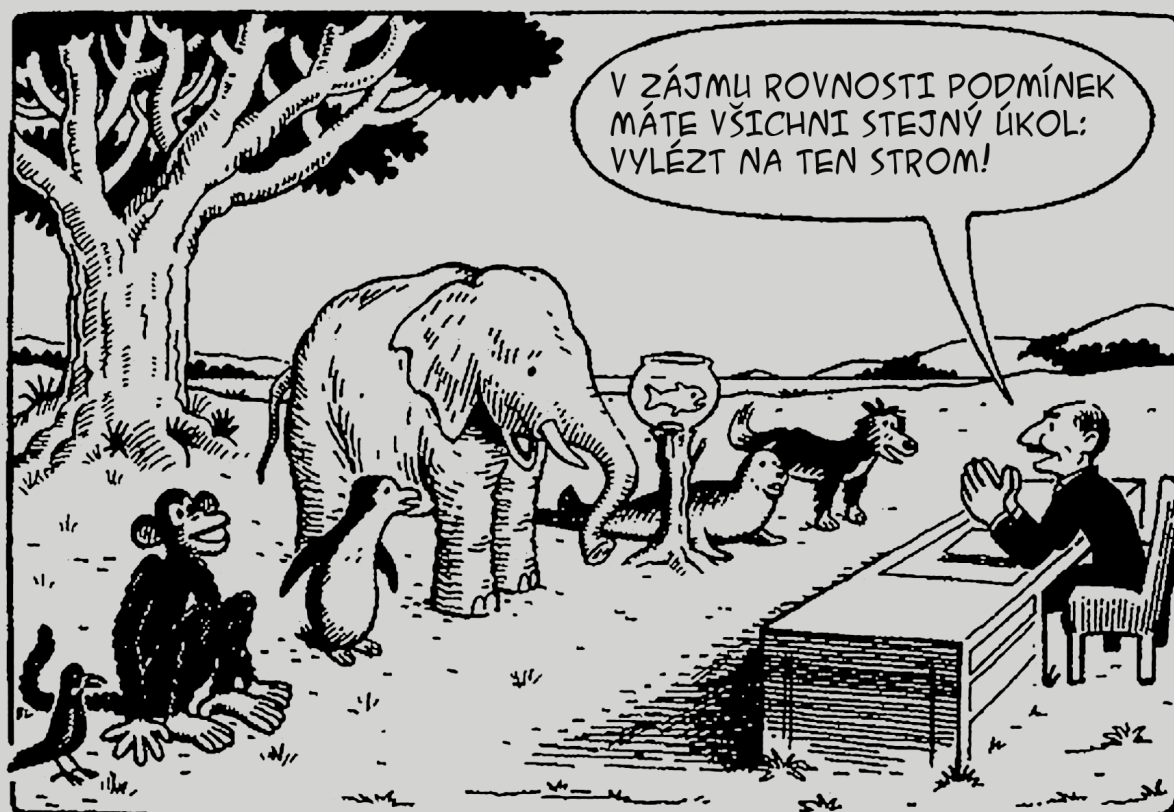




# MODERNÍ TRENDY V OBLASTI TESTOVÁNÍ PŘI VÝUCE MEDICÍNY A NELÉKAŘSKÝCH ZDRAVOTNICKÝCH OBORŮ

PŘÍRUČKA K WORKSHOPŮM O PŘÍPRAVĚ TESTOVÉ AGENDY  
NA LÉKAŘSKÉ FAKULTĚ MASARYKOVY UNIVERZITY







# **MODERNÍ TRENDY V OBLASTI TESTOVÁNÍ PŘI VÝUCE MEDICÍNY A NELÉKAŘSKÝCH ZDRAVOTNICKÝCH OBORŮ**

PŘÍRUČKA K WORKSHOPŮM O PŘÍPRAVĚ TESTOVÉ AGENDY  
NA LÉKAŘSKÉ FAKULTĚ MASARYKOVY UNIVERZITY



Podpořeno Fondem rozvoje Masarykovy univerzity  
(projekt č.: MUNI/FR/0057/2014)



# OBSAH



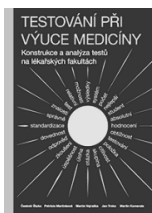
<b>ÚVODEM</b>	<b>5</b>
<b>AUTORSKÝ A LEKTORSKÝ KOLEKTIV</b>	<b>6</b>
<b>DŘÍVE NEŽ ZAČNETE</b>	<b>9</b>
Proč zkoušet	9
Co zkoušet	9
Jak zkoušet	10
S kým zkoušet	11
<b>CYKLUS PŘÍPRAVY TESTU</b>	<b>12</b>
Plánování testu (blueprinting)	13
Oponentura otázek/položek	14
Pilotování testu	14
Standardizace a normování testu	15
Realizace testů	16
Klasifikace studentů	16
Analýza výsledků a hodnocení kvality testu	17
<b>OSNOVA SEMINÁŘE</b>	<b>18</b>



# ÚVODEM



Testování znalostí je klíčovým momentem v procesu vzdělávání a často hraje zásadní roli nejen pro jednotlivé vzdělávací instituce, ale i pro celou společnost. V mnoha případech právě forma a obsah testu rozhodují o tom, na které znalosti či dovednosti se student při studiu zaměří a osvojí si je natolik, že je o jejich naučení se schopen přesvědčit také zkoušející a následně odborníky v oboru, který studuje a bude v budoucnu profesně vykonávat. S ohledem na fakt, že je problematika testování na vysokých školách stále aktuálnější, byla z projektových zdrojů Masarykovy univerzity podpořena sada edukačních seminářů cílených na podporu obecného povědomí a prohloubení pedagogických kompetencí akademických pracovníků v oblasti přípravy testové agendy. Semináře jsou záměrně určeny pro omezený počet posluchačů (v ideálním případě v rámci jednoho pracoviště) tak, aby bylo možno ve spolupráci s odbornými lektory prakticky realizovat vybrané fáze životního cyklu testové agendy, který zahrnuje plánování testu (blueprinting), identifikaci typů otázek/položek a jejich vytváření, oponenturu otázek/položek, pilotování testu, standardizaci a normování testu a v neposlední řadě také samotnou realizaci, analýzu výsledků testování a hodnocení kvality testu, potažmo evaluaci kvality vzdělávacího procesu. Účastníci seminářů se podrobněji seznámí se způsoby hodnocení znalostí, do detailu rozeberou možnosti, které poskytuje písemné či elektronické zkoušení. Pozornost bude věnována vlastní tvorbě testů s důrazem na jednotlivé druhy položek, jejich výhody i rizika použití. Akcentovány budou nové formáty otázek/položek, které umožňují ověřovat a zkoušet nejen izolovanou znalost faktů, ale i míru pochopení tématu a probírané látky. Jak již bylo uvedeno, pozornost nebude věnována pouze tvorbě otázek a položek, ale také celému cyklu přípravy testu: stanovení cílů výuky a plánování testu, přípravě položek, jejich oponování a pilotního odzkoušení, vyhodnocování kvality položek a celého testu, standardizaci a převodu výsledků testů do formy klasifikace. Pro přípravu kvalitních baterií otázek/položek a testů je nezbytná koordinovaná spolupráce týmu odborníků, jejichž investovaný čas se promítá do prostředků vynaložených na položku. V neposlední řadě budou zmíněny také vybrané softwarové nástroje pro realizaci testů a hodnocení kvality jejich položek s ohledem na použitelnost v prostředí Masarykovy univerzity.



