

Typologie metaevaluací

Typology of Metaevaluations

Jiří Remr*

Abstrakt

Text v první řadě definuje metaevaluace, stručně poukazuje na okolnosti vzniku a rozvoje metaevaluací a nastiňuje typické cíle, ke kterým metaevaluační projekty spějí. Hlavní část textu je věnována doposud nepublikovanému rozboru jednotlivých typů metaevaluací; každý z přístupů je nejprve stručně charakterizován a následně jsou uváděny aplikační přednosti a rizika spojená s jeho uplatněním. Na podrobnou typologii metaevaluací a způsobu jejich provedení navazuje samostatná část textu, která je věnována kritériím používaným v metaevaluacích a způsobu jejich stanovení. Text v neposlední řadě poukazuje na vybrané metodologické okolnosti provádění metaevaluací, přičemž hlavní pozornost je věnována výběru primárních evaluací určených k provedení metaevaluací, volbě podkladů, na jejichž základě jsou metaevaluace prováděny a způsobu hodnocení jednotlivých kritérií.

Abstract

The text is trying to define what metaevaluation is and it also briefly points-out the key aspects of its introduction and development. The article also presents typical objectives that metaevaluations anticipate. Main part of the text is devoted to analysis of types of metaevaluations; each of the approaches is described and its main strengths and weaknesses are critically reviewed. Detailed typology of metaevaluations is followed by explication of criteria used for metaevaluations and ways how such criteria are defined. The text presents also selected methodological aspects of conducting metaevaluations; attention is paid to sampling and criteria assessment.

Klíčová slova

metaevaluace, evaluační standardy, zvyšování kvality

Keywords

metaevaluation, evaluation standards, improving quality

Tento článek byl zpracován v rámci projektu podpořeného Grantovou agenturou ČR, grant č. 404/11/0949. Práce byly provedeny v rámci Programu rozvoje vědních oblastí na Univerzitě Karlově (PRVOUK) č. P07.

1. Úvod

Jedním z důležitých cílů evaluací je poskytovat poučení z realizovaných aktivit, jakými mohou být např. implementace programů či projektů nebo uskutečňování konkrétních intervencí či politik (Fitzpatrick et al. 1999; House, Howe 1999; Owen 2006). V tomto ohledu je zřetelná snaha evaluací (a zejména pak evaluací formativních) přinášet užitečné a využitelné náměty na zlepšení prováděných intervencí, odstranění či snížení výskytu implementačních selhání a na zvyšování efektivity prováděných programů (Clarke 1999). Otázkou zůstává, nakolik reflexivní jsou evaluace samotné, resp. do jaké míry sami evaluátoři mohou čerpat podněty ke svému vlastnímu zlepšení a rozvoji evaluačních kompetencí ze svých vlastních aktivit.

Z hodnocení vývoje evaluačního oboru od konce 60. let je patrný soustavný zájem o získávání zpětné vazby týkající se prováděných evaluací adresované evaluátorům samotným. Mnohé evaluační přístupy navrhované prominentními evaluátory (např. Scriven 1973, Stufflebeam 2001, Patton 2008 či House 1980) považují za přirozenou součást rozvoje evaluačních kompetencí a dovedností, že jednotlivé subjekty zapojené v evaluačním procesu¹ vytvářejí specifické postupy a mechanismy umožňující systematické získávání a vyhodnocování zpětné vazby k prováděným evaluacím.

¹ Nemusí se jednat pouze o realizátory evaluací, ale také o pracovníky interních evaluačních jednotek či o zadavatele, případně další uživatele evaluací.

Metaevaluace (dále jen MEV) v tomto ohledu představují zásadní činnost, kterou v druhé polovině šedesátých let popsal M. Scriven (1991) a definoval ji jako evaluaci již provedené evaluace. Bazálním cílem MEV je ověřit, v jakých ohledech a do jaké míry provedené evaluace odpovídají zadání, očekávání či obecně definovaným standardům (Schwandt, Halpern 1988). Nicméně MEV jsou multifunkční, tzn., že obvykle sledují větší počet cílů, a tak MEV mohou kromě již uvedeného poskytovat rozbor kvality provedené evaluace, mohou provádět kontrolu zaměřenou na identifikaci konkrétních nedostatků či nepřesností nebo mohou anticipovat edukativní cíle a přinášet poučení realizátorům evaluací i uživatelům výsledků provedených evaluací.

MEV mohou být rovněž vedeny snahou o provádění evaluace samotných evaluátorů, tedy jako evaluaci evaluátorů. Takováto forma MEV vychází z předpokladu, že samotní evaluátoři jsou jednou ze zainteresovaných stran (Guba, Lincoln 1989) a tak by i oni měli být evaluováni, stejně tak jako ostatní subjekty. Klíčovým motivem je v tomto ohledu snaha eliminovat či alespoň kontrolovat subjektivní prvek v primárních evaluacích a dosáhnout tak vyšší kredibility získávaných evaluačních závěrů². Původní Scrivenova představa o roli a významu MEV byla dále rozpracována a rozšířena D. Stufflebeamem, který chápal MEV nejen jako nástroj hodnocení způsobu provedení primárních evaluací, ale především jako ucelený soubor opatření sloužících k tomu, aby evaluace přinášely jasná zjištění a jednoznačné závěry (Stufflebeam, Shinkfield 2007). V jeho pojetí je patrná snaha pojmout MEV jako praktickou pomůcku určenou samotným realizátorům evaluací.

² Je třeba mít stále na paměti, že se na základě evaluací uskutečňují mnohá rozhodnutí, která navíc mohou mít zásadní význam pro fungování evaluovaných subjektů. Není proto divu, že se rozhodovatelé snaží ujistit, že jejich rozhodnutí stojí na spolehlivých, validních a robustních základech (Stufflebeam, 2001).

2. Typy metaevaluací

Typologie a podrobný rozbor jednotlivých variant a způsobů uspořádání MEV doposud nebyl proveden, a to ani v české a dokonce ne ani v zahraniční literatuře. Aktuální prameny jsou většinou jen kazuistické – popisují konkrétní design, jenž byl v daných podmínkách zvolen, popisují se charakteristiky primárních evaluací. Použité metody a techniky jsou však popsány jen rámcově a nesměřují k formulaci obecných doporučení či návodů, obvykle chybí zobecněné poučení či návod, jak koncipovat MEV v dalších podobných situacích. Tento text je tedy v tomto ohledu prvním nástinem obecněji koncipovaného pojednání o typech MEV.

