

Studentské hodnocení výuky – zkušenosti z VVŠ PV ve Vyškově

Eva Šmelová, Oldřich Kříž

Anotace:

Příspěvek shrnuje zkušenosti z oblasti hodnocení výuky studenty realizované na Fakultě ekonomiky obrany státu a logistiky Vysoké vojenské školy pozemního vojska ve Vyškově. Blíže se zabývá také statistickými metodami, které byly použity pro vyhodnocení získaných výsledků.

Obecná východiska hodnocení výuky studenty a příprava experimentu

Specifickou oblastí evaluace výuky je *hodnocení kvality vysokoškolské výuky studenty*. Je možné zde využívat anonymních dotazníků s položkami škály (rating) hodnotícími jednotlivé kurzy, jednotlivé vyučující apod.

Názory na tento způsob hodnocení výuky jsou rozdílné. V 80. letech se tento typ evaluace prosazoval např. na některých vysokých školách v USA, později se od něj ustupovalo, a to zejména z důvodů potíží vznikajících 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“.

Například Shaw se ve své studii (4) zabývá identifikací charakteristických rysů tříd efektivních vysokoškolských a univerzitních učitelů. Poukazuje na skutečnost, že se zatím vyučující a výzkumní pracovníci neshodli na jasných odpovědích na otázky *Co je efektivní vyučování?* a *Jak lze efektivní vyučování měřit?*. Studentskému hodnocení byla věnována velká pozornost. U nás se touto problematikou zabývá např. Mareš, Chráska. Platnost studentského hodnocení byla pozorně analyzována a obecně podpořena v literatuře koncem 20. století (Centra, 1994; Cohen, 1981, 1987; Feldman, 1989; Marsh, 1987; Marsh-Bailey, 1993). Bylo zjištěno, že hodnocení výuky studenty velmi souvisí s charakteristickými rysy osobnosti učitelů (Feldman, 1986; Murray, Rushton-Paunonen, 1990; Renaud-Murray, 1996). Pokusy dr. Foxe v 70. letech (March, 1987; Naftalin, Ware-Murray, 1996) demonstrovaly, že studenti hodnotili charismatické a výrazné učitele jako velmi efektivní, a to bez ohledu na podstatu obsahu přednášky. Murray a kolektiv (1990) srovnal klasifikační hodnocení charakterových vlastností osobnosti se studentským hodnocením a zjistil, že charakterové vlastnosti se lišily

s různými typy kurzů. Renauld a Muray (1996) také zkoumali vztah mezi osobností učitele a studentským hodnocením, aniž by dělali rozdíly v typech kurzů. Výsledné vztahy byly silnější, než ty, ke kterým dříve dospěl ve své studii Murray a kolektiv, ale to mohlo být způsobeno omezenými rozsahy v dřívějších studiích.

Jak dále uvádí Shaw, bylo zjištěno, že výsledky studentů a hodnocení studentů spolu souvisí. Byl zkoumán vztah mezi studentským hodnocením a výsledky závěrečných zkoušek. Autor článku ukazuje na zjištění, že pokud spolu souvisí studentské hodnocení a výsledky zkoušek, vyšší hodnocení studentů může způsobit:

- efektivnější vyučování, které podněcuje studenty, aby se více učili a současně tak byli lépe hodnoceni,
- vyšší spokojenost studentů s vyššími známkami, které studenty podněcují k tomu, aby to učitelům „vrátili“ v podobě vyššího hodnocení bez ohledu na efektivitu vyučování nebo na to, jak moc se studenti učí.

March a Roche v roce 1997 (4) shrnuli výzkum shovívavosti při klasifikování vyjádřením, že podjatost existuje, ale její vlivy jsou nedůležité. Greenwald a Gillmore v roce 1997 rozebrali vztah mezi studentským hodnocením a klasifikováním učitelů a došli k výsledku, že silný vztah mezi nimi je náznakem přirozeného vztahu, shovívavost při klasifikaci vede k tomu, že studenti hodnotí učitele pozitivněji.

