

Marie Vasilešková a Hana Marvánová

## **Rukověť autora testových úloh II – chemie**

Vydalo: Centrum pro zjišťování výsledků vzdělávání

Jeruzalémská 957/12

110 00 Praha 1

*www.cermat.cz*

Návrh obálky: Jan Jiskra

Jazyková korektura: Mgr. Radomír Silber, Ph.D.

Technická redakce: Bc. Monika Vítková

Sazba a tisk: MAAG promotion, a. s.

Praha 2006

Vydání první

## 4 VYHODNOCENÍ A OVĚŘOVÁNÍ TESTŮ

Vyhodnocení a ověření testu je poslední a nezanedbatelnou fází tvorby didaktického testu následující po fázích plánování a konstrukce testu.<sup>2</sup> Díky ní je možno určit, do jaké míry test vyhovuje předem daným kritériím a zda je vhodný k dalšímu použití. Díky statistické analýze lze také upravit, popř. nahradit nevhovující úlohy a upravit test do jeho konečné podoby.

Při analýze didaktických testů byly sledovány tyto charakteristiky: skóre testu a četnosti, výběrové charakteristiky, vlastnosti testů a vlastnosti jednotlivých testových úloh.

### A) SKÓRE TESTU A ČETNOSTI

Prvním krokem při vyhodnocování testů je hodnocení jednotlivých testových úloh, neboli vyhodnocování správných odpovědí. K ohodnocení úloh se zpravidla užívá bodů. Součet všech bodů, které daný žák v testu získal, se nazývá **skóre**(e), jinak řečeno skóre udává **celkový bodový výkon** žáka v testu. Existuje více možností, jak položky hodnotit, ale nejčastěji se užívá způsob, při kterém je správná odpověď hodnocena 1 bodem, nesprávná či vynechaná (chybějící) odpověď 0 body. V tomto případě pak zjišťujeme tzv. **hrubé skóre**, které se rovná počtu správně řešených úloh. Méně běžný způsob přiřazuje správné odpovědi 1 bod, nesprávné a vynechané odpovědi -1 bod. Potom se k vyjádření výsledku užívá tzv. **korigované skóre**, u kterého je snížena pravděpodobnost ovlivnění výsledku uhádnutím správné odpovědi -1 bod.

Informaci o tom, kolik žáků/procent žáků dosáhlo daného skóre, podává **absolutní/relativní četnost (frekvence) skóre**.

### B) VÝBĚROVÉ CHARAKTERISTIKY

Jde o veličiny, které informují o zjištěných hodnotách skóre a daných četnostech. Lze je rozdělit do dvou skupin, na charakteristiky polohy a charakteristiky rozptylu.

1) CHARAKTERISTIKY POLOHY (střední hodnoty) slouží k popisu rozložení četností a k jejich vzájemnému srovnávání.

**Aritmetický průměr** je nejpoužívanější střední hodnota. Je to průměrná hodnota sledovaného znaku (skóre) vztažená na celkový soubor (počet testů). Jeho nevýhodou je, že je významně ovlivněn několika jedinci souboru, kteří mají velmi velké nebo malé hodnoty znaku.

<sup>2</sup> Plánování testu v sobě zahrnuje určení struktury učiva, které má být testováno, dále určení počtu úloh v testu a určení úrovně osvojení poznatků, kterou mají úlohy ověřovat. Konstrukce testu pak zahrnuje návrh testových úloh a návrh prototypu didaktického testu. Blíže se tímto etapám tvorby didaktického testu tato práce nevěnuje a odkazuje na literaturu uvedenou v seznamu.

**Modus** je hodnota znaku (skóre), která se v testovaném souboru vyskytuje nejčastěji, tj. má největší četnost.

**Medián** je teoreticky prostřední hodnota znaku – to znamená, že vedle ní najdeme v souboru stejně množství jedinců s hodnotami jak vyššími, tak nižšími. Na rozdíl od aritmetického průměru podléhá medián podstatně méně vlivu jedinců s extrémními hodnotami.

2) CHARAKTERISTIKY ROZPTYLU (míry rozptýlenosti) udávají míru rozptýlení hodnot kolem průměru (jejich odchylku od průměrné hodnoty).

**Rozptyl (variance)** je průměr čtverců odchylek všech zjištěných hodnot od aritmetického průměru. Platí, že čím častěji se jednotlivé hodnoty odchylují od průměru, tím je rozptyl větší.

**Směrodatná (standardní) odchylka** je nejčastěji používanou mírou rozptýlenosti, protože na rozdíl od jiných měr (např. průměrné odchylky) zohledňuje větší odchylky více než malé.