Obsah souhrnné příručky vychází z poznatků publikovaných v recenzované knize o metodách testování s názvem **Testování při výuce medicíny, Konstrukce a analýza testů na lékařských fakultách**.

Autoři: Štuka, Č., Martinková, P., Vejražka, M., Trnka, J., Komenda, M.

Praha: Karolinum, 2013, ISBN 978-80-246-2369-6.

V elektronické formě k dispozici ke stažení na adrese: [www.wikiskripta.eu/Testy](http://www.wikiskripta.eu/Testy)

Autoři zmíněné publikace se testováním v medicíně zabývají již několik let. Poprvé společně vystoupili v rámci konference MEFANET, kde se uskutečnil workshop na téma elektronické testování. V návaznosti na tuto akci vznikla pod záštitou sítě MEFANET pracovní skupina pro kontinuální podporu testové agendy. Edukační projekt MEFANET (MEdicinal FACulties NETwork) je zaměřený na budování a posílení spolupráce lékařských a nelékařských zdravotnických fakult ČR a SR při rozvoji výuky s využitím moderních informačních a komunikačních technologií. Členové týmu zastupují jednotlivé lékařské fakulty a postupnými dílčími aktivitami se snaží získávat nové zkušenosti a znalosti nejen v tuzemsku, ale především v zahraničí. V případě zájmu lze kontaktovat buď vybraného autora, nebo použít pro hromadné postřehy, připomínky a případné náměty na spolupráci kontaktní email [testy@wikiskripta.eu](mailto:testy@wikiskripta.eu).

# AUTORSKÝ A LEKTORSKÝ KOLEKTIV



## **Čestmír Štuka** (stuka@cesnet.cz)

Vystudoval fyziku na Matematicko-fyzikální fakultě Univerzity Karlovy v Praze. Ve Všeobecné fakultní nemocnici v Praze se zabýval výzkumem rázových vln v medicíně a zavedením neinvazivní litotrypse do praxe. Spolupracoval na rozvoji informatiky ve VFN, nyní vede Oddělení výpočetní techniky I. LF UK v Praze. Pracuje na elektronické podpoře výuky medicíny a byl u zrodu spolupráce lékařských fakult MEFANET a projektu sdílené tvorby výukových materiálů WikiSkripta. Testování se věnuje již od své dizertace, zaměřené na analýzu přijímacího řízení.



## **Patrícia Martinková** (martinkova@cs.cas.cz)

Vystudovala matematiku na Matematicko-fyzikální fakultě Univerzity Karlovy v Praze, tamtéž absolvovala také doktorské studium v oboru pravděpodobnost a matematická statistika. Mezi její hlavní vědecké zájmy patří biostatistika a psychometrie. Od roku 2001 působí na Ústavu informatiky AV ČR, kde se zabývá vývojem modelů pro hodnocení kvality psychologických a znalostních testů. Pokročilé postupy aplikuje v praxi při analýze medicínských a didaktických dat. Podílí se také na výuce medicínské statistiky na úrovni pregraduálního i postgraduálního studia.



## **Martin Vejražka** (martin.vejrazka@lf1.cuni.cz)

Vystudoval všeobecné lékařství na 1. lékařské fakultě Univerzity Karlovy v Praze, na těžší fakultě posléze absolvoval i doktorské studium v oboru biochemie a patobiochemie. Ve výzkumu se zabývá úlohou volných radikálů a reaktivních forem kyslíku na vzniku některých onemocnění i ve fyziologických procesech. Kromě toho, že vyučuje lékařskou biochemii, se věnuje také elektronické podpoře výuky medicíny. Mimo jiné stál u zrodu WikiSkript, dnes nejnavštěvovanějšího webu pro pregraduální studium medicíny.





**Jan Trnka (jan.trnka@lf3.cuni.cz)**

Vystudoval lékařství na 3. lékařské fakultě Univerzity Karlovy v Praze, doktorské studium v oboru biochemie absolvoval na University of Cambridge a tamtéž studoval i historii a filozofii vědy. V současné době vede na 3. LF UK Laboratoř metabolismu a bioenergetiky, kde se zabývá mitochondriální funkcí a dysfunkcí způsobenou nutričními faktory. Věnuje se také výuce biochemie a tu a tam i filozofie vědy. K problematice testování se dostal už v rámci výuky v Cambridge, ale hlavně na 3. LF UK, kde se v rámci revize kurikula spolupodílel na tvorbě nového testovacího systému.



**Martin Komenda (komenda@iba.muni.cz)**

Vystudoval aplikovanou informatiku na Fakultě informatiky Masarykovy univerzity v Brně. Dlouhodobě se pohybuje v oblasti elektronické podpory a rozvoje výuky na lékařských fakultách. Na Institutu biostatistiky a analýz MU, kde působí jako odborný analytik, se věnuje především analýze a návrhu webově orientovaných nástrojů pro vzdělávání, optimalizaci medicínského kurikula s využitím moderních informačních technologií a vizualizacím zdravotnických dat.



