkával se stejnými výhradami jako většinou v zahraničí.

Celkově se v evaluaci výuky projevuje určitá nevyrovnanost v tom, že většina těchto hodnocení se zaměřuje na vyučovací činnosti učitelů, kdežto hodnocení učení žáků, jejich učebních aktivit, je zřídka prováděno. Možná je zde hlavní překážkou to, že v případě učitele se jedná o jeden subjekt, jehož činnosti ve výuce jsou snadněji registrovatelné, kdežto v případě žáků se jedná o větší či menší skupinu subjektů, jejichž individuální činnosti lze obtížně registrovat ve stejném čase (srov. pozorovací systém OPTIC - aplikace A (15)).

V zahraniční literatuře jsou sice popsány systémy, které mají výraz "evaluace učení" i ve svém názvu, ale ve skutečnosti registrují výsledky učení, nikoliv přímo jeho průběh a obsah. Nejznámější (nikoli u nás, ale v zahraničí z těchto systémů je SOLO Taxonomy (Structure of the Observed Learning Outcome-Struktura pozorovaného výsledku učení, vypracovaná australskými odborníky (Biggs, Collis, 1982; Biggs, Moore, 1993). Tento evaluační systém (obdobně konstruovaný jako známá taxonomie cílů výuky B. S. Blooma) je založen na klasifikaci kognitivních operací žáka podle Piagetovy teorie. Má však spíše výzkumnou aplikaci, pro praktickou evaluaci je nesnadno použitelný.

Někdy se objevují dotazy směřující k tomu, které techniky evaluace výuky jsou "lepší"-objektivní (observační) nebo subjektivní (participační)? Z porovnání předností a nevýhod obou vyplývá: nelze jednoznačně stanovit, že jenom jeden z obou přístupů je lepší než druhý. Optimální je kombinace obou: Chceme-li dosáhnout co nejspolehlivějšího hodnocení výuky, je nutno používat oba uvedené přístupy.

Uvedeme nyní příklady aplikace, v nichž jsou procedury obou přístupů uplatňovány.

EVALUACE VZDĚLÁVACÍCH VÝSLEDKŮ

Jak jsme uvedli výše (v kap. 1), hodnocení vzdělávacích výsledků představuje nejrozsáhlejší součást pedagogické evaluace. Je tomu tak proto, že - z hlediska historického vývoje pedagogického výzkumu - první pokusy o soustavné a exaktní hodnocení pedagogických jevů byly spjaty s hodnocením vzdělávacích výsledků žáků a škol.

Za průkopníky evaluace vzdělávacích výsledků možno považovat dva americké pedagogy 19. století: Henri Barnard (1811-1900) a Horace Mann (1796-1859). Oba se nezávisle na sobě snažili realizovat myšlenku, že "pracovníci řízení školství mohou provádět užitečná rozhodnutí pouze tehdy, mají-li k dispozici soubor spolehlivých informací vystihujících stav vzdělávání ve školách" (cit. podle Encyclopedia of Educational Research, 1992, s.384).

H. Mann ve státě Massachusetts (USA) se přičinil o to, že již v r. 1837 se na základě školského zákona vytvořil výbor odborníků, který se staral o sběr dat týkajících se vzdělávacích výsledků škol. Od té doby má stát Massachusetts (a později i další státy) každoročně publikovaná podrobná data o výsledcích všech svých primárních a sekundárních škol. H. Barnard ve státě Connecticut (USA) založil r. 1855 časopis American Journal of Education (vychází doposud), ve kterém začaly být publikovány zprávy o vzdělávacích výsledcích v amerických školách.

Zajímavý je vývoj evaluačních procedur při hodnocení vzdělávacích výsledků. Thorndike a Hagen (1967) uvádějí, že zhruba do poloviny 19. století se veškeré hodnocení vzdělávacích výsledků v USA provádělo na základě ústních zkoušek. Učitelé nebo examinační kladli otázky a žáci na ně odpovídali. Bylo s tím spojeno mnoho subjektivního hodnocení a nebyla tím také vytvářena možnost pro srovnávání škol aj. Proto se v 2. polovině 19. století prosadilo písemné hodnocení, tj. prostřednictvím esejí: Skupinám žáků byly předkládány stejné otázky a poskytován stejný čas na jejich zodpovězení. Avšak i při tomto způsobu evaluace se projevoval subjektivismus při skórování písemných odpovědí. V důsledku toho se od počátku 20. století postupně prosazovaly standardy hodnocení písemných odpovědí, tj. jednotné ukazatele výkonu vztahující se k určité věkové skupině nebo ročníku školy. Vyvinulo se testování vzdělávacích výsledků.