**Variační koeficient** udává, jakou měrou se podílí směrodatná odchylka na aritmetickém průměru.

### C) VLASTNOSTI TESTU

V kapitole 1.1 již byly uvedeny požadavky, které jsou na kvalitní didaktický test kladeny.

Při statistické analýze se zjišťuje zejména obtížnost, reliabilita, citlivost a validita testu. **Obtížnost testu** se posuzuje indexem obtížnosti P, který může nabývat hodnot od 0 do 100 %. U dobrého testu by se měla hodnota P zhruba blížit ideální hodnotě  $P_{id}$ , přičemž platí:

- je-li  $P \gg P_{id}$ , pak je test příliš snadný;
- je-li  $P = P_{id} (\pm 10 \%)$ , pak má test odpovídající (vyhovující) obtížnost;
- je-li  $P \ll P_{id}$ , pak je test příliš obtížný.

**Reliabilita testu** se posuzuje prostřednictvím koeficientu reliability (korelačního koeficientu)  $r_T$ , který v praxi nabývá hodnot od 0 (pro testy naprosto nespolehlivé a nepřesné) až po hodnoty blízké 1 (pro testy s dokonalou spolehlivostí a přesností). U standardizovaných testů je požadováno, aby měl koeficient reliability minimálně hodnotu 0,80. Obecně platí, že čím má test více úloh, tím větší má reliabilitu. Z toho vyplývá, že reliabilitu téhož testu lze zvýšit zvýšením počtu testových úloh. Uvádí se, že dobrý didaktický test by měl obsahovat minimálně 10–15 úloh (viz kapitola 2).

**Citlivost testu** lze posoudit nepřímou buďto podle rozložení četností skóre, nebo podle variability skóre. V prvním případě by se grafické znázornění rozložení četností mělo přibližovat Gaussovu rozložení. V druhém případě se vychází z faktu, že citlivost testu je přímo úměrná variačnímu koeficientu, který u dobře konstruovaného testu přesahuje hodnotu 0,20 (viz kapitola 1.1).

**Validita testu** se posuzuje na základě rozboru jednotlivých testových úloh a porovnáním jejich obsahu s danými učebními osnovami a texty, což v praxi zpravidla provádí příslušný odborník, popř. skupina odborníků (viz kapitola 1.1).

#### D) POLOŽKOVÁ ANALÝZA

Analýza vlastností testových úloh je cenným zdrojem informací o kvalitě jednotlivých úloh, ale i o vědomostech žáků. Analýza se zaměřuje hlavně na obtížnost a citlivost jednotlivých úloh a také na nenormované odpovědi.

**Obtížnost úlohy** je určena **indexem obtížnosti**  $P_p$  (procento žáků ve skupině, kteří danou úlohu zodpověděli správně)<sup>3</sup>, přičemž platí:

- je-li  $P_p < 15\%$ , pak je úloha příliš obtížná;
- je-li  $15\% < P_p < 85\%$ , pak má úloha vyhovující obtížnost;
- je-li  $P_p > 85\%$ , pak je úloha příliš snadná.

U většiny úloh v testu by se měl index obtížnosti pohybovat zhruba kolem 50%.

Citlivost úlohy je rozlišovací schopnost dané úlohy, tj. udává, nakolik je úloha schopna rozlišit žáky s lepšími vědomostmi od žáků s horšími vědomostmi. Dobrá úloha by totiž měla zvýhodňovat dobré žáky před žáky slabšími. K určení citlivosti úlohy se užívá několik ukazatelů, z nichž nejběžnější je **diskriminační koeficient D** (koeficient ULI). Ten nabývá hodnot od -1 do +1, přičemž platí:

- je-li  $D > 0$ , pak úloha rozlišila lepší žáky od horších;
- je-li  $D = 0$ , pak úloha nerozlišila lepší žáky od horších;
- je-li  $D < 0$ , pak úlohu řešilo více horších žáků než lepších a je nutné přezkoumat formulace jednotlivých alternativ této položky.

**Analýza nenormovaných odpovědí (doplňková položková analýza)** je rozbor odpovědí vynechaných a nesprávných, při kterém se zjišťuje počet odpovědí na všechny nabízené distraktory. Platí, že všechny nabídnuté možnosti odpovědi, by měly být pro žáky stejně atraktivní, a vyskytne-li se distraktor, který nevolil nikdo či minimum z testovaných, je třeba jej vyloučit a nahradit vhodnějším. Doplnkovou analýzou lze navíc získat podrobnější informace o konkrétních nedostacích žáků a příčinách jejich neúspěchu při řešení dané úlohy, a z nich pak vyvodit závěry pro další výuku (např. test na výpočty odhalil nedostatky v názvosloví a nutnost znovu je s žáky procvičit). Přehled všech uvedených veličin spolu se vztahy pro jejich výpočet uvádí tabulka 1.