**Andrea Pokorná (apokorna@med.muni.cz)**

Vystudovala obor učitelství pro zdravotnické obory na Palackého univerzitě v Olomouci, přičemž se již v té době zaměřovala na testování znalostí u studentů ošetrovatelství. Doktorské studium v oboru Hygiena a preventivní lékařství absolvovala na Masarykově univerzitě v Brně a habilitační řízení na Jihočeské univerzitě v Českých Budějovicích. Dlouhodobě se věnuje možnostem elektronického vzdělávání v nelékařských zdravotnických oborech a spolupracuje s Institutem biostatistiky a analýz MU v této oblasti.



# DŘÍVE NEŽ ZAČNETE



Hodnocení studentů je běžnou součástí činnosti vysokoškolského pedagoga. Může se zdát, že na něm není nic složitého. Mezi nejčastější úskalí a klíčové problematické oblasti však patří fakt stále se zvyšujícího počtu studentů a nároků na rozsah osvojovaných poznatků a z nich vyplývající dotazy. Jak vyzkoušet velké množství studentů v krátkém čase? Nebo jak prokázat objektivitu a reprodukovatelnost zkoušení? Nejen na tyto otázky, ale také na mnoho dalších se pokusí odpovědět následující kapitoly, které poskytnou čtenáři základní povědomí o jednotlivých tématech.

## PROČ ZKOUŠET

Zkoušení je realizováno proto, abychom zjistili, zda student dostatečně zvládl náplň daného předmětu, případně, zda může či nemůže postoupit do dalšího stupně studia. Tento typ zkoušení se označuje jako výstupní neboli sumativní a jeho primárním výstupem je hodnocení výkonu studenta (splnil/nesplnil, známka, počet bodů či umístění v rámci testované skupiny).

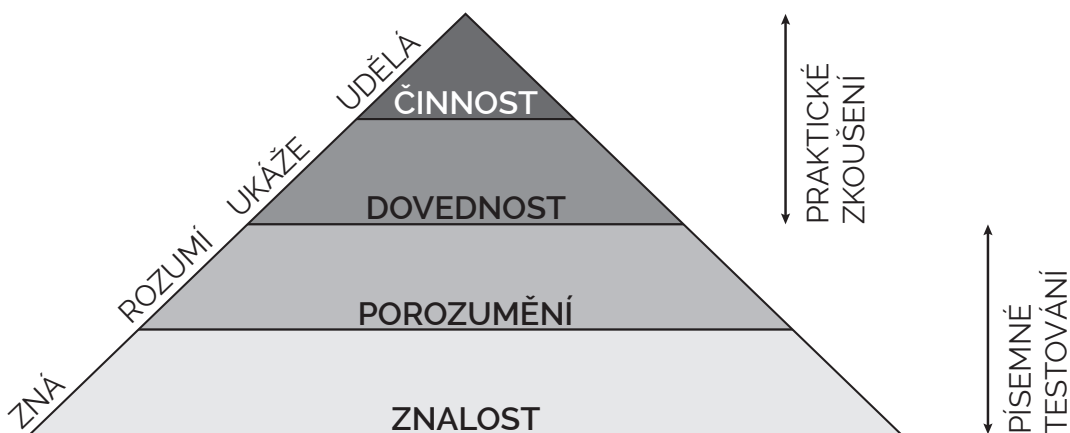
Různé formy zkoušení však také mohou sloužit jako zdroj zpětné vazby pro učitele či studenty. Například nás může zajímat, nakolik studenti v průběhu výuky vstřebávají a chápou předkládaný obsah, či které oblasti jim činí největší potíže a kde tedy učitel musí změnit strategii výuky. Můžeme také studenty testovat proto, aby oni sami našli svoje slabé stránky a mohli na nich zapracovat. Tomuto typu zkoušení se říká průběžné, nebo těž formativní.

V rámci jednoho předmětu, kurzu či ročníku je možné a vhodné oba typy zkoušení kombinovat, zejména na začátku studia a u rozsáhlejších předmětů. Účel zkoušení také napomáhá určit požadovanou úroveň znalostí – jiná kritéria lze uplatnit u jednoho z mnoha průběžných testů a jinak u závěrečné zkoušky, která může rozhodnout o ukončení studia.

## CO ZKOUŠET

Jednou z nejdůležitějších otázek, které je vhodné zodpovědět je, co přesně chceme zkoušet. Jednoduchá odpověď by mohla znít, že v biochemii budeme zkoušet biochemii a v patologii – co jiného než patologii. Z hlediska plánování formy zkoušení je však nutné se na věc podívat detailněji.

V průběhu výuky na vysoké škole se vyučující snaží studentům předávat soubor znalostí, dovedností a postojů „namixovaných“ v různých poměrech podle konkrétního předmětu. V biochemii se tak studenti učí intermediáty Krebsova cyklu (znalosti), chemické výpočty, a třeba i základy experimentální práce v laboratorním praktiku (dovednosti, případně i správné postoje k poctivému nakládání s experimentálními daty). V kurzech komunikace či etiky získávají dovednosti nutné k dobré komunikaci s pacienty, svými rodinnými příslušníky anebo kolegy, osvojují si profesionální postoje nutné k řešení složitých situací a jistě získají i teoretické poznatky o komunikačních technikách a etických systémech. Zásadní otázkou tedy je, co vše musí studenti prokazatelně zvládnout, aby daný předmět absolvovali, a co tedy budeme chtít zkoušet. Obvykle to nebude celý obsah kurzu, ale jen jeho centrální část. „Co se považuje za důležité“ by se mělo odvíjet od představy, jak má vypadat absolvent celého studia, a od potřeb dalších navazujících předmětů.



Millerova pyramida úrovní znalostí a dovedností umožňuje přehledně zobrazit úrovně vhodné pro písemné testování. Písemné testování (stejně jako např. testování pomocí počítače) lze použít pro hodnocení znalostí a porozumění. Vyšší úrovně dovedností je nezbytné zkoušet jinými způsoby.