Celkově výzkum efektivity učitelů indikuje přiměřeně pozitivní vztah mezi hodnocením výuky studenty a výsledky studentů. Podle McKeachie jsou studenti ve třídě téměř každý den a ví, o co jde. Jsou to oni, koho se snaží pedagogové ovlivnit, a sami vnímají, zda se něčemu učí. V roce 1997 opět McKeachie byl důsledný ve svých názorech, když tvrdil, že studenti budou vždy nejvíce ovlivněni vyučováním, proto jejich hodnocení výuky budou vždy využitelná.

Z uvedených skutečností je zřejmé, že studentské hodnocení je vždy značně subjektivní, ale pokud budeme s uvedenými možnými vlivy počítat, potom může být pro nás i zdrojem kvalitní zpětné vazby.

Po prostudování našich i zahraničních zkušeností z oblasti studentského posuzování výuky jsme připravili a následně realizovali posuzování výuky studenty na FEOSL VVŠ PV ve Vyškově. Jako výzkumná metoda bylo zvoleno škálování, upravená verze posuzovací škály SEEQ Marešem a Chráskou (2). Do hodnocení se na základě dobrovolnosti mohli zapojit všichni studenti prezenčního studia na zmíněné fakultě, které hodnocený učitel vyučoval v předcházejícím semestru. Bylo třeba provést následující kroky:

- stanovit období, za které mají být učitelé hodnoceni,

- určit učitele, kteří mají být hodnoceni (rozhodnutí v tomto kroku provedl děkan fakulty),
- seznámit učitele s připravovanou akcí (na kolegiu děkana byli informováni vedoucí kateder o připravované akci, jejich prostřednictvím byli dále informováni vyučující),
- oslovit studenty a seznámit je s připravovanou akcí (prostřednictvím akademického senátu a informační tabule v prostorách fakulty).

Po uskutečnění uvedených kroků byl realizován sběr dat. Studenti byli seznámeni blíže s cílem šetření a požádáni o spolupráci. Zde byla zejména zdůrazněna anonymita šetření a zapojení do hodnocení na základě dobrovolnosti (objevovala se často ze strany studentů nedůvěra v zachování anonymity). Hodnocení bylo ověřováno na vzorku:

- **5 učitelů,**
- které posuzovalo **158 studentů.**

Respektovali jsme, aby každého učitele hodnotilo alespoň 80 % studentů z každé studijní skupiny. Vycházíme zde z E. A. McBeana a W. C. Leanoxe (4), kteří uvádějí orientační údaje v procentech a poukazují na nezbytnost brát v úvahu velikost skupiny studentů. V případě, že se jedná o skupinu menší než 30 studentů, měli bychom shromáždit názory alespoň od 80 % osob. Jde-li o skupinu 30 osob a více, je třeba získat data alespoň od 50 % osob.

Metoda výzkumného šetření a statistické metody pro ověřování platnosti hypotéz

Zkušenosti ukazují, že nejvhodnějším prostředkem studentského hodnocení kvality vysokoškolské výuky jsou posuzovací škály (v odborné literatuře označované jako rating, škály Likertova typu, bodové škály apod.). Likertovy škály se používají na měření postojů a názorů lidí. Skládají se z výroků a stupnice. Na stupnici člověk vyjadřuje stupeň svého souhlasu resp. nesouhlasu s výrokem (1).

Posuzovací škála je nástroj, který umožňuje zjišťovat míru vlastností jevu nebo jeho intenzitu. Posuzovatel vyjadřuje stupeň svého souhlasu resp. nesouhlasu s výrokem prostřednictvím určení polohy na škále. Posuzovací škály mají většinou 3, 5, 7, případně 9 stupňů. Počet stupňů ovlivňuje jemnost posouzení. Při třech stupních je jev posouzen hruběji, s růstem počtu stupňů narůstá jemnost posouzení. V praxi se často upřednostňují liché počty stupňů před sudými, protože lze lépe "nastavit" prostřední úroveň.

