

Výzkum učitele chemie na ZŠ (studijní materiál)

Hana Cídllová

2006

Obsah

1	Statistická analýza.....	4
1.1	Náhodný výběr.....	4
1.2	Vybrané metody statistického usuzování při náhodném výběru.....	5
1.2.1	Ověření předpokladů o datech.....	5
1.2.2	Postup testování statistické hypotézy o dvou výběrech.....	5
1.2.2.1	Odhad parametru polohy a parametru rozptýlení.....	6
1.2.2.1.1	Ověření nezávislosti prvků výběru.....	7
1.2.2.1.2	Ověření homogenity výběru.....	7
1.2.2.1.3	Ověření normality rozdělení výběru.....	7
1.2.2.1.4	Testy shody parametru rozptýlení.....	8
1.2.2.1.5	Testy shody parametru polohy.....	8
1.3	Korelace.....	11
1.3.1	Pearsonův párový korelační koeficient.....	11
1.3.2	Spearmanův pořadový korelační koeficient.....	11
1.3.3	Cronbachův korelační koeficient γ spolehlivosti výsledku.....	12
2	Didaktické testy.....	13
2.1	Typy didaktických testů.....	13
2.2	Charakteristiky didaktického testu.....	14
2.2.1	Objektivnost.....	15
2.2.2	Standardnost.....	15
2.2.3	Přiměřená obtížnost.....	15
2.2.4	Spolehlivost (reliabilita).....	16
2.2.4.1	Způsoby stanovení reliability.....	16
2.2.4.2	Doporučená hodnota reliability.....	18
2.2.5	Validita.....	18
2.2.6	Kvalitativní a kvantitativní interpretovatelnost a normalizace.....	19
2.2.7	Senzibilita (citlivost).....	19
2.2.8	Úspornost.....	20
2.2.9	Administrace didaktického testu.....	20
2.3	Základní kroky tvorby didaktického testu.....	21
2.3.1	Příprava didaktických testů.....	21
2.3.2	Tvorba učitelského ověřovacího didaktického testu.....	21
2.4	Testové položky.....	22
2.4.1	Typy testových položek podle formy.....	22
2.4.2	Zásady pro navrhování testových úloh.....	23
2.4.2.1	Obecné zásady pro navrhování všech druhů testových úloh.....	23
2.4.2.2	Návrh otevřených úloh se stručnou odpovědí.....	24
2.4.2.3	Některé zásady pro návrh přiřazovacích úloh.....	25
2.4.2.4	Některé zásady tvorby úloh s výběrem uspořádání.....	26
2.4.3	Typy položek podle náročnosti myšlenkových operací.....	26
2.4.4	Počet položek v testu.....	28
2.5	Analýza didaktických testů.....	28
2.5.1	Kvalitativní a kvantitativní analýza.....	28
2.5.2	Skórování odpovědí.....	29
2.5.3	Vlastnosti testových položek.....	30
2.5.3.1	Obtížnost položky.....	30
2.5.3.2	Citlivost položky.....	30
2.5.3.3	Analýza nenormovaných odpovědí.....	31
2.5.3.4	Vliv jedné položky na celkové skóre testu.....	32

2.6	<i>Ověřování a optimalizace didaktického testu</i>	32
2.6.1	Identifikace nevhodných testových položek.....	32
2.6.2	Vytvoření definitivní podoby didaktického testu	32
2.6.3	Standardizace didaktického testu.....	33
2.6.3.1	Percentilová škála.....	33
2.6.3.2	C-škála.....	33
2.6.3.3	Škála STANIN	34
2.6.3.4	z-škála.....	34
2.6.3.5	Z-škála.....	34
2.6.3.6	T-ŠKÁLA.....	35
2.7	<i>Využití didaktických testů k posouzení úrovně zvládnutí různých úrovní myšlenkových operací s učivem</i> 35	
2.7.1	Didaktickologické testy	35
2.7.2	Test představivosti	36
2.7.3	Cílené využití různých forem položek.....	38
3	Použitá literatura	39

1 Statistická analýza

1.1 Náhodný výběr

(podle Hniličkové [24])

Zajímají-li nás vlastnosti testu platné pro velké množství respondentů (např. zvolenou populaci v zemi), není z praktického hlediska možné zjistit je přímým měřením se všemi respondenty. V takovém případě je nutno provést výběr tzv. **reprezentativního vzorku**, jehož výsledky by bylo možno rozšířit na celý základní soubor.