Americké evaluační techniky byly přenášeny do evropských zemí (včetně Českého království před rokem 1918). Ovšem je třeba zdůraznit, že evropské země (Německo, Anglie, Francie a další) přispívaly k rozvoji pedagogické evaluace vzdělávacích výsledků i vlastním výzkumem. Např. když byla zřízena r. 1799 první katedra pedagogiky na univerzitě v Halle (patrně první svého druhu nejen v Německu, ale ve světě vůbec), vedl ji Ernst Christian Trapp (1745-1818), který zaváděl empirické postupy při hodnocení vzdělávacích výsledků. Také v Anglii docházelo k pokusům o měření vzdělávacích výsledků již od poloviny 19. století. Mezi průkopníky evaluace bývá uváděn George Fisher, jenž v roce 1864 vydal tzv. scale book, tj. soubor písemných testů s hodnotícími stupnicemi pro evaluaci výsledků školního vzdělávání.

Během 20. století se ve vývoji evaluace vzdělávacích výsledků prosazovala značná diferenciací a specializací:

. Na jedné straně se postupně rozvíjelo konstruování testů a jiných evaluačních nástrojů pro jednotlivé vyučovací předměty (nebo skupiny předmětů). Takto se rozvinuly testy pro matematiku a zvláště pro cizí jazyky (language

testing), které dnes představují samy o sobě rozsáhlé a metodologicky hluboce propracované oblasti (o historii vývoje jazykových testů viz v knížce Bubeníkové a kol., 1975).

. Na druhé straně se evaluace vzdělávacích výsledků postupně rozšiřovala teritoriálně: Zpočátku šlo jen o hodnocení a srovnávání výsledků jednotlivých tříd a škol, později (po 2. svět. válce) již o hodnocení na národní úrovni a v současných letech se provádějí mezinárodní evaluace vzdělávacích výsledků zahrnující i země různých kontinentů.

Vzhledem k této faktické rozrůzněnosti a obrovské šíři evaluace vzdělávacích výsledků musíme náš další výklad rozčlenit následujícím způsobem:

Nejprve objasníme evaluační procedury vztahující se k typům škol či skupinám žáků - část 8.1 .

V dalších částech se budeme zabývat evaluacemi vzdělávacích výsledků na národní úrovni, a to jednak v zemích západní Evropy a v USA - část 8.2, jednak na mezinárodní úrovni (komparace vzdělávacích výsledků různých zemí- část 8.3.

V poslední části 8.4 popíšeme dosavadní evaluace vzdělávacích výsledků, které byly provedeny v České republice (resp. v bývalém Československu).

8.1

EVALUACE VZDĚLÁVACÍCH VÝSLEDKŮ V RŮZNÝCH DRUŽÍCH ŠKOL/POPULACÍCH ŽÁKŮ

V české i zahraniční pedagogické evaluaci se nejčastěji setkáváme s hodnocením vzdělávacích výsledků charakterizujících školy určitého typu (stupně) nebo třídy určitého ročníku. Používají se přitom vesměs techniky (hlavně testy) zjišťující kognitivní výsledky (vědomosti) a méně často i některé dovednosti a postoje vzdělávajících se subjektů. Tyto vzdělávací výsledky jsou zjišťovány většinou v jednotlivých vyučovacích předmětech, výjimečně ve více předmětech společně.

Můžeme zde zavést pracovní typologii těchto evaluačních výzkumů, s uvedením příkladů, které jsou podrobněji popsány in Průcha (1989 a):

. Srovnávání vzdělávacích výsledků ve stejném předmětu mezi různými druhy škol

Jde o zjišťování toho, jak jsou vzdělávací výsledky determinovány určitým druhem školy. K tomu účelu musí být evaluace prováděna na srovnatelném objektu, tj. ve stejném předmětu, resp. na stejném učivu.