Při výzkumu „tohoto“ typu využíváme **ordinální stupnice**. Měřit na ordinální stupnici znamená jednotlivým variantám odpovědi přiřadit čísla vyjadřující „větší“ nebo „menší“ úroveň sledovaného znaku. Údaj o tom, jakou úroveň znaku nese daný objekt, vypovídá totiž o jeho pořadí v rámci uspořádaných prvků základního souboru nebo výběru. Více však z těchto čísel vyvozovat nelze. V množině uspořádané podle ordinální stupnice nelze stanovit, jak vel-

ká je „vzdálenost“ mezi dvěma sousedními objekty. Příkladem ordinální stupnice je např. klasifikační stupnice, nebo následující posloupnost výroků vyjadřující názor respondenta na položenou otázku, např. přednáška měla jasnou a srozumitelnou strukturu:

- 1) naprosto nesouhlasím
- 2) nesouhlasím
- 3) nemám vyhraněný názor
- 4) souhlasím
- 5) naprosto souhlasím

U takto konstruovaných stupnic je možné rozhodnout, že se dvě úrovně nerovnájí, a navíc můžeme prohlásit, že např. obměna 4) představuje vyšší úroveň souhlasu než obměna 2). Problém se vzdálenostmi dokládají i krajní stupně, které obvykle zahrnují celý zbývající rozsah možných poloh. Odtud vyplývá potřeba věnovat pozornost výběru metod určených pro zpracování dat ordinálního typu.

Pro vyjádření výsledků jsme použili způsob, kdy je škála chápána jako kontinuum. Jednotlivým obměnám na škále jsme přisoudili koeficienty a pro určení střední úrovně byl použit medián (5). Postup byl následující:

- jednotlivým stupňům na škále byly přiřazeny číselné hodnoty od 1 (nejméně příznivý postoj) po 5 (nejvíce příznivý postoj),
- potom jsme zjistili, kolik respondentů volilo jednotlivé odpovědi (frekvence voleb),
- střední úroveň sledovaného znaku byla vyjádřena prostřednictvím *mediánu*.

Námi použitý posuzovací nástroj obsahuje celkem 24 škálových položek (otázek), které lze rozdělit do osmi částí. Každá část má společné obsahové vlastnosti, které odborně nazýváme *dimenze*. Pro každou dimenzi existuje několik škálových položek. V našem případě se jedná o šest dimenzí, které obsahují 3-4 škálové položky (položky č. 1 - č. 22). Položky č. 23 a 24 jsou souhrnným zhodnocením výuky a učitele. Použité byly u všech škálových položek 5stupňové posuzovací škály. Jednotlivé škálové položky sledující hlavní zvolené dimenze, jsou strukturované takto:

- | | |
|-------------------------------|----------------------|
| 1. studentovo učení | (položky č. 1 – 3) |
| 2. učitelovo nadšení | (položky č. 4 – 7) |
| 3. organizace výuky | (položky č. 8 – 11) |
| 4. interakce učitel - student | (položky č. 12 – 15) |
| 5. individualizace výuky | (položky č. 16 – 18) |
| 6. učitelův rozhled | (položky č. 19 – 22) |

7. souhrnné hodnocení výuky (položka č. 23)

8. souhrnné hodnocení učitele (položka č. 24)

Pro vyhodnocení získaných dat použijeme metody odpovídající typu zpracovávaných dat. Předpokládejme nejprve, že 1 sledovaný znak u 1 učitele je posuzovaný prostřednictvím ordinální stupnice, která má K kategorií (obměn, stupňů). V průběhu výpočtů budeme využívat absolutní četnosti (zastoupení) k -té kategorie n_k , $k = 1, 2, \dots, K$.

Předpokládejme, že při hodnocení jednoho znaku chceme ověřit, zda podíl respondentů hovořících ve prospěch jedné a druhé kategorie se významně liší. Úkolem je tedy v rámci jedné otázky zjistit, zda rozdíl mezi absolutními četnostmi n_i a n_j těchto dvou kategorií je statisticky významný. Pokud je podíl respondentů hovořících ve prospěch i -té a j -té kategorie shodný, lze očekávat, že $n_i / (n_i + n_j) \approx 1/2$. Čím větší bude rozdíl od $1/2$, tím spíše se přikloníme k tvrzení, že podíl respondentů ve prospěch jedné a druhé kategorie se významně liší. Jedná se tedy o binomický model s parametry $\pi = 1/2$ a $n = n_i + n_j$.