Následující typologie proto vznikala jednak odvozením z jiných technik používaných v analogických situacích buď přímo v oblasti evaluací (viz rozlišování sumativních a formativních MEV) nebo v jiných oblastech společenskovedního výzkumu (zejm. míra zapojení realizátora), případně přenesením vybraných přístupů z jiných oborů (např. předmět evaluace). Významným zdrojem prezentovaných typů jsou rovněž praktické zkušenosti nabyté v rámci dlouhodobého projektu zaměřeného na precizaci metaevaluačních postupů. Tyto poznatky vycházející z praxe umožnily posoudit relevanci jednotlivých dichotomií, zpřesnit jejich význam a formulovat základní metodologická doporučení. Lze tedy rozlišovat typy MEV podle následujících šesti kritérií:

a) dle iniciátora

V první řadě stojí za pozornost, kdo provedení MEV iniciuje. Provedení MEV může být iniciováno samotnými realizátory evaluace nebo může být požadováno další zainteresovanou stranou – typicky zadavatelem či některým z uživatelů. Lze se setkat i se situací, kdy iniciátorem MEV je třetí strana, např. kontrolní orgán, arbitr, koordinační pracoviště, akademická instituce nebo potenciální zadavatel evaluací, který se provedením MEV snaží poučit z již provedených evaluací.

V tomto ohledu je třeba připomenout, že iniciátor MEV je významnou zainteresovanou stranou, která svým záměrem deklaruje svůj zájem ve vztahu k primární evaluaci (tedy k té, jež je předmětem MEV). To má pochopitelně vliv na metodiku MEV, na její celkový design, na volbu kritérií, jež by měla být použita pro hodnocení primární evaluace a na způsob nakládání s výsledky MEV.

V případě, že iniciátorem je zadavatel, je obvyklou snahou vytěžit ze způsobu provádění evaluací nové poznatky využitelné v procesu zadávání nových evaluací. V případě, že iniciátorem MEV je subjekt, který je častým zadavatelem (a uživatelem) evaluací, pak se zájem obrací k identifikaci charakteristik, jež jsou realizovaným evaluacím společné, a které mohou být nezamýšleným důsledkem způsobu zadávání evaluací či stylem řízení a kontroly jednotlivých evaluací. V některých případech (nemusí to však být pravidlem) je MEV ze strany zadavatele iniciována jako nástroj získávání informací použitelných při následném řízení kvality prováděných evaluací (Benson et al. 2001). Pokud je iniciátorem MEV samotný realizátor evaluace, je obvyklým důvodem snaha identifikovat okolnosti daného evaluačního projektu, které jsou odlišné od ostatních evaluací (např. významně delší čas nutný k realizaci hloubkových rozhovorů, odchylky v návratnosti v prováděném dotazníkovém šetření, počet a rozsah připomínek zadavatele k první verzi zprávy apod.).

Specifikovat u provedené MEV iniciátora je významné, neboť tato informace souvisí s rozsahem informací dostupných pro realizaci MEV (zadavelské MEV se mohou více zaměřit na okolnosti zadání evaluací a jejich využití, naproti tomu realizátorské MEV se mohou více soustředit na vlastní metodiku evaluace a způsob jejího provedení; MEV iniciované třetími stranami se obvykle zaměřují na posouzení širšího rámce, ve kterém jsou MEV prováděny, na hodnocení kontextu, popř. na soulad s obecnými standardy či očekáváním.

b) dle účelu

Další z významných dichotomií rozlišuje MEV sumativní a formativní. Formativní MEV lze spojovat se snahou o zlepšování a rozvíjení hodnocených evaluací. Výsledky takto pojatých MEV tedy představují průběžnou zpětnou vazbu, která je využívána pro korigování běžících evaluací a pro zvýšení jejich kvality. Formativní MEV umožňují včas odhalit případné nedostatky, zjednat jejich nápravu a minimalizovat případné negativní dopady (Bingham, Felbinger 2002). Dále platí, že podrobná analýza nedostatků právě prováděných evaluací umožňuje hlubší a přesnější pochopení fungování používaných postupů a metod přímo v reálných podmínkách. Formativní MEV tedy mají obvykle (neplatí to však vždy) interní povahu, kdy evaluátor hodnotí svou vlastní evaluaci v průběhu jejího zpracovávání. V tomto případě MEV umožňuje lepší pochopení, jak různá

rozhodnutí v rámci daného evaluačního designu ovlivňují kvalitu výsledků a jak tato rozhodnutí dopadají na daný evaluační projekt. To dává evaluátorům příležitost ke zlepšení vlastních činností, a to ještě předtím, než jsou výsledky evaluace prezentovány zadavateli.

Sumativní MEV jsou používány v situacích, kdy je třeba posoudit dosažené výsledky a způsob provedení evaluace. Cílem evaluací není v tomto případě hledat návrhy na zlepšení, nýbrž konstatovat a osvědčit dosažení či nedosažení vytyčených cílů (Daponte 2008). Sumativní MEV jsou prováděny buď po skončení dané evaluace, nebo po přesně definované etapě (např. po předání vstupní zprávy, po dokončení terénních prací, po předání draftu evaluační zprávy apod.). Vzhledem k tomu, že takovéto MEV bývají často prováděny po dokončení a akceptaci primární evaluace zadavatelem, případné návrhy na její zlepšení stejně nemohly být zohledněny. Z podobného důvodu si sumativní MEV nekladou za cíl hledat v procesu realizace primární evaluace příležitosti a možnosti, jak dosáhnout zlepšení.