Tento typ evaluace provedla u nás např. Kaligová (1988). Didaktickým testem měřila vzdělávací výsledky ve fyzice na souboru 739 žáků ze tří druhů škol: gymnázium, střední průmyslová škola, střední odborné učiliště. Zjistila signifikantní rozdíly ve vzdělávacích výsledcích v závislosti na druhu školy (nejúspěšnější byli žáci gymnázia, nejméně úspěšní žáci z učiliště). Vedle toho ještě vypočítávala " ukazatel efektivity výukového procesu "; při jehož výpočtu se bere do úvahy jak počet správných odpovědí v testu, tak čas (počet vyučovacích hodin) věnovaný na učivo, které bylo testováno. Podle tohoto ukazatele byla nejvyšší efektivnost výuky dosahována ve střední průmyslové škole, nižší v gymnáziu a nejnižší v učilišti.

. Srovnávání vzdělávacích výsledků ve stejném druhu škol mezi jednotlivými předměty

Zatímco předchozí typ evaluace (v jednom předmětu) se provádí často, evaluace srovnávající výsledky ve dvou či více předmětech současně jsou prováděny zřídka. U nás realizoval takovouto evaluaci Konečný (1986). Měřil výsledky odborné přípravy žáků učebního oboru zahradník, a to v 8 vyučovacích předmětech (biologie, zahradnická výroba, ovocnictví, aj.). Testoval celkem 852 žáků z 2. ročníku a 977 žáků z 3. ročníku SOU.

Evaluace Konečného (1986) je důležitá nejen proto, že prokázala významné rozdíly ve výsledcích mezi jednotlivými předměty na téže druhu školy, ale také vliv různých vstupních a procesuálních determinant (srov. empirický model efektivity v části 2.2) na úroveň vzdělávacích výsledků: např. kvalifikovanost pedagogických pracovníků na určité škole; časové ztráty ve výuce; materiálně technická vybavenost výuky.

. Hodnocení vzdělávacích výsledků absolventů škol

Některé evaluace se zaměřují na zjišťování vzdělávacích výsledků u žáků, kteří absolvovali základní školu a přicházejí na střední školu nebo u absolventů středních škol vstupujících na vysoké školy. Většinou je tato evaluace prováděna na základě vstupních testů. U nás prováděli takovéto evaluace např. Blažek (1985-86), Švanda (1982), aj.

Závažné jsou evaluace zjišťující vliv absolvované střední školy na studijní úspěšnost na vysoké škole. V 80. letech byla u nás realizována řada výzkumů, které prokazovaly, že tato studijní úspěšnost je skutečně závislá na tom, jaký druh střední školy předtím studenti absolvovali (Kolda, Macháčová, 1986; Šikula, Čermáková, 1987; Kolek, Byčkovský, Lev, Štelcichová, 1984, aj.).

Zvlášť průkazný v tomto ohledu byl výzkum Kloudové a Provazníkové (1987). Hodnoceními subjekty byli studenti 1. ročníku pěti oborů vysokých škol: přírodovědecké fakulty, lékařské fakulty, stavební fakulty, elektrotechnické fakulty, Vysoké školy ekonomické - celkem 1 005 studentů. Zjistilo se, že ačkoliv se neprojevoval rozdíl mezi studenty v intelektových schopnostech, druh předchozího vzdělání (gymnázium, střední odborná škola, střední odborné učiliště s maturitou) měl vliv na studijní úspěšnost na vysoké škole.

. Komplexní evaluace vzdělávacích výsledků a jejich determinant

Výše zmíněné typy evaluace jsou vesměs toho typu, kdy se hodnotí vzdělávací výsledky samy o sobě, tj. předpokládá se, že jsou produktem učiva, které bylo testovaným subjektům nějakým způsobem exponováno a

kteří si v nějaké úrovni osvojili. Stále důrazněji však vyvstává problém, co všechno ovlivňuje (kromě samotného učiva) vytváření určitých vzdělávacích výsledků u jedince. Je to problém komplexu determinantů osobních na vstupu a v průběhu vzdělávacího procesu (viz v části 2.2).

Se zaměřením na tento problém byly u nás již dříve provedeny některé evaluace: Např. Švanda (1989) prokazoval vliv některých postojových, osobnostních a sociálních charakteristik žáků různých druhů středních škol na vzdělávací výsledky v několika předmětech. Šteffl, Botlík, Kalous (1988) ve výzkumu vzdělávacích výsledků v matematice zjišťovali vliv některých sociodemografických determinantů rodinného prostředí žáků (zvl. stupeň vzdělání rodičů) na výsledky žáků gymnázií aj.