## ÚROVNĚ ZNALOSTÍ A DOVEDNOSTÍ

Stejně jako musíme být schopní definovat obsahovou náplň zkoušky či testu, je také třeba přesně odpovědět na otázku, jakou úroveň znalostí nebo dovedností chceme testovat. Zkoušíme-li určité odborné téma, můžeme po studentovi požadovat, aby prokázal:

- **Znalost** (student zná; v anglicky psané literatuře knowledge, úroveň knows). Zkoušíme diagnostiku plicní embolie. Za znalost považujeme, že student má nastudované teoretické údaje o plicní embolii, zná nejčastější zdroje embolizace, rizikové faktory, patofyziologii změn, k nimž embolie vede atd.
- **Porozumění** (student ví jak; competence, knows how). Zkoušený dokáže znalosti z předchozí úrovně zapojit do kontextu – dokáže určit, jaké výsledky zobrazovacích metod jsou kompatibilní s diagnózou plicní embolie, jaké jsou očekávané výsledky jednotlivých laboratorních a klinických vyšetření apod.
- **Dovednost** (dokáže ukázat jak; performance, shows how). Dovednost je již komplexní, zkoušený se „sám vyzná“ a kombinuje široké spektrum znalostí a schopností, kterých často nabyl v různých předmětech a částech studia. Určí diagnózu, nebo vysloví podezření na plicní embolii na základě předloženého popisu konkrétního případu, rentgenových snímků a výsledků laboratorních vyšetření.
- **Činnost** (v praxi provádí správně veškeré potřebné úkony; action, does). Této úrovně by měl dosáhnout např. kandidát u státní závěrečné zkoušky nebo u atestace. Dokáže v praxi od pacienta odebrat anamnézu, fyzikálně jej vyšetřit, správně zajistit a předepsat adekvátní vyšetření, interpretovat jejich výsledky, ordinovat správnou léčbu atd.

## JAK ZKOUŠET

Pokud máme jasno v tom, jaké znalosti, dovednosti či postoje chceme zkoušet, můžeme přemýšlet o vhodných formách zkoušení. Zhruba je můžeme rozdělit na ústní a písemné. V tomto textu jako svébytnou skupinu vyčleníme ještě praktické zkoušení, které hraje v zdravotnickém vzdělávání a medicíně velmi významnou roli. Každá z uvedených tří forem má své výhody a nevýhody.

**Písemné či počítačové zkoušení** je velmi vhodné pro hodnocení velkého množství studentů a velkého rozsahu látky. Jeho hlavní přednost spočívá v hodnocení znalostí od přehledového zkoušení velkých souborů faktů, přes porozumění jejich souvislostem až po simulované řešení problémů či klinických situací. Principiální výhodou je jeho zpětná přezkoumatelnost a možnosti statistické analýzy výsledků. Z metodického hlediska je do značné míry lhostejné, zda se zkoušení provádí písemně (na papíře) nebo elektronicky na počítači.

Ústní zkoušení je vhodnější ke zjišťování schopnosti řešit problémy, zejména ty hůře strukturované, a tedy podobné reálným situacím. Nevýhody ústního zkoušení zahrnují především velkou časovou a personální náročnost, problematickou standardizaci, obtížnost srovnání jednotlivých zkušebních komisí a termínů a nemožnost zpětného přezkoumání. Relativní nevýhodou je také riziko antipatií a předsudků. Výhodou je naopak možnost eliminovat případné nedorozumění v pochopení zadání otázky či odpovědi na ni vzájemnou komunikací studenta a pedagoga.

**Praktické zkoušení** budeme v tomto textu považovat za specifickou problematiku, které se budeme věnovat jen okrajově. Praktické zkoušení hraje významnou úlohu zejména v klinické části studia. Mělo by být zásadní součástí závěrečných zkoušek a později i zkoušek konaných např. v rámci specializačního vzdělávání. Praktická zkouška má často několik součástí, při nichž se hodnotí výkon zkoušeného při určité činnosti. Často ale zkoušený odpovídá i na otázky, i praktická zkouška tak nese prvky písemného či ústního zkoušení. Více informací o praktickém zkoušení naleznete v příloze.

Konkrétní formu zkoušení zvolíme podle toho, co chceme zkoušet, jaký je rozsah zkoušené látky, kolik studentů je třeba ohodnotit, jaké k tomu máme personální či technické podmínky a jak spravedlivě či přesně potřebujeme v dané situaci zkoušet. Pět studentů pravděpodobně vyzkoušíme mnohem rychleji ústně než písemně, pokud započteme čas nutný na přípravu kvalitního testu. Naopak někdy můžeme být nuceni použít písemné zkoušení z praktických důvodů (nedostatečné personální zajištění), i když v daném případě nemusí jít o optimální formu.

Obecně se dá říci, že písemné či počítačové testování je vhodné pouze pro zkoušení znalostí a porozumění. Vyšší úroveň dovedností je třeba hodnotit principiálně jinými metodami (tedy již zmíněným praktickým zkoušením). Jistě není třeba zdůrazňovat, že znalosti a porozumění jsou nutnou, nikoliv však postačující úrovní pro úspěšného absolventa mnoha částí studia; písemné testování tak nutně nemůže být jedinou metodou hodnocení, v určité fázi na něj musí navazovat jiné přístupy ověřování.

## S KÝM ZKOUŠET

Předpokladem efektivního a relevantního zkoušení je jeho dobrá organizace. U předmětů s větším rozsahem, významem v rámci studijního programu a s velkým počtem studentů je prakticky nezbytné vytvořit dobře fungující tým, který se bude organizaci zkoušení systematicky věnovat, a to včetně zpětné kontroly jeho kvality.

Výhody zkušebního týmu jsou zřejmé: rozloží se nápor práce a umožní se účinná kontrola kvality a vzájemná podpora. Týmová spolupráce je také nezbytná pro standardizaci testů. Některé zahraniční univerzity do zkušebních týmů zařazují i externí členy z jiných institucí, čímž posilují nestrannost komisí a zajišťují srovnatelnost zkušebních standardů mezi jednotlivými vzdělávacími institucemi.

# CYKLUS PŘÍPRAVY TESTU



Podívejme se, s čím při přípravě, realizaci a vyhodnocení testů můžeme být konfrontováni. Předpokládejme, že jsme se rozhodli otestovat znalost skupiny studentů pomocí písemného testu. Příprava písemného testu je náročnější než samotné ústní zkoušení; musíme tedy pro takové rozhodnutí mít nějaký důvod. Může to být potřeba vyzkoušet v omezeném čase velké množství studentů, či potřeba zajistit spolehlivé a reprodukovatelné hodnocení.

