

TESTOVÁNÍ (testing) a HODNOCENÍ (assessment)

Test je jednou z forem hodnocení studenta.

Existují různé druhy testů = (a samozřejmě také zkoušek): placement (rozřazovací, vstupní) test, finální testy; testy pro určitou jazykovou úroveň, nebo jen zkoumající určité konkrétní znalosti/dovednosti = tzv. průběžné testy (the Achievement Tests), který má za úkol zjistit úroveň zvládnutí učiva a užívá se k měření výsledku vzdělávání.

ASTRA = Adaptive Software for Testing (UJOP UK): <https://www.astra-testy.cz/>

Jedná se o nově vytvořenou formu rozřazovacího testu pro kurzy UK v Praze. Nicméně je může využívat široká veřejnost ke zjištění své jazykové úrovně v češtině. Končí úrovní C1 a ve spolupráci AI se snaží vytvořit text „na míru“, tedy podle vstupních dat a odpovědí na první otázku vytvářet otázky další tak, aby byly adekvátní vaší úrovni znalosti češtiny. Jedná se o tzv. mnohaúrovňový test. Tasové úkoly sestávají z porozumění poslechu a čtení, gramlexu a otevřených otázek, které substituuji produkci.

Hned na začátku je důležité si říct, **co je cílem testu?** (= **Co** chci ověřovat a **jak?**) Jakou informaci o studentovi a jeho znalosti češtiny potřebujeme získat?

PROČ testujeme? Monitorujeme pokrok studenta v kurzu; diagnostikujeme obtíže, které v češtině má; poskytujeme nám zpětnou vazbu za účelem zlepšení ať studenta samotného, učitel, nebo výuky. Celkově vede k vyšší efektivitě práce.

DOBŘÍ TEST by měl mít pozitivní účinek; být praktický (ekonomický co do času a finančních nákladů na jeho tvorbu); přesně měřit schopnosti, které jsou středem našeho zájmu (účel testu).

VLASTNOSTI didaktického testu:

OBJEKTIVITA – snaha vytvořit takové zadání, které nebude zvyhodňovat žádnou skupinu studentů. Dále stanovení přesných pravidel hodnocení, aby nemohlo docházet k „subjektivnímu“ hodnocení studenta. „*K zajištění objektivitě přispívá jednoznačná formulace úloh testu, shodné podmínky při jeho zadávání a především přesná a pro všechny stejná pravidla hodnocení žákovských odpovědí.*“ Rukověť, str. 11.

VALIDITA – měří test to, co má měřit? Slouží svému účelu? Jedná se tedy o shodu mezi výsledkem testu a jeho účelem. Pokud je nízká, výsledky, které získáme, nám vlastně nic neřeknou. „*Validita je nejdůležitějším hlediskem při tvorbě a hodnocení testu.*“ Rukověť, str. 12.

Obsahová validita = zda test měří tu konkrétní znalost/dovednost, kterou chceme měřit, zda ji svým obsahem dostatečně pokrývá.

Validita empirická – skóre dosažené v testu je porovnáváno s jinými kritérii, jsme schopni ji také nějak číselně vyjádřit.

Zjevná validita – posouzení tzv. na první pohled. Do jaké míry je test, jeho struktura a grafika, shodná s jinými testy téže série. Např. na letní škole máte každý rok skupina A2 a každý rok by měl být test stejně dlouhý, stejný prostor věnován stejným jevům, které hodnotíte, a měli byste hodnotit tytéž jevy. Jde také o to, abyste tytéž jevy každý rok v kurzu probírali. Tohle vše zajistí stejnou obtížnost, obtížnost adekvátní úrovni kurzu.

RELIABILITA (spolehlivost) – vysoká míra reliability nám umožňuje, že se můžeme na test spolehnout. Vedle spolehlivosti je zde důležitá také PŘESNOST. Obecně se jedná o veličiny, jichž je velice těžké dosáhnout. Měly by být maximálně eliminovány všechny vnější vlivy, které mohou výsledek testu

zkreslovat. Jednodušší je udržet vysokou reliabilitu v částech, které jsou homogenní a úzce spojené s látkou. Nižší reliabilita je napří u poslechu nebo čtení, kde jsou jednotlivé úlohy mnohem náchylnější k rozdílné obtížnosti.