Sumativní MEV je prováděna třetí (nezávislou) stranou za účelem posílení legitimacy výsledků MEV a z důvodu snazší identifikace zdrojů chyb vyskytnuvších se v primární evaluaci. Tento přístup má pozitivní efekty pro budoucí evaluační projekty a je faktickým spojením mezi samostatnými evaluacemi. Iniciátorem sumativní evaluace může být kterýkoliv subjekt zmíněný v předchozí části (tedy zadavatel, realizátor i případná třetí strana), typickým iniciátorem je však zadavatel evaluace.

c) dle předmětu MEV

Předmětem MEV mohou být na jedné straně výstupy provedené evaluace (tj. typicky evaluační zprávy) nebo, na straně druhé, celý proces primární evaluace, a to včetně jednotlivých procesů vykonávaných během její přípravy, v rámci identifikace a analýzy zainteresovaných stran účastnících se zadání primární evaluace, při výběru jejího realizátora atd. Rozhodnutí o předmětu MEV je zásadní, neboť má vliv na volbu metod provedení MEV. V případě, že cílem MEV je posuzování finálních výstupů provedené evaluace, pak MEV bude mít velmi pravděpodobně podobu desk research, kdy metaevaluátor provede rozbor obsahu evaluační zprávy s využitím metod a technik obsahové, popř. výpovědní analýzy.

Kromě samotných závěrečných výstupů však mohou být předmětem MEV také další dokumenty ovlivňující způsob provedení primární evaluace a její charakter (např. zadávací dokumentace, smlouva o realizaci evaluace apod.). S použitými metodami souvisí také volba a definice jednotlivých kritérií, podle kterých jsou při provádění MEV primární evaluace hodnoceny.

V situaci, kdy předmětem MEV je posouzení odborné úrovně primární evaluace či její metodologické správnosti, pak MEV může být obohacena o prvky peer-review, kdy nezávislí odborníci a specialisté poskytnou vlastní odborná stanoviska ke konkrétním charakteristikám primární evaluace (Combs, Falletta 2000). Takovýmto způsobem lze posuzovat mj. vhodnost a korektnost identifikace zainteresovaných stran, volbu adekvátních evaluačních metod, zvolený způsob analýzy informací či transparentnost zpracování údajů a dovození evaluačních závěrů.

Pokud je cílem MEV komplexní vyhodnocení způsobu provedení primární evaluace, je velmi pravděpodobné, že při realizaci MEV bude desk research doplněn terénním výzkumem, kdy metaevaluátor provede např. individuální hloubkové rozhovory se zadavateli primární evaluace, se zástupci evaluovaných cílových skupin, popř. s vybranými členy evaluačního týmu.

d) dle realizátora MEV

Toto hledisko je třeba rozlišit od iniciace MEV, neboť se v aplikační praxi lze setkat s tím, že iniciátorem a realizátorem MEV jsou rozdílné subjekty. Velmi často zadavatel primární evaluace iniciuje provedení MEV, nicméně pověří touto činností třetí stranu. Častá je situace, kdy iniciátorem MEV je realizátor primární evaluace, který svěří její provedení třetí straně. Vyskytují se i případy, kdy zadavatel požaduje provedení MEV po realizátorovi primární evaluace formou jakési auto-evaluace. Východiskem rozlišování konkrétních subjektů provádějících MEV je fakt, že některé organizace – zadavatelé evaluací mohou k provádění MEV využívat své vlastní pracovníky. Třetí strana je subjektem, který neparticipoval na primární evaluaci a nebyl ani angažován zadavatelem primární evaluace.

Důvodů, proč někteří zadavatelé primárních evaluací využívají k realizaci MEV vlastní pracovníky namísto toho, aby realizovaly jednotlivé metaevaluační projekty na smluvním základě s externími subjekty, je hned

několik. U mnohých zadavatelů mohou být MEV považovány za součást vlastního řízení, a tak vzniká potřeba po průběžné MEV jednotlivých primárních evaluací (Preskill, Russ-eft 2005). Některé organizace navíc zadávají relativně vysoký počet primárních evaluací, což vytváří předpoklady pro systematické provádění MEV a pro vyškolení vlastních pracovníků pro provádění takovéto činnosti. V těchto případech tato forma MEV umožňuje dané organizaci redukovat náklady na realizaci vysokého množství, relativně standardizovaných MEV. Hlavní výhody provádění MEV zadavateli primárních evaluací lze dovodit ze souhrnu Khakeeho (2003), který v podobném kontextu poukazuje především na přesnou a detailní znalost prostředí, cílů a záměrů – interní pracovník je zkrátka součástí dané organizační struktury a kultury, a tak případné riziko plynoucí z nepochopení cílů MEV a účelu jejího provedení je relativně nízké. Tato výhoda se dále může promítnout do volby adekvátnějších postupů použitých při MEV a do efektivnější komunikace získaných výsledků.

Naproti tomu výhodou MEV prováděných třetími stranami (tedy obvykle nezávislými evaluátory) je potenciálně vyšší profesionalita, případně specializace na některé konkrétní typy a metody MEV. Pokud je tedy požadován specifický či méně standardní typ MEV, je pravděpodobnější, že třetí strana (externí evaluátor) bude v tomto ohledu dosahovat vyšší excelence než vlastní pracovník zadavatelské organizace (Smith, Brandon 2008). Pochopitelným důvodem spolupráce se třetími stranami je realizace sumativních MEV; v tomto případě je účast interních pracovníků s ohledem na možný konflikt zájmů přinejmenším nevhodná a nežádoucí.

Specifickým případem provádění MEV jsou šetření prováděná samotnými realizátory, kdy jde *de facto* o vlastní hodnocení, neboť MEV provádí přímo realizátor primární evaluace. Ress-eft a Bober (2007) zmiňují základní charakteristiky, jež jsme v kontextu MEV dále doplnili a rozšířili. Jde především o tyto:

reflexivní praxe

MEV provedená samotným realizátorem je činností, která využívá na schopnosti jednotlivých subjektů provést reflexi sebe sama. Vlastní hodnocení vede dotčené subjekty k tomu, aby věnovaly pozornost konkrétním aspektům své činnosti.

průběžné zjišťování kvality

Subjekty provádějící vlastní hodnocení systematicky hledají poučení ve své činnosti a posuzují adekvátnost jednotlivých opatření vzhledem k cílům, kterých dosáhly. Smysl vlastního hodnocení z hlediska evaluační praxe je třeba hledat také ve stimulaci postupů a kroků, které umožní soustavně vyhledávat a vyhodnocovat zpětnou vazbu týkající se vykonávané evaluační činnosti; vlastní hodnocení tak posiluje odpovědnost za konkrétní kroky, které jsou na úrovni jednotlivých subjektů prováděny. Vlastní hodnocení lze využít v situacích, kdy je dosažení definovaných cílů či ověření jiných cílů dané evaluace snadno identifikovatelné (a potenciálně snadno přezkoumatelné) a kdy není třeba používat komplexní mapu sofistikovaných indikátorů. Vlastní hodnocení vychází z předpokladu, že realizátoři mají zájem na vlastním zlepšování.