Nejjednodušší (tzv. nestandardizovaný) písemný test lze sestavit ad hoc, pouze na základě zkušeností vyučujícího. Není na tom nic špatného, pokud je účelem testu pouhé poskytnutí zpětné vazby studentům nebo vyučujícím. Má-li však být výstupem klasifikace nebo rozhodnutí se závažnějšími důsledky (např. rozhodování o postupu studenta do dalšího studia), měla by být přípravě testu věnována patřičná pozornost, aby bylo hodnocení validní, objektivní a reprodukovatelné.

Jak by měl vypadat cyklus přípravy testu? Jeho základní kroky odhadneme i intuitivně: Máme-li promyšlené zadání testu, můžeme podle něj test navrhnout. Vytvoříme otázky, které si během recenze necháme zkontrolovat erudovanými kolegy. Poté můžeme test realizovat, studenty oznámkovat a statisticky zhodnotit i samotný test během jeho analýzy. Poučíme se, promítneme zpětnou vazbu do celého cyklu přípravy a můžeme se pustit do přípravy dalšího testu. Na následujícím schématu je zobrazen úplný cyklus testové agendy. Jeho jednotlivým etapám se přehledově věnují navazující kapitoly.

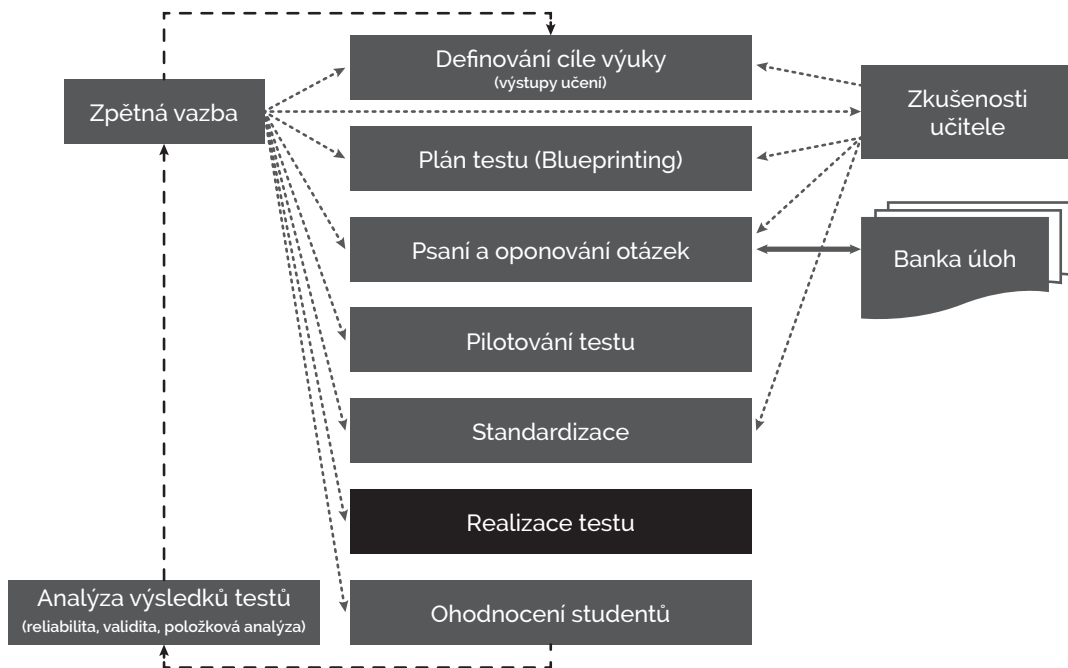


Schéma testového cyklu

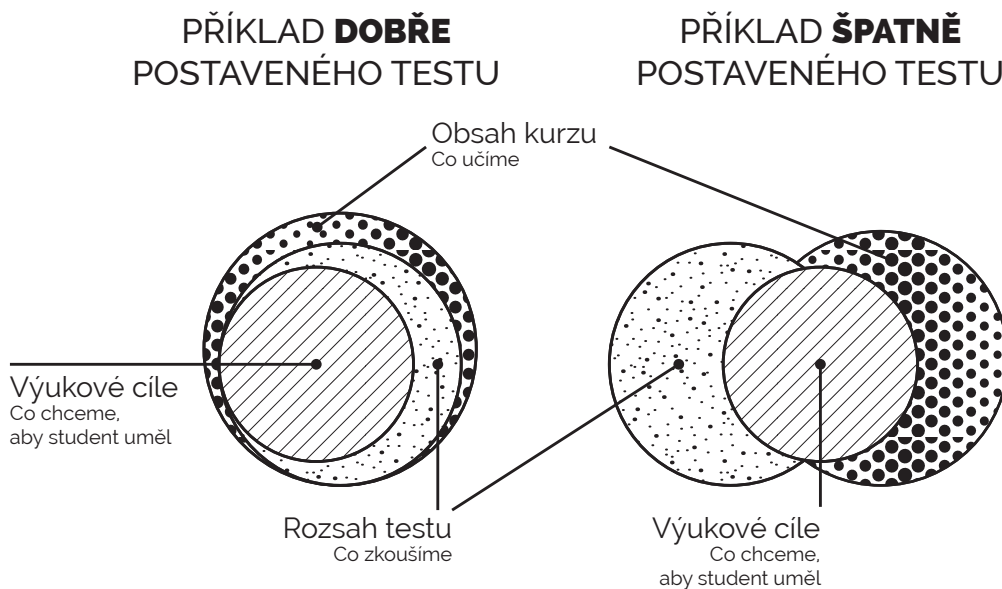
# PLÁNOVÁNÍ TESTU (BLUEPRINTING)

Účelem plánování testu (anglicky blueprinting) je definovat obsah testu a skladbu použitých metod a forem zkoušení. V této fázi přípravy testu hledáme odpověď na otázky:

- co přesně se bude zkoušet,
- jak velký podíl testu se bude věnovat kterému tématu,
- jakým typem úloh se bude zkoušet určitý typ znalostí či dovedností.

Jinými slovy, při plánování testu je nezbytné navrhnout, kolik bude test obsahovat položek ze kterého okruhu a jakého typu tyto položky budou. Zvažujeme přitom význam jednotlivých témat v rámci celého předmětu. Pečlivým naplánováním testu se vyhneme situacím, kdy se po většinu kurzu výuka věnuje určitým problémům, které učitelé považují za zvláště důležité, v testu se však k těmto tématům objeví jen málo otázek a zbytek se zabývá jinými okruhy. Testované osoby jsou pak zaskočeny a výsledky obvykle neodpovídají ani názoru vyučujících. (Nejlepší preventivní strategií a samozřejmostí u dobrého pedagoga by mělo být to, že studenti budou s cíli výuky a požadavky na ukončení předmětu předem seznámeni.) Naopak dobrý plán testování umožní objektivně a srovnatelně klasifikovat i několik skupin studentů pomocí různých variant testů.