Označme π pravděpodobnost toho, že respondent zvolí i -tou kategorii. Situaci, kdy mezi kategoriemi rozdíl neexistuje, budeme formulovat jako nulovou hypotézu $H_0: \pi = 1/2$. Proti ní postavíme alternativní hypotézu $H_1: \pi \neq 1/2$. Pro $n_i \geq n_j$ hypotézu H_0 zamítneme, když

$$\frac{n_i}{n_j + 1} \geq F_{1-\alpha/2}(v_1, v_2), \text{ kde}$$

$F_{1-\alpha/2}(v_1, v_2)$ je kvantil Fisherova rozdělení a $v_1 = 2(n_j + 1)$ a $v_2 = 2n_i$ jsou stupně volnosti.

Pokud chceme prokázat, že podíl respondentů hovořících ve prospěch i -té kategorie významně překračuje podíl respondentů hovořících ve prospěch j -té kategorie, použijeme jednostranný test. Jako nulovou hypotézu vyjádříme opět tvrzení $H_0: \pi = 1/2$ proti alternativní hypotéze $H_1: \pi > 1/2$. Pro $n_i \geq n_j$ potom hypotézu H zamítneme, když

$$\frac{n_i}{n_j + 1} \geq F_{1-\alpha}(v_1, v_2), \text{ kde}$$

$F_{1-\alpha}(v_1, v_2)$ je kvantil Fisherova rozdělení s v_1 a v_2 stupňů volnosti.

Pokud chceme srovnávání rozšířit na L kategorií, použijeme test o shodě pravděpodobností L kategorií ($L \leq K$). Tento test je možné užít za předpokladu, že pro všechny kategoriální četnosti platí $n_1 + n_2 + \dots + n_L > 30$. Předpokládejme, že chceme ověřit, zda podíl respondentů hovořících ve prospěch jednotlivých kategorií se významně liší. Potom budeme formulovat hypotézu o shodě pravděpodobností jednotlivých kategorií proti alternativě o jejich neshodě, tj. $H_0: \pi_1 = \pi_2 = \dots = \pi_L$ proti $H_1: \pi_i \neq \pi_j$ pro libovolné $i \neq j$. Testové kritérium

$$\chi^2 = \frac{L \cdot \sum_{k=1}^L n_k^2}{\sum_{k=1}^L n_k} - \sum_{k=1}^L n_k$$

má při planosti hypotézy H_0 Pearsonovo rozdělení s $L-1$ stupni volnosti. Se $100(1-\alpha)\%$ spolehlivostí potom zamítneme H_0 , když bude platit

$$\chi^2 \geq \chi_{1-\alpha}^2(L-1), \text{ kde}$$

$\chi_{1-\alpha}^2(L-1)$ je kvantil Pearsonova rozdělení s $L-1$ stupni volnosti.

Vhodnou metodou pro srovnání sledovaného znaku u více jako dvou subjektů (učitelů) může být Kruskal-Wallisův test. Předpokládejme, že máme hodnotit ve sledované vlastnosti R subjektů ($R \geq 3$), a počet respondentů, kteří je hodnotili, je alespoň 5. Obecně se postupuje tak, že každému subjektu přiřadíme náhodný výběr reprezentovaný výsledky hodnotitelů a ze všech jednotek vytvoříme sdružený (spojený) výběr o rozsahu $n = n_1 + n_2 + \dots + n_R$ uspořádaný vzestupně podle velikosti. Jednotlivým hodnotám se přiřadí pořadová čísla. Pro každý subjekt pak vypočítáme součet pořadí jednotek příslušejících do r -tého výběru a označíme jej T_r , $r = 1, 2, \dots, R$.