aplikační omezení

Zásadním nedostatkem vlastního hodnocení může být jeho subjektivní charakter, kvůli kterému je tento přístup diskvalifikován z některých aplikačních situací; MEV prováděné samotnými realizátory tak nelze např. využít, pokud jsou sledovány sumativní cíle. Vlastní hodnocení má tedy spíše formativní charakter a jeho využitelnost je tak přímo závislá na důvěře mezi realizátorem MEV a primární evaluací na jedné straně a zadavatelem na straně druhé.

e) dle plánovitosti

MEV lze rovněž rozlišovat podle toho, zda se s provedením MEV počítalo už při zadávání primární evaluace (či dokonce před její přípravou) nebo zdali bylo takového rozhodnutí učiněno až později (např. poté, kdy je primární evaluace dokončena). V tomto ohledu jde tedy o to, zdali je MEV součástí projektu připravované primární evaluace či zdali je samostatným projektem realizovaným víceméně nezávisle na primární evaluaci.

V praxi významně častěji převažuje druhý přístup, kdy se rozhodnutí o provedení MEV uskutečňuje častokrát až v návaznosti na výsledky dokončené evaluace (Depoy, French 2003). Takováto praxe však vede k tomu, že zadávané MEV mají sumativní charakter. Častokrát jsou jakými „auditem“, od něhož se očekává, že identifikuje případné chyby a sehlání při realizaci primární evaluace a poskytne tak argument k odmítnutí

„nevhodných“ závěrů primární evaluace. Pokud k rozhodnutí o realizaci MEV dochází až po dokončení primární evaluace, může být obtížné stanovit adekvátní kritéria, a pokud se to nakonec podaří, může být obtížné (či dokonce nemožné) získat potřebné informace a údaje. Z těchto důvodů je vhodné plánovat provedení MEV už při přípravě primární evaluace.

f) dle participativnosti

Participativní MEV probíhá ve spolupráci s ostatními zainteresovanými stranami, zejména pak s realizátorem primární evaluace a s jejím zadavatelem. Participativní MEV činí z těchto zainteresovaných stran „kolegy-evaluátory“, kteří ovlivňují design prováděné MEV a jsou tak vtaženi do procesu rozhodování o postupu evaluační činnosti, sběru dat i analýzy a vyhodnocení výsledků. Participativní MEV vycházejí z představy plnohodnotného, přímého a aktivního zapojení zainteresovaných stran, což činí z metaevaluačního procesu husté předivo interaktivních činností. Realizátor při participativní MEV vystupuje jako partner jednotlivých zainteresovaných stran, jako konzultant či kouč, který stimuluje, rozvíjí a vybízí jednotlivé subjekty k aktivitě v rámci jednotlivých etap MEV.

Participativní MEV tak vedou nejen k samotné evaluaci, ale napomáhají jednotlivým zainteresovaným stranám rozumět principům MEV, pomáhají jim uvědomovat si souvislosti jednotlivých činností, a mohou tak zlepšit zapojení dotčených zainteresovaných stran do procesů, jež jsou předmětem MEV (Alkin 2004).

Participativní MEV jsou častokrát mylně spojovány s kvalitativními metodami, patrně proto, že se v jejich případě očekává vyšší míra pochopení, vzhledu a obecně větší prostor, jež má zkoumaný subjekt k dispozici ke svému vlastnímu vyjádření. Nicméně platí, že participativní MEV se k žádné konkrétní metodě nepřimykají, navíc někteří představitelé (Bradley 2007; Hall, Hall 2004; Weiss 1998) sami vyvracejí názor, že kvalitativní výzkumné techniky vedou k vyšší míře zapojení zainteresovaných stran. Poukazují přitom na fakt, že účastníci kvalitativních šetření nemohou ovlivnit jejich design (což je základem participace) stejně tak jako účastníci kvantitativních šetření.

Z řady příkladů konkrétní aplikační praxe se ukazuje, že participativní MEV nejenže zapojuje jednotlivé zainteresované strany do procesu hod-

nocení, ale dokonce může přispět i k vyšší míře využívání dosažených výsledků (Caracelli, Preskill 2000). V tomto ohledu se totiž ukazuje, že zapojení jednotlivých zainteresovaných stran posiluje důvěryhodnost získaných závěrů, což se pozitivně odráží ve vyšší míře jejich využití.

Z uvedeného přiblížení participativního přístupu se může zdát, že opačný, tedy neparticipativní, přístup je nevhodný. Nicméně je třeba upozornit na skutečnost, že oba přístupy jsou vyvážené a nelze žádný z nich automaticky preferovat. Výhodou neparticipativních přístupů je, že obvykle vedou k vyšší validitě získaných závěrů. Participativní MEV totiž v důsledku své primární orientace na podporu zapojení co zainteresovaných stran, zakládají své závěry na metodách a postupech, které sice mohou být akceptovány mezi realizátorem MEV a realizátory primární evaluace, ale už nemusí být přijatelné pro zadavatele MEV.

Aplikační potenciál participativních MEV je při praktické realizaci omezen zájmem (či spíše nezájmem) zainteresovaných stran zapojit se aktivně do procesu MEV. Nelze považovat za samozřejmé, že se realizátor bude chtít zapojit do MEV, kterou může vnímat jako zásah a narušení svých zájmů; není možné samozřejmě předpokládat, že realizátor primární evaluace bude mít zájem poskytovat dodatečné (častokrát interní či citlivé) údaje o způsobu provedení evaluace (Batini, Scannapieco 2006) či si osvojovat způsoby provádění vlastního hodnocení. Rovněž se ukazuje, že nedostatečná schopnost zainteresovaných stran spolu komunikovat neumožňuje dosáhnout uspokojivého výsledku při využití participativních MEV.

3. Prameny kritérií pro provádění MEV

Předchozí část se zaměřila na vybrané typy MEV, zatímco v této části je pozornost věnována především obsahu MEV a zejména pak zdrojům, ze kterých je možné čerpat kritéria hodnocení používaných při MEV. Kromě celého spektra potenciálně využitelných dokumentů zahrnujícího zadávací dokumentaci primárních evaluací, metodiky, manuály a příručky je zvláštní pozornost věnována oborovým normám coby samoregulačním nástrojům zvyšování kvality prováděných evaluací.