Připomeňme na tomto místě, že klíčovou informací pro plánování testu by mělo být to, co má absolvent umět (tj. „cíle výuky“ - „learning objectives“). Tomu by měl odpovídat jak obsah a rozsah kurzu, tak i obsah a rozsah testu. V praxi se obsah a rozsah znalostí a dovedností vyučovaných, zkoušených a žádaných vždy poněkud liší. Při plánování testu bychom tedy měli vycházet z cílů výuky, je však třeba přihlížet i ke skutečné náplni kurzu/předmětu. Výsledky testu se samozřejmě mohou a mají stát zpětnou vazbou pro zdokonalování kurzu v dalších semestrech výuky.



Cíle výuky, skutečně vyučovaná látka a obsah a rozsah testu se v praxi vždy poněkud liší. Velkým rozdílem mezi cíli výuky, skutečně probranou látkou a náplní testu má zabránit plánování testu. V dobře naplánovaném testu jsou všechna požadovaná a vyučovaná témata zastoupena proporcionálně, test je vyrovnaný. Nevznikají nežádoucí „překvapení“ vyplývající z toho, že studenti jsou testováni z něčeho jiného, než očekávali. A s čím neměli zkušenost z výuky. Z hlediska andragogiky a principů vysokoškolského studia je ovšem samozřejmé, že ne všechny okruhy budou probrány do důsledků a je vyžadována určitá míra samostudia, ale o tom musí být student informován.

## OPONENTURA OTÁZEK/POLOŽEK

Nedílnou součástí přípravy testů je oponentura. Je rozdělena do několika fází, které se vždy zaměřují na specifickou oblast. Jejím cílem je odhalení nedostatků, které zpravidla testy ve své počáteční podobě obsahují. Provedení oponentury je totožné pro všechny formy zkoušení. Stejný postup tedy lze použít při elektronických i písemných variantách testování. Motivací k provádění revize je objektivní zajištění správnosti, optimalizace testu a odstranění subjektivních vlivů. I přesto, že oponentura bývá zpočátku časově a organizačně poněkud náročnější, její přínos je nepopíratelný a roste s významem testu. Po úspěšném zvládnutí všech níže uvedených revizí (obsahová revize, revize férovosti, redakční revize) by měl finální podobu jednotlivých položek/úloh znovu projít autorský tým a všechny provedené změny odsouhlasit. Teprve potom přichází na řadu pilotní administrace testu na malé skupině studentů.

Oponentura, podobně jako příprava kompletní testové agendy, je založena na týmové spolupráci. Několik zainteresovaných odborníků nezávisle na sobě posuzuje vhodnost jednotlivých otázek a společnými silami se snaží o odstranění všech nedostatků, které by mohly při praktické realizaci působit potíže. Týmová spolupráce tedy hraje při oponování databanky položek/úloh zcela klíčovou roli. Subjektivní pohled autora dané položky může mnohdy negativně ovlivnit formulaci zadání tzv. kmen položky), případně také nabízených odpovědí (distraktorů), a tedy i správnost celé položky. Konstruktivní oponentura několika vyučujících zvyšuje především objektivitu otázek a významně tak přispívá ke zkvalitnění celého testu.

Celý proces oponentury testu lze rozdělit do tří fází, kterými oponenta provede formulář pro recenzenty úloh (podrobněji rozebrán dále v textu).

- **Obsahová revize:** Jsou odpovědi správně a přesně formulované? Nejsou distraktory diskutabilní?
- **Revize férovosti:** Testují otázky pouze požadovanou konkrétní znalost a nic jiného?
- **Redakční revize:** Jsou otázky dostatečně srozumitelné, typograficky jednotné a bezchybně jazykově i obsahově formulované?

## PILOTOVÁNÍ TESTU

Důvěryhodné testování výsledků výuky, zvláště pokud ovlivňuje další postup studentů ve studiu, předpokládá, že vlastnosti používaného testu budeme znát ještě před jeho „ostrým“ tedy netestovacím použitím. K odhadu vlastností testu slouží pilotní testování a pretestování. Oba pojmy se částečně překrývají; termín pilotní testování se v této práci většinou používá jako širší označení obou kroků. Pokud je třeba oba kroky rozlišit, rozumí se pojmem pilotní testování obecnější „proof of concept“ – jakási studie proveditelnosti, která na malé skupině studentů odhaluje případné chyby v konceptu a designu testu a může přinést i užitečnou subjektivní zpětnou vazbu. Termínem pretest se pak myslí formálnější a podrobnější předběžné prověření testu, které umožňuje odhadnout psychometrické vlastnosti otázek, jejich obtížnost, schopnost rozlišit mezi dobrými a slabými účastníky/absolventy testu a které umožňuje získat subjektivní i objektivní zpětnou vazbu od testované skupiny. Pretestování je srovnatelné s kroky, které se provádějí při vyvozování závěrů z „ostrého“ testování. Zatímco pro samotný pilotní běh testu stačí menší skupina studentů (například 20) s odpovídající úrovní znalostí a motivací, jako má cílová skupina, pro pretest, sloužící k výpočtu statistických parametrů položek, je třeba skupina větší, nejméně 100 respondentů. Vzhledem k nárokům na sestavení relevantní skupiny testovaných osob a mnohdy i časově náročnosti, se jako pretest často používá první „ostrý“ běh samotného testování. Podněty získané z vyhodnocení



předběžných testů je zapotřebí zpracovat v návrhu ostré verze testu. Zpravidla je třeba upravit některé položky. Pokud pretest prokáže významné nedostatky, může být nezbytné i přepracování celé koncepce testu.

- Subjektivní zpětná vazba: dotazník, diskusní fórum, diskuze ve frontální výuce (v případě menšího množství studentů, při větším počtu se tato varianta stává neefektivní), poznámky v testu nebo tzv. přemýšlení nahlas.
- Objektivní zpětná vazba: zhodnocení obtížnosti testových úloh (identifikace snadných a obtížných úloh, nevyhovujících otázek, možnost uspořádání úloh podle obtížnosti), určení citlivosti jednotlivých úloh (analýza a korekce nebo vyřazení úloh s nevyhovující citlivostí), vyhodnocení kvality testu jako celku, především jeho reliability a validity.