CITLIVOST – jedná se o schopnost rozlišovat mezi žáky s různou úrovní skutečných znalostí a dovedností. Je-li test citlivý, měly by být výsledky žáků přiměřeně rozprostřeny po celé bodové škále. (Většinou se nejvíce žáků pohybuje kolem středu nebo mírně nad ním, několik z nich má výborné výsledky atakující plný počet bodů a několik z nich výsledky špatné, blížící se spodnímu konci bodování.)

FORMÁT testu – celková délka testu, počet testových úloh a jejich délka (počet hodnocených jevů), upřesnění obsahu testu, a tedy i druhu úloh, s tím souvisí celková forma testu a jeho zadávání. Časová dotace na celý test, popř. na jeho části. Charakter textů pro čtení a poslech musí být volen tak, aby opět nezvýhodňoval určitou skupinu studentů a netestoval obecné znalosti, které nesouvisí se znalostí jazyka. Velice důležité je ZADÁNÍ testových úloh, INSTRUKCE musí být zcela jasné a jednoznačně nejlépe vč. příkladu. Jak jsme již hovořili, jednotlivé úlohy by měly testovat jen jednu „věc“, např. znalost pádů, nikoli zároveň znalost lexika. Pokud k tomuto přesto dojde, MUSÍ na to být brán ohled u HODNOCENÍ (např. testování minulého času většinou spojuje kromě znalosti formy participia, také znalost týkající se použité pomocného slovesa a slovosled. Negativní formy mají také svá specifika. = to vše by hodnocení mělo zohledňovat).

Existují různé testovací techniky, které volíme podle očekávaného výstupu.

Je také možné vytvořit různé úrovně jednoho testu. Důležité je spíše pokud máme více variant jednoho testu (paralelní verze testu), aby byly vzájemně srovnatelné z hlediska obtížnosti (a samozřejmě také délka aj.).

KRITERIA HODNOCENÍ – jak již bylo uvedeno výše, jsou velice důležitá a to vč. tzv. vážení – co je považováno za chybu, kolik bodů je potřeba, aby student získal, resp. kolik procent z testu je ještě akceptovatelných. Zde se mluví o tzv. MEZNÍ HRANICE ÚSPĚŠNOSTI (= minimální standard).

Aby byla testovací úloha alespoň minimálně validní, měla by mít minim. deset položek.

Základní druhy testových úloh (také cvičení v učebnici):

A/ vícenásobná volba (multiple choice) – výběr ze 4 (minimálně 3) položek

B/ dichotomická volba (ANO/NE, pravda – nepravda)

C/ přiřazování (slov, částí věty, slovo a jeho význam, slovo/situace a obrázek, mezititulků k textu aj.)

D/ seřazování (části textu, pracovní postup, popis děje aj.)

E/ doplňování (zde může být s předem daným výběrem/bez výběru, tzn. jde o otevřeno úlohy, která hodnotí vlastní kreativitu/produkcii/aktivní znalost jazyka)

F/ krátké (jednoslovné) odpovědi

G/ větné transformace

H/ oprava chyb (vhodné až u vyšších úrovní, a i tak je sporné praktické využití)

I/ shrnutí (uvést hlavní body, souhrn, přehled, rekapitulace); **resumé** (Summary) = stručné shrnutí obsahu textu (článku) a jeho hlavních myšlenek. Dnes se místo toho častěji používá anglický výraz

Abstract, který však o něco rozsáhlejší a používá se např. u diplomových prací či odborných textů/příspěvků do sborníků, odborných časopisů apod. Často je doprovázen výčtem klíčových slov (key words).

J/ otevřené (dlouhé) odpovědi

Dále pak různé diktáty, překlady aj.

Celkově testové úlohy dělíme na:

UZAVŘENÉ – jsou objektivní, rychle hodnocení. Nevýhodou je, že hodnotí je pasivní znalost a je zde poměrně vysoká šance, že se student tzv. trefí. Také se snadněji opisuje. Jejich tvoření není úplně jednoduché!

- úlohy dichotomické
- úlohy s výběrem z více odpovědí
- úlohy přiřazovací
- úlohy uspořádací

OTEVŘENÉ – testují produktivní schopnosti, kreativitu. Snáze se rozpoznají nedostatky v zadání úlohy, které musí být zcela jednoznačné, aby se předešlo mylným interpretacím ze strany studentů. Těžké zajistit její objektivní hodnocení (co je dobře, a co už ne). Náročně se sestavují kritéria hodnocení, neexistuje klíč, který by pokryl všechna ne/možná řešení. Hodnocení je časově velmi náročné.