Můžeme tyto výzkumy nazývat komplexní evaluace vzdělávacích výsledků (protože zahrnují i vliv různých determinantů) a na jeden z nich se podíváme podrobněji v aplikaci A (17).

~ APLIKACE

A(17): KOMPLEXNÍ EVALUACE VZDĚLÁVACÍCH VÝSLEDKŮ

Komplexní evaluaci vzdělávacích výsledků ve vysokoškolském vzdělávání provedli u nás Mareš a Hartmanová (1986) a publikovali ji v práci nazvané příznačně "Širší pohled na výsledky výuky": Tato práce se pokoušela měnit situaci, kdy se ve výzkumech vzdělávacích výsledků různé determinanty "obvykle jen vyjmenují a jejich zkoumání se odsunuje s poukazem na to, že je jich mnoho, že jejich působení je složité, že se obtížně zjišťují (S. 1 %).

Jak byla tato evaluace pojata?

V daném výzkumu byly předmětem evaluační analýzy vzdělávací výsledky v předmětu biochemie dosahované studenty 2. ročníku na lékařské fakultě (šlo o soubor 30 studentů).

Jako proměnné kontextu, tj. faktory determinující tyto vzdělávací výsledky, byly do analýzy zahrnuty:

(1) charakteristiky studenta - např. kvalita průběžné přípravy, celková studijní zdatnost, vztah k učení, aktivity během výuky

(2) charakteristiky učitele - např. učitelovy znalosti a dovednosti, náročnost na studenty, vážnost učitele u studentů, učitelova odpovědnost za výsledky výuky

(3) výukové cíle předmětu

(4) charakteristiky průběhu výuky (podle hodnocení samotných studentů)

(5) charakteristiky učebnice - stupeň objektivně změřené obtížnosti, subjektivní názory studentů na vlastnosti učebnice

(6) charakteristiky didaktických testů - hlediska spolehlivosti, citlivosti a obtížnosti.

Vliv různých determinantů (proměnných) na vzdělávací výsledky studentů byl odvozován z korelačních koeficientů mezi výsledky a proměnnými. I když konkrétní data zde mají omezenou platnost, tento typ evaluace má metodologickou důležitost, protože ukazuje jak je nutno a možno vysvětlovat vzdělávací výsledky i působením různých determinantů na vstupu a v průběhu vzdělávacího procesu.

Právě popsaný výzkum není u nás jediný, jenž se pokouší vysvětlovat vzdělávací výsledky z působení různých jejich determinantů. Již dříve např. Macák (1983) a Švanda (1989) zahrnoval do evaluace vzdělávacích výsledků postojové charakteristiky žáků, a to

. profesionální perspektiva (vyhraněnost vztahu žáků k budoucímu povolání a vliv, který má představa o budoucím povolání na postoj žáků ke vzdělávání)

. vztah žáků ke střední škole (důvěra žáků ke škole, hodnocení vlastních výkonů ve škole, potřeba úspěšného výkonu, akceptace působení učitelů aj.).

Byly publikovány i některé další komplexní evaluace vzdělávacích výsledků, včetně slovenských výzkumů (viz podrobněji in Průcha, 1989 a). Celkově se ovšem nedaří "uchopit" všechny podstatné determinanty ovlivňující vzdělávací výsledky, neboť jejich počet a charakter působení se zatím vymyká našim výzkumným možnostem - srov. výklad J. Pelikána (1995) o složitosti činitelů působících v procesu výchovy.

8.4 EVALUACE VZDĚLÁVACÍCH VÝSLEDKŮ: ČESKÁ REPUBLIKA

Po roce 1989 se ve vzdělávací politice České republiky začal intenzivně prosazovat trend zaměřený na to, aby se i u nás prováděly evaluace vzdělávacích výsledků. Tento trend je zřetelný v několika směrech:

. Ve vládních, ministerských a jiných dokumentech publikovaných po listopadu 1989 se požadavek evaluace vzdělávacích výsledků objevoval zprvu jako jeden z mnoha nespecifikovaných návrhů. Radikální změna se objevuje v dokumentu Program rozvoje výchovně vzdělávací soustavy v České republice, 1: krok (MŠMT ČR, leden 1992). Tento programový dokument zavádí požadavek evaluace ve významu "státní kontroly nad výsledky škol"; "nad úroveň žáků ze škol vycházejících" a poprvé se zde objevuje požadavek "státních standardů":