## STANDARDIZACE A NORMOVÁNÍ TESTU

Standardizace testu znamená zajištění rovnosti podmínek testovaných, porovnatelnosti jejich výsledků navzájem a porovnatelnosti výsledků z různých běhů testu tak, aby zkoušení bylo maximálně spravedlivé, objektivní a reprodukovatelné. Standardizované testy nabízejí všem respondentům stejný test za stejných (nebo přiměřeně rovných) podmínek, a jsou proto vnímány jako spravedlivější než jiná hodnocení, která používají nesrovnatelné otázky a podmínky pro studenty skládající zkoušky v různých termínech nebo u různých examinátorů.

Jednou z hlavních předností standardizovaného testování je, že výsledky mohou být objektivně dokumentovány a mají dostatečný stupeň spolehlivosti (reliability) a správnosti (validity). Výsledky standardizovaného testování jsou zobecnitelné a opakovatelné, což je odlišuje od běžného školního hodnocení, které je závislé na jednotlivém učiteli. Bez standardizovaného testování by bylo obtížné objektivizovat rozdíly ve vzdělávání jednotlivých škol či učitelů. Dobře navržený standardizovaný test poskytuje nejen informaci o znalostech jednatelce, ale při agregaci výsledků celých testovaných skupin může poskytovat další užitečné informace – např. možnost poměrně přesně porovnat výsledky různých tříd, škol nebo jiných skupin v časové ose.

Podle některých autorů „standardizované“ testy nemohou měřit iniciativu, tvořivost, představivost, koncepční myšlení, zvědavost, úsilí, ironii, úsudek, angažovanost, dobrou vůli, etické reflexe a celou řadu dalších hodnotných dispozic a atributů. To, co mohou měřit, jsou konkrétní dovednosti a znalosti, tedy nejméně zajímavé a nejméně významné aspekty učení. Kritici standardizovaných testů poukazují na uniformitu takového vzdělávacího modelu a produkování absolventů „jako na montážní lince“. Další námitkou je, že nadužívání a zneužívání standardizovaných testů poškozují výuku, neboť zužuje osnovy. Použití standardizovaného testování bez ohledu na cíle výuky totiž způsobuje, že co není testováno, se neučí; způsob zkoušení se pak stává vzorem toho, jak předmět učit. (Příkladem může být snaha o standardizaci maturitních zkoušek v ČR a rozsáhlá diskuse na toto téma v odborných kruzích a médiích.) Do jaké míry jsou tyto výhrady relevantní pro testování v rámci studia na lékařských fakultách, je ovšem zatím nezodpovězená otázka. Příznivci standardizovaného testování reagují, že nejde o kritiku standardizovaného testování jako takového, ale o kritiku špatně navržených testů.

- Relativní standardizace se neváže na obsah testu, ale hodnotí jednotlivé účastníky mezi sebou. Výhodou tedy je, že zabraňuje inflaci nejvyšších hodnocení, zřetelně odliší nejlepší studenty a není nutné individuálně standardizovat každý test zvlášť. Mezi nevýhody relativního hodnocení patří kolísání kvality úspěšných studentů podle kvality dané skupiny. Zejména u menších skupin se tedy

může stát, že uspějí i studenti s úrovní znalostí, která neodpovídá stanoveným či předpokládaným požadavkům. A naopak, část studentů nemůže v testu uspět, ani kdyby látku uměli sebelépe. Hodnocení studentů podle relativní standardizace odrazuje od spolupráce a týmové práce, protože si studenti uvědomují, že si navzájem konkurují v důsledku omezeného počtu nejvyšších hodnocení. Snižuje to i motivaci studentů oslabením vztahu mezi jejich úsilím a výslednou známkou, protože ta závisí nejen na jejich vlastním výkonu, ale i na výkonu ostatních. Zvláště v menších a homogenních skupinách může relativní standardizace zvětšit nepodstatné rozdíly. S ohledem na tato omezení bychom o užití relativního hodnocení měli uvažovat především ve velkých heterogenních skupinách, v nichž se nepředpokládá spolupráce.

- Absolutní hodnocení závisí jen na tom, co se student naučil, nikoli na jeho pozici mezi ostatními. Jeho nevýhodou je nutnost stanovovat kritéria úspěchu pro každý test zvlášť. Musí být nastavena tak, aby rozlišovala mezi studenty, kteří danou oblast dostatečně zvládli, a těmi, jejichž znalosti či dovednosti nejsou dostatečné k dalšímu postupu.
- Kombinovaná standardizace spojuje do jisté míry výhody absolutního hodnocení s kompetitivním aspektem hodnocení relativního. Studenti, kteří dosáhli absolutní hranice pro úspěšné složení zkoušky, jsou rozřazeni do skupin a podle dosažených bodů jsou jim přiděleny známky.

## REALIZACE TESTŮ

Moderní testování zhodnocuje současné technologie a obsahuje prvky inženýrského přístupu. Nahlíží se na něj jako na integrovaný systém systémů. Myslí se tím, že na systém pro tvorbu položek může navazovat systém pro jejich pilotování, poté systém pro předkládání testů a tak dále. Podle míry elektronizace testů rozlišujeme:

- počítačové testování (computer based assessment, CBA), v němž se celý cyklus hodnocení odehrává v elektronické podobě;
- počítačem podporované testování (computer assisted assessment, CAA), které zahrnuje i kroky prováděné „ručně“, tedy např. vytvoření a zpracování papírové formy testů.

V rámci Masarykovy univerzity je možné využít obě výše zmíněné varianty. Elektronické testování v počítačových učebnách lze realizovat přímo v lokálním Informačním systému Masarykovy univerzity, počítačem podporované testování = papírové testování s automatickým vyhodnocením formou skenování.

## KLASIFIKACE STUDENTŮ

Klasifikace, čili ohodnocení – oznámkování výkonu studentů v testu, je pokračováním či rozšířením standardizace testu. Konstrukce klasifikační stupnice, respektive nastavení relace mezi výkonem v testu a hodnotícím – klasifikačním stupněm, je jediný subjektivní prvek, který do celého testování vstupuje. Je mu tedy třeba věnovat náležitou pozornost. Pro nastavení relace mezi výkonem v testu a klasifikačními stupni je nutno si ujasnit, co je vlastně v daném případě smyslem vysokoškolského vzdělávání.