- úlohy se stručnou odpovědí
- úlohy se širokou odpovědí

TESTOVÁNÍ čtení s porozuměním (receptivní dovednosti, nejjednodušší, není tam časový stres jako u poslechu s porozuměním). Možnost objektivního hodnocení. (Mnohem vyšší než u produktivních dovedností typu psaní/mluvení.)

Pro tvůrce není zcela jednoduché vytvořit testové úlohy tak, aby netestovaly obecné znalosti, vybrat vhodný text, který bude splňovat kritérium obecnosti, a ten dále upravit směrem **k distraktorům**. (! **Zjistěte si, co to je?**)

Co testujeme?

čtení globální = za účelem celkové orientace v textu, pochopit jádro textu

čtení selektivní = informativní čtení (referenční příčka), získat specifické informace

čtení detailní = čtení a plnění pokynů, dopodrobna porozumět všemu, pochopit hlubší (nevyřčené) významy čteného (čtení mezi řádky) atd.

čtení pro radost

TESTOVÁNÍ poslechu s porozuměním (receptivní dovednosti, náročné kvůli časovému stresu - max. 2x se opakuje). Možnost objektivního hodnocení. (Mnohem vyšší než u produktivních dovedností typu psaní/mluvení.)

Pro tvůrce se opět jedná o těžké testové úlohy (viz čtení s porozuměním).

Důležité, kromě vlastního tématu, textu a úkolů, je také to **KDO je mluvčí** – věk, pohlaví, jazyk, profesionální mluvčí či nikoli; dialogy MUŽ a ŽENA; nutný je přesný scénář a další podklady pro nahrávání.

Co testujeme?

poslech globální = pochopit jádro věci

poslech selektivní = zachytil jednotlivé specifické informace

poslech detailní = dopodrobna všemu porozumět, popř. pochopit hlubší (nevyřčené) souvislosti atd.

Často se využívají různá veřejná hlášení (info, pokyny, varování); média (rozhlas nebo TV – zprávy, počasí, hlášení o situaci na silnicích aj.); nahrávky, filmy – pro vyšší úrovně.

Je možné využít i živého publika např. v divadle, na veřejných schůzích, na přednáškách/konferencích,

na zábavné akce nebo promo akce nabízející různé výrobky/produkty; někdy i nahodile vyslechnuté rozhovory.

Všechny testy by měly ideálně projít tzv. PRETESTINGEM (pilotáží testů), která odhalí nepřesnosti v zadání aj. Každý test musí být opatřený klíčem (uzavřené úlohy), nebo popisem správné odpovědi u otevřených úloh. (Zde je to opravdu náročné, zejm. pokud víme, že test bude využívat někdo jiný než sám autor testu).

! DU – Zjistit, co znamená tzv. TASK-BASED TESTING?

Vaší pozornosti doporučuji **RUKOVĚŤ autora testových úloh** od Radka Schindlera a kol. z r. 2006, která je dostupná zde: <https://docplayer.cz/17382733-Rukovet-autora-testovych-uloh.html>

Níže vkládám ještě dva slidy z prezentace paní dr. Kateřiny Vodičkové z VTC UJOP UK Praha.



Ústav jazykové
a odborné přípravy
Univerzita Karlova

Začínajícím autorům testů

- Měří váš test to, co skutečně měřit má?
- Nezjišťuje váš test především formálnosti a nepodstatné věci?
- Obsahuje váš test alespoň deset a více položek?
- Jsou vámi navržené úlohy navzájem nezávislé?
- Neobsahují úlohy vašeho testu nezamýšlené odpovědi?
- Vyhýbají se úlohy vašeho testu neadekvátním zdrojům obtížnosti?
- Neobsahují vaše testy chytáky?
- Jsou zadání vašich úloh srozumitelná nejen pro vás, ale i pro vaše studenty?
- Je váš test primárně určen opravdu k měření výsledků výuky?
- Přikládáte výsledkům svých testů adekvátní význam?



Čeho se v hodnocení vyvarovat

- Haló efekt = určitá vlastnost působí v oblastech, se kterými vůbec nesouvisí (např. kdo je učiteli sympatičtější)
- Perservační tendence = učitel při pozdějších posuzování už dříve vytvořený úsudek nezmění
- Efekt řady a kontrastu = učitel hodnotí studenta vzhledem k ostatním, nikoliv vzhledem k standardu
- Posuzovací tendence = princip mírnosti, princip přísnosti, tendence středu, tendence k extrémním následkům
- Chyby způsobené vědomím následků = učitel si uvědomuje, jaké následky může mít studentův neúspěch u testu/zkoušky