V posledním období je v podmínkách ČR patrný sílící zájem o zjišťování, měření a řízení kvality evaluací, a to nejen v oblasti evaluace strukturálních fondů, ale také v oblasti evaluace projektů zahraniční rozvojové spolupráce a některých dalších oblastech, které evaluacemi aktivně využívají. Tyto tendence je třeba chápat jako přirozenou reakci na „první generaci“ prováděných evaluací. Shromážděné evaluační zprávy a dokončené evaluační projekty představují již poměrně rozsáhlý soubor podkladů, který poskytuje vhodné informace využitelné pro budoucí nastavení ukazatelů kvality realizovaných evaluací. V této souvislosti je třeba uvést, že MEV nejsou primárně nástrojem řízení kvality, nicméně mohou v těchto snahách o zvyšování kvality pomoci.

Výchozím pramenem, od něhož se odvíjejí kritéria hodnocení primárních evaluací, je samotný cíl, který vede dotčené subjekty k realizaci MEV. Zvolená kritéria jsou tak odvozeny od cílů MEV a zohledňují konkrétní potřeby jejich zadavatelů (iniciátorů). Jiným využitelným pramenem, z něhož jsou kritéria odvozována, je zadávací dokumentace primární evaluace. Takovýmto způsobem jsou kritéria stanovována v případech, kdy je cílem MEV stanovit míru souladu provedené evaluace a jejího zadání; MEV se zaměřuje na posouzení, do jaké míry naplnila provedená evaluace původní zadání, zdali navržený design evaluace směřoval k zodpovězení jednotlivých evaluačních otázek, zdali byla použita adekvátní metodika apod.

Významným pramenem kritérií pro MEV jsou dále metodiky vydávané zadavateli, neboť tyto dokumenty vytyčují oblast použitelných metodik zpracování evaluace³ a stanovují okruh akceptovatelných postupů (Uusikyla, Virtanen 2000). V rámci MEV se pak sleduje, do jaké míry a jakým konkrétním způsobem byly uplatněny postupy popisované v metodice.

Dalším pramenem, z něhož lze čerpat kritéria pro provedení MEV, jsou manuály a příručky – tedy prameny stanovující doporučené postupy pro tvorbu jednotlivých výstupů a evaluačních zpráv. Tyto zdroje obvykle stanovují význam klíčových pojmů, obracejí pozornost k jednotlivým metodám využitelných v rámci provádění evaluací a slouží tak realizáto-

³ To je v evaluacích mimořádně významné, neboť evaluace jsou transdisciplinárním oborem, v němž se setkávají metody a techniky různých oborů (ekonomie, pedagogického výzkumu, sociologie, demografie a mnohých dalších.).

rům evaluací jako pramen informací o uplatnitelných metodách a postupech využitelných při provádění evaluací. Ze své podstaty sice nemají závazný charakter (na rozdíl od metodik), nicméně působí silou své odborné autority, neboť zadavatelé evaluací očekávají, že jednotlivé činnosti prováděné v rámci zadávaných evaluací budou korespondovat s odborným standardem, který je v těchto manuálech a příručkách popsán. Přestože manuály a příručky mají nižší závaznost než metodiky, nicméně i tento typ dokumentů může být využit při případě kritérií dané MEV.

Specifickým typem dokumentů, které mohou představovat obsahovou základnu, z níž MEV mohou čerpat a fakticky také čerpají kritéria pro hodnocení primárních evaluací, jsou samoregulační nástroje. V tomto případě jde o dokumenty, jimiž se realizátoři sami zavazují k respektování konkrétních pravidel (Shadish et al. 1995). Samoregulační nástroje typicky zahrnují etické kodexy a procesně zaměřené normy či soubory standardů. Společným jmenovatelem těchto dokumentů je, že jsou připraveny přímo samotnými realizátory evaluací, obvykle na půdě oborových asociací a sdružení. Častokrát mají charakter souboru pravidel, jež jsou považovány za důležité. Obvykle pokrývají jednotlivá významná témata, která souvisí jak s realizací evaluací, tak také s jejich zadáváním a utilizací.

V rámci evaluačního oboru jsou klíčovým dokumentem tohoto typu Standardy evaluace programů (Program Evaluation Standards), které vznikly na půdě Společného výboru pro standardy evaluací ve vzdělávání (Joint Committee on Standards for Educational Evaluation), nicméně velmi rychle se rozšířily mimo rámec vzdělávacích evaluací a byly přijaty mnohými evaluačními asociacemi a společnostmi po celém světě – kromě USA, Kanady a Austrálie také ve Švýcarsku a v Německu. Rovněž Česká evaluační společnost tyto standardy přijala, a to na sklonku roku 2013. V současné podobě je komplex jednotlivých standardů rozdělen do čtyř oblastí definujících požadavky na primární evaluace s ohledem k jejich užitečnosti, proveditelnosti, korektnosti a přesnosti (Sanders 1994). Jednotlivé oblasti dále zahrnují několik konkrétních charakteristik, které by při designu a provádění evaluací měly být zohledněny.

V rámci MEV mohou být pro každou charakteristiku (či pro každý jednotlivý standard) identifikována konkrétní kritéria (ukazatele), jež vystihují smysl a účel jednotlivých standardů. Pokud je např. cílem MEV ověřit, do jaké míry naplňuje primární evaluace standard A9, tj. ospravedlnitel-

nost závěrů, lze se v rámci MEV zaměřit na kritéria odrážející mj. skutečnost, zdali primární evaluace cituje zdroje informací, o něž se opírá, zdali jsou formulované závěry vztahovány pouze k příslušnému období a kontextu a vyhýbají se nepřipustným generalizacím, zdali jsou uváděny také další alternativní závěry, které lze z používaných faktů dovodit a další.

V souvislosti s odvozením kritérií a s jejich konstrukcí je třeba upozornit, že reliabilitu prováděné MEV posilují především kritéria faktografické povahy, zatímco kritéria postojového charakteru reliabilitu naopak snižují (Byrn 2002). Podobně také kritéria binárního charakteru (ano/ne) či kritéria s nízkým počtem kategorií (3–4) obecně přispívají k vyšší reliabilitě MEV, zatímco v případě kritérií s vyšším počtem záznamových kategorií je třeba počítat s nižší spolehlivostí danou nižšími očekávanými četnostmi připadajícími na jednotlivé kategorie.