"Státní standard ve výsledcích výchovy a vzdělání tvoří státem stanovený soubor znalostí a dovedností žáka (studenta). Je formulován pro obligatorní soubor tak, že zahrnuje nejvíce 75 % rozsahu probíraného učiva" (s. 22)

Od doby publikování tohoto dokumentu se požadavky na evaluaci vzdělávacích výsledků soustavně objevovaly jak v různých ministerských dokumentech, tak v elaborátech některých expertů a pedagogických iniciativ (JČMF, IDEA, aj.). Tyto požadavky nakonec vyústily do konstruování vzdělávacích standardů (viz podrobněji v kap. 9), jako naplnění úsilí hodnotit "kvalitu vzdělávací soustavy".

Postupně se stávalo zřejmé - také z hlediska mezinárodních expertíz, jimiž bylo české školství posuzováno - že i v České republice bude nutno provádět národní evaluace vzdělávacích výsledků. Explicitně byla tato potřeba vyjádřena v programovém dokumentu Kvalita a odpovědnost (MŠMT ČR, říjen 1994, uveřejněném též v Učitelských novinách, 1994, č. 42). V tomto dokumentu, vytyčujícím určité kroky k rozvoji české vzdělávací soustavy, si ministerstvo ukládá v oblasti "hodnocení a péče o kvalitu" (š.4):

"Uskuteční zásadní kvalitativní změnu v práci České školní inspekce, jejíž hlavní funkcí se stane hodnocení práce škol vzhledem k podmínkám průběhu a cílům vymezeným obecně závaznými právními předpisy a jejich vzdělávacím programem . . .

Výstupy školní inspekce i ostatní výsledky evaluace budou jako veřejně přístupné informace sloužit nejen odborné, ale i občanské veřejnosti, například rodičům . . .

Zajistí provádění národních průzkumů zaměřených na zjištění výsledků vzdělávání a analýzu jejich souvislostí s požadavky státu, s metodikou a cíly výuky, profesionalitou učitelů, pojetím a kvalitou učebnic, očekáváním veřejnosti a s dalšími faktory. Zapojí se rovněž do mezinárodních projektů hodnocení podmínek a výsledků vzdělávání . . . (podtrhl J. P.)

Také v jiném dokumentu, Otevřenost, rozmanitost a kvalita - Zásady rozvoje české vzdělávací soustavy (Učitelské noviny, 1994, č. 14, s. 12-

21 ; autor neuveden, avšak dokument má oficiální charakter, neboť byl vypracován pro komisi Poslanecké sněmovny Parlamentu ČR), se jasně formuluje požadavek na národní evaluaci vzdělávání:

"Zavést mechanismy pravidelného vyhodnocování výsledků vzdělávání na jednotlivých školách a úrovních vzdělávací soustavy . . .

Je třeba, abychom se urychleně aktivně zapojili do všech kvalitních mezinárodních srovnávacích výzkumů . . . seznámili širokou veřejnost s jejich výsledky a vyvodili z nich závěry"

Jak je z těchto (a dalších) dokladů patrné, česká vzdělávací politika velmi zřetelně formuluje nutnost provádět evaluace vzdělávacích výsledků na národní úrovni. Formulace postulátů je ovšem jedna věc a uskutečnění evaluací je věc druhá. A právě na to se soustředíme v následujícím výkladu. Popíšeme stručně dosud provedené evaluace na celostátní úrovni a to

. sonda k evaluaci vzdělávacích výsledků na základních školách: ČŠI 1994

. evaluace vzdělávacích výsledků na základních školách: projekt Kalibro, 1995

. srovnávací evaluace vzdělávacích výsledků na státních a soukromých gymnáziích: výzkum ČŠI, 1995

. evaluace čtenářských dovedností žáků základních škol (replika výzkumu IEA): výzkum VÚP, 1995

Každá z těchto evaluací má odlišný obsah a zaměření, společné je jim to, že poskytují obraz o určité složce vzdělávacích výsledků na národní úrovni, a to s využitím objektivních nástrojů měření (testy aj.).