### KLASIFIKACE POROVNÁVAJÍCÍ VÝKONNOST VE SKUPINĚ (RELATIVNÍ KLASIFIKACE)

Prvním přístupem je klasifikace založená na relativním výkonu ve skupině. Stanovuje známku jako funkci pořadí studenta v rámci určité skupiny. Pomocí testu je vytvořeno pořadí studentů a těm rozdělen

klasifikační stupeň podle předem stanovených procentuálních mezí. Relativní klasifikace je založena na předpokladu, že výkonnost všech studijních skupin (například prostorem a časem) je v zásadě stejná. Z pohledu studenta obsahuje tento způsob klasifikace zjevnou nespravedlnost, protože hodnocení nezávisí jen na výkonu studenta, ale i na výkonech ostatních. Je tedy možné, že se stejnou mírou znalostí by student byl v jednom roce klasifikován lépe než v roce jiném.

## KLASIFIKACE ZALOŽENÁ NA KRITÉRIÍCH (ABSOLUTNÍ KLASIFIKACE)

Naproti tomu klasifikace založená na kritériích měří úspěšnost studenta ve vztahu ke kritériím vyžadovaným pro dosažení toho kterého klasifikačního stupně. Na rozdíl od normativní klasifikace není předem stanoveno rozdělení klasifikačních stupňů a studentovo oznámkování není ovlivněno výkonností ostatních. Pokud bychom teoreticky zkoušeli výrazně nadprůměrnou skupinu studentů, mohou všichni dostat dobré známky, a naopak, pokud by se náhodou sešla velmi slabá skupina studentů, nemusí dobré známky dostat nikdo. Cílem klasifikace založené na kritériích je zhodnotit objektivní výsledky studenta nezávisle na výsledcích ostatních studentů.

## ANALÝZA VÝSLEDKŮ A HODNOCENÍ KVALITY TESTU

Výsledky testování mohou mít v praxi dalekosáhlé dopady a důsledky – například rozhodnutí o přijetí, rozhodnutí o udělení atestace nebo o udělení titulu. Použijeme-li nevhodné testy, závěry učiněné z jejich výsledků mohou být zavádějící. Kvalita použitých testů a jejich položek je proto zásadní a je důležité ji pravidelně ověřovat. Jaké vlastnosti by měl mít kvalitní znalostní test? V první řadě by měl měřit znalost, kterou měřit chceme. Dále by měl měřit co nejpřesněji a výsledky by měly být reprodukovatelné, zadáme-li studentovi jinou verzi téhož testu. Testy a jednotlivé položky by měly být spravedlivé a neměly by zvýhodňovat některé skupiny. Jednotlivé položky by měly mít vhodnou obtížnost a měly by mít schopnost dobře rozlišovat mezi různými úrovněmi znalostí studentů. Jak ověřit, zda náš test všechny tyto vlastnosti splňuje? Mnohé napoví analýza výsledků testu.

Jak analýzu výsledků provést? Prvním krokem analýzy výsledků by měl být souhrnný popis a vhodné grafické vyobrazení celkových výsledků všech studentů. Taková sumarizace se často vyžaduje při vykazování výsledků nebo pro sledování vývoje výsledků u stejného testu v průběhu let. Ukážeme si ale, že může být také prvním nástrojem k identifikaci problémů testu, například při nečekaném prozrazení („vynesení“) části úloh. V dalším kroku je důležité prozkoumat vlastnosti testu jako celku, především jeho reliabilitu neboli spolehlivost, a jeho validitu, tedy zda měří znalost, kterou měřit chceme. Ke kvalitě celého testu přispívají jednotlivé položky (otázky v testu), je proto potřeba sledovat také vlastnosti jednotlivých položek a nabízených distraktorů. Tato položková analýza v sobě zahrnuje odhady obtížnosti a citlivosti jednotlivých položek. Je důležitá také pro tvorbu dvou nebo více srovnatelných verzí testu.

# OSNOVA SEMINÁŘE



Předpokládaná délka semináře je 90 minut + závěrečná diskuse.

<b>Úvod</b>	
▸ představení projektu, motivace, pracovní skupina, osnova semináře	10 min
<b>Dříve než začnete</b>	
▸ benefity a možná úskalí testování	10 min
▸ proč, co, jak a s kým testovat	
<b>Životní cyklus přípravy testu</b>	
▸ představení kompletního schématu testování	5 min
<b>Plánování testu (blueprinting)</b>	
▸ obecný úvod a motivace	15 min
▸ nejčastější typy otázek + hlavní doporučení pro volbu otázek	
▸ ukázka vyplnění formuláře pro plánování testu	
▸ vytvořte ve skupinách jednu SBA otázku	
<b>Oponentura otázek</b>	
▸ obecný úvod a motivace	15 min
▸ představení jednotlivých typů oponentur	
▸ prostřednictvím formuláře zrecenzujte SBA	
<b>Standardizace a normování testu</b>	
▸ obecný úvod a motivace	15 min
▸ představení jednotlivých typů standardizace	
▸ schematický příklad ilustrující hodnocení studentů při relativní a absolutní standardizaci	
▸ prostřednictvím Angoffovy metody určete klasifikaci pro vytvořené otázky	
<b>Klasifikace studentů</b>	
▸ obecný úvod a motivace	10 min
▸ relativní a absolutní klasifikace v praxi	
<b>Realizace testů</b>	
▸ představení e-agendy na Masarykově univerzitě	5 min
<b>Analýza testu</b>	
▸ obecný úvod a motivace	5 min
▸ přehled metod pro analýzu testu	
<b>Závěr a diskuse</b>	-

×





# **MODERNÍ TRENDY V OBLASTI TESTOVÁNÍ PŘI VÝUCE MEDICÍNY A NELÉKAŘSKÝCH ZDRAVOTNICKÝCH OBORŮ**

**PŘÍRUČKA K WORKSHOPŮM O PŘÍPRAVĚ TESTOVÉ AGENDY  
NA LÉKAŘSKÉ FAKULTĚ MASARYKOVY UNIVERZITY**

Editor: Martin Komenda

Sazba: Radim Šustr

Vydal Institut biostatistiky a analýz, Masarykova univerzita v roce 2014