Rovněž je při přípravě kritérií MEV třeba počítat se záznamem celé řady klasifikačních kritérií (počet stran evaluační zprávy, charakteristiky autora zprávy, rozpočet, doba provedení apod.) a dalších pomocných znaků, které při následné analýze usnadňují charakterizaci typu posuzované evaluace.

Specifickými kritérii mohou být kvalitativní charakteristiky získávané pomocí hloubkových rozhovorů. V rámci MEV tyto rozhovory bývají používány s cílem ověřit správnost kvantitativních kritérií, získat explikaci dosažených hodnot a zachytit kontext, v němž byly primární evaluace prováděny.

4. Některé metodologické souvislosti MEV

V rámci tohoto textu není možné věnovat pozornost všem relevantním nuancím metaevaluačního designu a prezentovat zde jednotlivé metody a techniky využívané při MEV. Proto se pozornost zaměřuje na tři vybrané okolnosti, jež významným způsobem formují charakter prováděných MEV. Mezi tyto okolnosti patří volba podkladů používaných při MEV; ta pochopitelně vychází z cílů MEV, nicméně s ohledem k jejich dostupnosti a nákladům spojených s jejich získáním může cíle MEV zpětně ovlivňovat (Isaac, Michael 1997). Druhou okolností je způsob výběru primárních evaluací určených k MEV, který je stejně důležitý jako volba kritérií MEV.

Třetí okolností je způsob provedení, resp. organizace vlastní metaevaluační činnosti. Vzhledem ke skutečnosti, že významná část MEV se zaměřuje na hodnocení evaluačních zpráv, je obsah této části přizpůsoben charakteristikám, které odrážejí vlastnosti těchto dokumentů.

4.1 Design MEV

Design MEV, a tím i její metodika, se odvíjí od cílů MEV. Platí, že MEV může být provedena mnohými způsoby a při její realizaci mohou být využity rozličné metody a techniky. Podobně také mohou být při MEV použity rozličné datové zdroje. Následující přehled poukazuje na některé podklady, se kterými se metaevaluátor setkává a na jejichž základě formuluje své závěry:

a) Evaluační zpráva

Naprostá většina MEV se opírá o evaluační zprávy, značná část MEV dokonce ani s jinými podklady nepracuje. V rámci takto koncipovaných MEV lze hodnotit obsah, strukturu evaluačních zpráv, splnění formálních náležitostí, které mají evaluační zprávy mít apod. Jde v tomto případě o nejužší vymezení MEV.

Evaluační zprávy jsou tedy používány jako zdroj informací o provedené evaluaci a na jejím základě je hodnocen samotný proces provedení primární evaluace. V tomto případě je třeba počítat s nižší validitou MEV závěrů, neboť v evaluační zprávě mohou mnohé údaje a informace chybět nebo mohou být popsány natolik vágně, že neumožňují rekonstruovat použitý evaluační design. Některé okolnosti prováděné primární evaluace však nelze pro účely MEV využít ani v případě velmi detailních evaluačních zpráv. Pokud např. není dostupná zadávací dokumentace a protokoly z kontrolních dnů, nelze rozhodnout, zdali je vysoký počet evaluačních otázek dán rozhodnutím zadavatele či iniciativou realizátora evaluace⁴. Metaevaluátor v takovém případě může jen porovnat počet evaluačních otázek s ostatními evaluacemi (v rámci benchmarkingu). Na druhé straně lze na evaluační zprávy pohlížet jako na výsledek procesu, při kterém zadavatelé specifikují své požadavky, připomínají dřívější verze závě-

⁴ Podobně také může být absence konceptuálních modelů důsledkem nízkého povědomí o tomto nástroji mezi realizátory evaluačních šetření, ale stejně tak může být důsledkem rozhodnutí zadavatele nezařadit takovýto výstup do závěrečné zprávy.

rečných zpráv a větší či menší měrou se spolupodílejí na jejím obsahu a zaměření. Zajímavou úlohou v tomto ohledu bývá určení, zdali vůbec v daném případě jde o evaluační zprávu či o dokument jiného druhu (výzkumnou zprávu, studii, analytický materiál, rešerši apod.), resp. jde o určení, zdali byla prováděná aktivita skutečnou evaluací.

b) další dokumenty

MEV může při využití evaluačních zpráv spolu s dalšími dokumenty vztahujícími se k prováděné evaluaci mnohem lépe zasadit danou evaluaci do konkrétního kontextu (Leeuw2002). Typicky lze v tomto ohledu využít zadávací dokumentaci, zápisy a protokoly z kontrolních dnů či protokoly o vypořádání připomínek. Takto pojatá MEV je sice komplexnější než MEV v předchozím případě, nicméně stále je založena zejména na metodách a technikách analýzy sekundárních dokumentů.

c) interview se zadavateli a realizátory

MEV se může v tomto případě zaměřovat i na ostatní okolnosti provedení primární evaluace jakými jsou mj. postup výběru realizátora, proces kontraktace, specifikace a konkretizace zadání, postupy a činnosti související s akceptací závěrečné zprávy, vzájemná komunikace zadavatele a realizátora atd.). MEV v tomto ohledu předpokládá získávání specifických primárních údajů (např. záznamy interview) a jejich vyhodnocení spolu s dostupnými dokumenty (Posavac, Carey 2003).

d) interview se zainteresovanými stranami a návštěvy in-situ

Nejrozsáhlejší forma MEV může zahrnovat dokonce i re-evaluaci, tedy zopakování provedených analýz, opakované návštěvy u respondentů dotazovaných v rámci primární evaluace či ověření prováděných výpočtů. MEV v takovém případě může přinést velmi podrobné poznatky o okolnostech realizace primární evaluace a může poukázat na nízkou spolehlivost některých evaluačních závěrů⁵.

⁵ V této souvislosti je třeba upozornit, že nízká spolehlivost nemusí být chybou realizátora primární evaluace, ale může být přirozeným důsledkem proměnlivosti prostředí, ve kterém byla primární evaluace prováděna.