#### LZE HOVOŘIT O NEDOSTACÍCH A RIZICÍCH DIDAKTICKÝCH TESTŮ?

Mezi nevýhody didaktických testů patří zejména velká odborná i časová náročnost jejich tvorby, která je kladena na autora, popř. autory testu. Další nevýhodou je, že didaktický test nelze užít vždy, ne vše lze testovat, a navíc lze testem těžko postihnout komplexnější znalosti a složitější dovednosti žáků (za předpokladu, že je v testu užito pouze objektivních testových položek).

<sup>3</sup> Někdy se ke stanovení obtížnosti úlohy používá též hodnota obtížnosti Q (procento žáků ve vzorku, kteří danou úlohu zodpověděli nesprávně nebo nezodpověděli vůbec).

Rizika vystávají také s vyhodnocováním testů. Výsledky testů mohou být nesprávně či nepřipustně interpretovány a také mnohdy absolutizovány, tzn. že se jim přikládá jiný význam, než jaký odpovídá jejich plánovanému účelu.

Tabulka 1: Přehled základních statistických veličin

Veličina	Symbol	Vztah pro výpočet	Proměnné
<b>Skóre testu, četnosti</b>			
Hrubé skóre	$X$	$X = n_s$	$n_s$ ... počet správně řešených položek
Korigované skóre	$X_{kor}$	$X_{kor} = n_s \cdot \frac{n_t}{m-1}$	$n_s$ ... počet správně řešených položek $n_t$ ... počet nesprávně řešených/neřešených položek $m$ ... počet alternativ v položce
<b>Charakteristiky polohy</b>			
Aritmetický průměr	$\bar{X}$	$\bar{X} = \frac{1}{k} \sum_{i=1}^k X_i n_i$	$X_i$ ... hodnota hrubého skóre s četností $n_i$ $n$ ... počet analyzovaných testů $k$ ... počet různých variant skóre
Modus	$M_o$	$M_o = X_n$	$X_n$ ... nejčastější hodnota skóre
Medián	$M_e$	$M_e = dh + i \cdot \frac{n/2 - F_d}{f_m}$	$dh$ ... dolní hranice matematického intervalu obsahujícího medián $n$ ... rozsah výběru $F_d$ ... součet všech četností nacházejících se pod $dh$ $f_m$ ... četnost intervalu, ve kterém se nachází medián
<b>Charakteristiky rozptylu</b>			
Rozptyl skóre	$s^2$	$s^2 = \frac{1}{n} \sum_{i=1}^n n_i X_i^2 - \bar{X}^2$	$n$ ... rozsah výběru $X_i$ ... hodnota hrubého skóre s četností $n_i$ $\bar{X}$ ... aritmetický průměr skóre
Směrodatná odchylka	$s$	$s = \sqrt{s^2}$	$s^2$ ... rozptyl skóre
Variační koeficient	$V$	$V = \frac{s}{\bar{X}}$	$s$ ... směrodatná odchylka $\bar{X}$ ... aritmetický průměr skóre
<b>Vlastnosti testu</b>			
Obtížnost testu - index obtížnosti	$P$	$P = \frac{\bar{X}}{X_{max}} \cdot 100$	$\bar{X}$ ... aritmetický průměr skóre $X_{max}$ ... nejvyšší dosažitelná hodnota skóre
Ideální obtížnost	$P_{id}$	$P_{id} = 50 \left( 1 + \frac{1}{m} \right)$	$m$ ... počet alternativ v položce
Reliabilita testu - korelační koeficient	$r_1$	$r_1 = \frac{k}{k-1} \left[ 1 - \frac{\sum p_i q_i}{s^2} \right]$	$k$ ... počet úloh v testu $p$ ... podíl žáků ve vzorku, kteří řešili určitou úlohu správně, $q = 1 - p$ $s$ ... směrodatná odchylka
Citlivost testu - určuje se nepřímou dle variačního koeficientu $V$		$V = \frac{s}{\bar{X}}$	$s$ ... směrodatná odchylka $\bar{X}$ ... aritmetický průměr skóre
<b>Položková analýza</b>			
Obtížnost položky - index obtížnosti	$P_p$	$P_p = \frac{R}{n} \cdot 100$	$R$ ... počet správných odpovědí na danou položku $n$ ... celkový počet analyzovaných testů
Citlivost položky - diskriminační koef.	$D$	$D = \frac{R_h - R_d}{f}$	$R_h$ ... počet správných odpovědí žáků horní skupiny $R_d$ ... počet správných odpovědí žáků dolní skupiny $f$ ... koeficient rovnající se (při dělení žáků na poloviny) 0,50 $n$ ... počet testovaných žáků