Testovaná hypotéza předpokládá, že mezi R subjekty není rozdíl v úrovni sledovaného znaku. Alternativa testu potom bude předpokládat, že existuje alespoň jeden subjekt, který se ve sledované R -tici v úrovni znaku liší od jiného nebo jiných subjektů. Testové kritérium Kruskal-Wallisova testu

$$Q_{KW} = \frac{12}{n \cdot (n+1)} \sum_{r=1}^R \frac{T_r^2}{n_r} - 3 \cdot (n+1)$$

má při platnosti hypotézy přibližně rozdělení χ^2 s $(R-1)$ stupni volnosti. Pokud je shodných pozorování ve všech výběrech více než 25 % (v našich průzkumech se % shodných pozorování bude blížit 100 % !), potom se doporučuje užít korigované testové kritérium

$$Q_{KW}^* = \frac{Q_{KW}}{1 - \frac{D}{n^3 - n}}, \text{ kde } D = \sum_{k=1}^K (n_k^3 - n_k),$$

K je počet obměn, R je počet subjektů, n_k je počet pozorování v k -té kategorii ve spojeném souboru, n_r je počet hodnotitelů v r -tém výběru a T_r je součet pořadí jednotek r -tého výběru.

Hypotézu se $100(1-\alpha)\%$ spolehlivostí zamítneme, pokud bude platit

$$Q_{KW}^* \geq \chi_{1-\alpha}^2(R-1), \text{ kde}$$

$\chi_{1-\alpha}^2(R-1)$ je $100(1-\alpha)\%$ kvantil Pearsonova rozdělení s $(R-1)$ stupni volnosti.

Shrnutí

V rámci tohoto experimentu jsme využili ověřenou posuzovací škálu, upravenou verzi SEEQ Marešem a Chráskou (2). Získané výsledky můžeme shrnout do následujících bodů:

1. Relevantní důležitost hodnocených položek:

~ lze konstatovat, že studenti se vyjadřovali k oblastem, které významně ovlivňují průběh a výsledky výuky.

2. Reprezentativnost a obsahová validita hodnocení:

~ stěží lze hodnotit výuku do všech detailů, je nutné se zaměřit pouze na určitý výběr proměnných, situací, vzorků osob;

~ může vystoupit do popředí otázka, zda měří zvolená metoda skutečně to, o čem se domníváme, že je důležité;

~ vycházíme ze skutečnosti, že posuzovací škála byla ověřena odborníky zabývajícími se problematikou studentského posuzování výuky.

3. Praktičnost a celková ekonomičnost hodnocení:

~ hodnocení realizované formou vyplňování hodnotících záznamů se jeví z časového hlediska jako velmi náročné a neekonomické.

4. Vícepohledovost, užití více metod:

~ pokud má být hodnocení multikriteriální, má-li poskytnout více pohledů na tutéž proměnnou, potom je nezbytné použít více metod pro hodnocení, získáme tak možnost výsledky vzájemně prověřovat;

~ domníváme se, že použitou posuzovací škálu lze využít pro studentské posuzování výuky, ale v žádném případě nemůže být jedinou metodou, prostřednictvím níž by byl pedagog hodnocen;

~ na základě použití pouze posuzovací škály za účelem hodnocení učitele nelze zformulovat objektivní závěry;

~ pokud budeme kombinovat více metod a na základě různých postupů získáme přibližně stejné údaje, teprve potom můžeme formulovat závěry hodnocení.

5. Predikační validita hodnocení:

~ při hodnocení výuky studenty je třeba mít jasně stanovený cíl;

~ je nezbytné si položit otázku, zda *použitá metoda odpovídá stanovenému účelu*, zda *získáme údaje, které nesou požadované informace*;

- ~ můžeme konstatovat, že metoda škálování i samotná posuzovací škála splňuje stanovený cíl, taktéž použité statistické metody lze vhodně využít a získat základní informaci o tom, jak je učitel studenty ve sledovaných oblastech vnímán.

6. Reliabilnost hodnocení:

- ~ použité metody pro hodnocení kvality výuky musí vést ke spolehlivým závěrům;
- ~ zvolené metody musí být vnitřně konzistentní, dobře strukturované;
- ~ při použití škálování za účelem posouzení pozorovaných jevů je mírou reliability shoda více zacvičených pozorovatelů;
- ~ reliabilita výzkumného nástroje je současně dána tím, že tatáž vlastnost je posuzována více položkami, nástroj tak získává větší vnitřní konzistenci, což použitá škála splňuje.