4.2 Výběr primárních evaluací pro MEV

V tomto případě je třeba za základní soubor považovat všechny provedené evaluace v rámci daného období (např. programovací období). S ohledem k jejich počtu je možné provést MEV:

a) jedné primární evaluace; MEV v takovém případě mívají charakter případových studií

b) souboru vybraných primárních evaluací; MEV v daném případě má charakter buď komparativní studie, která se zaměří na srovnání vybraných podsouborů primárních evaluací (např. evaluační zprávy připravené v rámci jednotlivých operačních programů) nebo je přehledovou studií, která se zaměřuje na kvantifikaci vybraných charakteristik odrážejících způsob provádění primárních evaluací

c) všech provedených primárních evaluací; MEV má podobně jako v přechodném případě, charakter přehledové studie. V tomto případě však není problematika výběru primárních evaluací relevantní.

V rámci provádění MEV je možné po zvážení konkrétních výhod a nevýhod zvolit některou z následujících výběrových technik:

Prostý náhodný výběr primárních evaluací

Výhodou prostého náhodného výběru je jednoduchost jeho provedení a možnost výpočtu standardní chyby výběru. Na druhé straně platí, že prostý náhodný výběr může vést k nesystematickým zkreslením a vychýlením, které nelze předem odhadnout. Dále platí, že v případě malého počtu primárních evaluací se riziko výběrového vychýlení dále zvyšuje.

Stratifikovaný náhodný výběr primárních evaluací

Výhodou stratifikovaného náhodného výběru je, že tento typ výběru lépe zohledňuje strukturu základního souboru (tedy souboru primárních evaluací) a snižuje riziko výběrového vychýlení. Stratifikovaný náhodný výběr může být nevhodný kvůli nedostatečným informacím obsažených ve výběrové opoře, která nemusí vždy obsahovat proměnné, které by bylo možné využít pro smysluplné nastavení strat.

Výběr primárních evaluací úsudkem

Tento typ výběru lze aplikovat jen v případě nízkého počtu primárních evaluací. Platí, že tento typ výběru je obtížně obhajitelný, neboť k výběru jednotek dochází na základě subjektivního posouzení. Lze ho akceptovat, pokud je MEV koncipována jako případová studie, nicméně i v tomto případě je nutná srozumitelná argumentace, proč (tj. na základě jakých znaků) byla daná primární evaluace vybrána.

Transparentní a metodologicky korektní výběr primárních evaluací, jež jsou určeny pro MEV je významnou okolností, která může ovlivnit kvalitu samotné MEV, a to zejména v případě přehledové či komparativně koncipovaných MEV, kdy použití nevhodné techniky výběru může způsobit posunutí (bias) metaevaluačních závěrů.

4.3 Postup hodnocení

Pokud jde o vlastní hodnocení jednotlivých parametrů primární evaluace, bývá používán jeden ze dvou typických postupů. Buď je hodnocení daného případu provedeno výlučně jedním metaevaluátorem nebo v rámci dané MEV provádí hodnocení každého jednotlivého kritéria větší počet hodnotitelů (Wholey et al. 2004). V tomto případě jsou jednotlivá kritéria zaznamenávána obvykle dvěma až pěti na sobě nezávislými hodnotiteli, kdy každý z nich postupuje samostatně podle stanovených instrukcí.

Výhodou prvního přístupu založeného na jednotlivém posouzení je vyšší konzistence hodnocení, neboť veškeré hodnocení provádí jedna osoba, a vyšší rychlost provedení MEV, jež je daná tím, že se provádí jen jedno hodnocení. Na druhé straně je třeba počítat s tím, že celková doba hodnocení je omezena individuální kapacitou daného hodnotitele (Splauding 2008). Zásadními nevýhodami jsou nižší reliabilita a vysoká míra subjektivity. Vzhledem k tomu nelze tento přístup použít, pokud je cílem provést sumativní MEV.

Výhodou druhého přístupu je vyšší reliabilita, objektivizace závěrů a vyšší přesnost hodnocení. Nevýhodou je především zdlouhavé zpracování, organizační náročnost a vyšší náklady spojené s realizací hodnocení. Specifickou okolností, která komplikuje přípravu takto koncipované MEV, je

jednak nutná kalibrace hodnotitelů zabezpečující jednotný způsob záznamu sledovaných kritérií⁶, a dále pak konsolidace většího počtu hodnocení.

V případě, že je pořizován vícečetný záznam sledovaných kritérií, je v další fázi nutné ověřit shodu individuálních hodnocení jednotlivých metaevaluatorů a v případě neshody postupovat podle předem stanovených kroků směřujících ke konvergenci individuálních hodnocení. V tomto ohledu jsou stávající hodnotitelé buď požádáni o dosažení konsenzu, nebo je daný případ hodnocen ještě dalším nezávislým hodnotitelem. V některých případech lze také přistoupit k opakovanému hodnocení stávajícími hodnotiteli (analogue test – re-test reliability). Tyto scénáře se mohou lišit v závislosti na typu kritéria; pokud jde o faktografické údaje (jako např. počet stran, rok přípravy zprávy apod.) lze danou informaci ověřit relativně snadno a také platí, že v tomto případě je míra shody mezi nezávislými hodnotiteli velmi vysoké (běžně přesahuje 90 procent).

Jiná je situace v případě měkkých, bodovaných údajů. Pokud v tomto případě dosahuje míra shody mezi hodnotiteli 75 – 80 procent, jde sice o vysokou míru shody, nicméně i tak je nutné vzniklé neshody v hodnocení uspokojivým způsobem vyřešit. Pokud míra shody mezi metaevaluátory činí méně než 50 procent, je třeba hledat zdroj takto vysoké neshody. Může jím být např. odlišný způsob hodnocení některého z hodnotitelů, vágně definované kritérium hodnocení či nepřesné pravidlo záznamu daného kritéria.

⁶ Jde např. o jednotné sledování počtu stran; záznam intenzity výskytu daného jevu či míry naplnění některého ze sledovaných kritérií.

Zdroje

- [1] ALKIN M. C., *Evaluation Roots*. London: SAGE Publication, 2004. ISBN: 9780761928942
- [2] BATINI C., SCANNAPIECO M., *Data Quality. Concepts, Methodologies and Techniques*. Berlin: Springer-Verlag. 2006. ISBN: 9783540331728
- [3] BENSON A. P., HINN D. M., LLOYD C. (eds.), *Visions of Quality: How Evaluators define, understand and represent program quality*. London: JAI. 2001. ISBN: 9780762307715
- [4] BINGHAM R. D., FELBINGER C. L., *Evaluation in Practice. A Methodological Approach*. New York: Seven Bridges Press. 2002. ISBN: 9781889119571
- [5] BRADLEY C. J. (ed.), *Process Use in Theory, Research, and Practice*. San Francisco: Jossey-Bass. 2007. ISBN: 9780470255261
- [6] BYRN D., *Interpreting Quantitative Data*. London: SAGE Publication. 2002. ISBN: 9780761962625
- [7] CARACELLI V. J., PRESKILL H. (eds.), *The Expanding Scope of Evaluation Use*. San Francisco: Jossey-Bass. 2000. ISBN: 9780787954338
- [8] CLARKE A., *Evaluation Research. An Introduction to Principles, Methods and Practice*. London: SAGE Publications. 1999. ISBN: 9780761950950
- [9] COMBS W. L., FALLETTA S. V., *The Targeted Evaluation Process*. Alexandria: ASTD. 2000. ISBN: 9781562861407
- [10] DAPONTE B. O., *Evaluation Essentials*. San Francisco: Jossey-Bass. 2008. ISBN: 9780787984397
- [11] DEPOY E., FRENCH G. S., *Evaluation Practice*. Toronto: Thomson. 2003. ISBN: 9780534543914
- [12] FITZPATRICK J. L., MORRIS M. (eds.), *Current and Emerging Ethical Challenges in Evaluation*. San Francisco: Jossey-Bass. 1999. ISBN: 9780787949020
- [13] GUBA E. G., LINCOLN Y. S., *Fourth Generation Evaluation*. London: SAGE Publishing. 1989. ISBN: 9780803932357
- [14] HALL D., HALL I., *Evaluation and social research – introducing small-scale practice*. New York: Palgrave Macmillan. 2004. ISBN: 9780333930953
- [15] HONG D. H., BODEN M., *R&D Programme Evaluation – Theory and Practice*. Hants: Ashgate. 2003. ISBN: 9780754632078
- [16] HOUSE E. R., *Evaluating with Validity*. London: SAGE Publishing. 1980. ISBN: 9780803914391
- [17] HOUSE E. R., HOWE K. R., *Values in Evaluation and Social Research*. London: SAGE Publishing. 1999. ISBN: 9780761911555

- [18] ISAAC S., MICHAEL W. B., *Handbook in Research and Evaluation*. California: Edits. 1997. ISBN: 9780912736327
- [19] KHAKEE A., *The Emerging Gap between Evaluation Research and Practice*. Evaluation 2003, 9, 340-352. 2003.
- [20] LEEUW F.L., *Evaluation in Europe 2000: Challenges to a Growth Industry*. Evaluation 2002, 8, 5-12. 2002.
- [21] OWEN J. M., *Program Evaluation. Forms and Approaches*. New York: The Guilford Press. 2006. ISBN: 9781593854065
- [22] PATTON M. Q., *Utilization-Focused Evaluation*. London: SAGE Publishing. 2008: ISBN: 9781412958615
- [23] POLLITT Ch., *Evaluation in Europe: Boom or Bubble?* Evaluation 1998, 4, 214-224. 1998.
- [24] POSAVAC E. J., CAREY R. G., *Program Evaluation. Methods and Case Studies*. New Jersey: Pearson Education Inc. 2003. ISBN: 9780132275606
- [25] PRESKILL H., RUSS-EFT D., *Building Evaluation Capacity*. London: SAGE Publishing. 2005. ISBN: 9780761928102
- [26] ROSSI P. H., LIPSEY M. W., FREEMAN H. E., *Evaluation. A Systematic Approach*. London: SAGE Publications. 2004. ISBN: 9780761908944
- [27] RUSS-EFT D., BOBER M. J., de la Teja I., FOXON M. J., KOSZALKA T. A., *Evaluators Competencies*. San Francisco: Jossey-Bass Publishers. 2007. ISBN: 9780787995997
- [28] SANDERS R. J., *The Program Evaluation Standards 2*. London: SAGE Publishing. 1994. ISBN: 9780803957329
- [29] SCHWANDT T. A., HALPERN E. S., *Linking Auditing and Metaevaluation*. London: SAGE Publishing. 1988. ISBN: 9780803929685
- [30] SCRIVEN M., *The Methodology of Evaluation*. In Worthen B. R., Sanders J. R. (eds.) *Educational Evaluation: Theory and Practice*. Belmont: Wadsworth. 1973. ISBN: 9780839600558
- [31] SCRIVEN M., *Evaluation Thesaurus*. London: SAGE Publishing. 1991. ISBN: 9780803943643
- [32] SHADISH W. R., NEWMAN D. L., SCHEIRER M. A., WYE Ch., (eds.), *Guiding Principles for Evaluators*. San Francisco: Jossey-Bass. 1995. ISBN: 9780787999247
- [33] SMITH N. L., BRANDON P. R., *Fundamental Issues in Evaluation*. New York: The Guilford Press. 2008. ISBN: 9781593853426
- [34] SPAULDING D. T., *Program Evaluation in Practice*. San Francisco: Jossey-Bass. 2008. ISBN: 9780787986858
- [35] STUFFLEBEAM D. L. (ed.), *Evaluation Models*. San Francisco: Jossey-Bass. 2001. ISBN: 9780787957551

- [36] STUFFLEBEAM D. L., SHINKFIELD A. J., *Evaluation, Theory, Models & Applications*. San Francisco: Jossey-Bass. 2007. ISBN: 9780787977658
- [37] UUSIKYLÄ P., VIRTANEN P., *Meta-Evaluation as a Tool for Learning: A Case Study of the European Structural Fund Evaluations in Finland*. Evaluation 2000, 6, 50-65. 2000.
- [38] WEISS C. H., *Evaluation*. New Jersey, Prentice Hall. 1998. ISBN: 9780133097252
- [39] WHOLEY J. S., HATRY H. P., NEWCOMER KE. (eds.), *Handbook of Practical Program Evaluation*. San Francisco: Jossey-Bass. 2004. ISBN: 9780787967130