7. Instrumentálnost hodnocení:

- ~ realizované hodnocení nesmí být nikdy samoučelné;
- ~ mělo by vést ke zkvalitnění výuky;
- ~ myslíme si, že uvedenou posuzovací škálu lze využít za účelem hodnocení výuky studenty, ale nemůže docházet k jejímu absolutizování;
- ~ užití kvantitativních metod při hodnocení kvality výuky (kde řadíme i škálování) nemůže nahradit další metody, jako je např. rozbor používaných studijních materiálů, hospitace ve výuce, diskuse s učiteli o výuce;
- ~ myslíme si, že studentské hodnocení výuky by bylo vhodnější brát jako určité ukazatele „spokojenosti studenta“.

Myslíme si, že by studentské hodnocení výuky mělo být bráno spíše jako zpětná vazba pro učitele. Pokud by měly získané výsledky sloužit jako podklady pro uskutečnění změn, je třeba je pečlivě analyzovat, zpracované údaje doplnit o možnost konzultací s nadřízeným pracovníkem či kolegy.

Co ještě nelze v této části opomenout, je finální etapa, tj. práce se získanými výsledky:

- předně zastáváme názor, že závěry o kvalitě výuky nelze vyvozovat pouze na základě studentského posuzování výuky,
- pokud škola studentské posuzování výuky realizuje, měla by mít také promyšleno, jak s údaji bude dále pracovat,
- učitel by neměl získat výsledky šetření pouze v podobě formalizovaného výstupu, např. počítačového výstupu, ale naopak by měl mít možnost dle potřeby o výsledcích šetření diskutovat; dosavadní výzkumy, jak uvádí (3), ukazují, že kvalita výuky se nezlepší poté,

co učitel získá informace o tom, jak je vnímán studenty, ale teprve potom, když tuto zpětnou vazbu analyzuje, prodiskutuje s kolegy,

- zpětnou vazbu zde můžeme vnímat jako určitý motiv ke změně, který spouští další proměnou, a to učitelovu snahu zlepšit výuku,
- další závažnou otázkou je, do jaké míry mají být výsledky studentského hodnocení výuky zveřejňovány, nelze zapomínat na kulturu hodnocení; myslíme si, že rozhodnutí o tom, zda budou výsledky šetření diskutovány např. na poradě katedry, či pouze individuálně, by mělo zůstat na rozhodnutí hodnoceného učitele.

Myslíme si, že studentské posuzování výuky a jeho vyhodnocení, tak jak bylo provedené na VVŠ PV ve Vyškově, lze využít. Je ale nezbytné počítat se všemi výše uvedenými aspekty. Nelze opomenout skutečnost, že jedině kompletní soubor dobře promyšlených a vzájemně se doplňujících metod hodnocení a jejich správné využití umožní vysokým školám zlepšovat kvalitu výuky.

Použité prameny a literatura

- 1 *GAVORA, P. Úvod do pedagogického výzkumu. Brno: Paido, 2000. 130 s. ISBN 80-85931-79-6.*
- 2 *CHRÁSKA, M. Závěrečná zpráva o řešení dílčího výzkumného úkolu v rámci vnitřního grantu UP. Aktuální teoretické problémy pedagogiky. Olomouc: PdF UP, 1995.*
- 3 *MAREŠ, J. Studentské posuzování jako jedna z metod hodnocení vysokoškolské výuky. Praha: SPN, 1991. 153 s. ISBN 80-7066-331-6.*
- 4 *SHAW, G., YOUNG, S. Profils of effctive college anhd university teachers. The Journal of Higher Education. Columbus, 1999. 70 s. ISSN 00221546*
- 5 *KOMENDA, S., KLEMENTA, J. Analýza náhodného v pedagogickém experimentu a praxi. Praha: SPN, 1981. 316 s.*

Kontakt:

PhDr. Eva Šmelová, PdF UP Olomouc, tel. 585635168, e-mail smelovaeva@seznam.cz

RNDr. Oldřich Kříž, VVŠ PV Vyškov, tel. 973452547, e-mail kriz@kmi.vvs-pv.cz