Výběr reprezentativního vzorku musí probíhat podle určitých pravidel. Nejjednodušší metoda je tzv. **prostý náhodný výběr**, při němž má každá jednotka základního souboru (např. respondent) stejnou náději, že bude do vzorku pojata.

Výběr z velkého množství malých stejnorodých snadno míchatelných objektů se provádí poměrně snadno (objekty promícháme a pak z nich náhodně vybereme požadované množství). Je-li však výběrový postup aplikován na takové jednotky, jako jsou osoby (tj. objekty relativně velké a nemíchatelné), musí být zprostředkován určitými značkami totožnosti, kterými jsou tyto jednotky při výběru zastupovány. Základní soubor složený z těchto značek (zástupců) nazýváme **oporou výběru**. Opora výběru musí mít tyto vlastnosti:

- 1) převod mezi značkami a jednotkami musí být jednoznačný
- 2) je znám způsob, jak ke značce najít jednotku a obráceně
- 3) značky jsou doprovázeny všemi údaji o jednotkách, které potřebujeme znát, abychom mohli vykonat výběr zvoleným způsobem.

Prostý náhodný výběr někdy neuspokojuje. Chceme-li např. znát vlastnosti různých dílčích souborů v nehomogenním souboru základním, je užitečné nejprve rozdělit základní soubor na dílčí soubory (např. podle typů škol) a teprve pak v nich provádět náhodný výběr některým z dříve popsaných způsobů. Můžeme postupovat proporcionálně (počet náhodně vybraných prvků z každého dílčího souboru je úměrný rozsahu tohoto dílčího souboru), pokud jsou rozsahy dílčích souborů přibližně stejně velké. Je-li rozsah některých dílčích souborů velmi malý, byly by vzorky vytvořené z nich proporcionálně velmi málo početné, takže by se z experimentu s nimi daly vyvodit jen velmi chudé závěry. V takovém případě je zapotřebí výběr z velmi malých dílčích souborů proti proporcionálnímu způsobu zvýhodnit.

V souvislosti s takto složitě vybraným vzorkem vzniká otázka, jak a nakolik je možno výsledky získané z vybraného vzorku zobecnit na celý základní soubor. Zde bude popsán jen postup zpracování výsledků při prostém náhodném výběru. Existují však i složitější metody výběru (např. **skupinkový výběr, vícestupňový výběr, zamítací výběr, systematický neboli mechanický výběr a řada jiných**).

Pokud ze základního souboru čísel (např. testové skóre) provedeme nějaký výběr určitého počtu prvků, je možno vypočítat průměrnou hodnotu těchto čísel. Výsledek nazveme **výběrový průměr** \bar{X} . Náhodných výběrů je možno ze základního souboru udělat obvykle velmi mnoho. Různým náhodným výběrům pak mohou náležet různé hodnoty výběrového průměru.

Jedním z důležitých výsledků teorie pravděpodobnosti je poznatek, že za velmi obecných podmínek, jež bývají v praxi splněny, mají výběrové průměry rozložení přibližně normální, je-li rozsah náhodných výběrů dostatečně velký. Tohoto poznatku bylo využito v různých metodách statistického usuzování. Některé z nich, významné pro zpracování didaktických testů, budou uvedeny v dalším textu.

1.2 Vybrané metody statistického usuzování při náhodném výběru

(podle [40])

1.2.1 Ověření předpokladů o datech

V praxi se většinou předpokládá, že data tvoří náhodný výběr. Reprezentativní náhodný výběr je charakterizován třemi důležitými předpoklady, které je před dalším rozbořením dat vždy nutno ověřit. Jsou to:

- ověření nezávislosti prvků výběru
- ověření homogenity výběru
- ověření normality rozdělení výběru

1.2.2 Postup testování statistické hypotézy o dvou výběrech

- Nejprve provedeme **odhad parametru polohy** a **odhad parametru rozptýlení**. Způsob vyhodnocení závisí na počtu prvků ve výběru.
- Ověříme, že oba srovnávané výběry jsou **navzájem nezávislé**.
- Ověříme, zda výběr je **homogenní**.
- Ověříme, že oba srovnávané výběry mají **normální rozdělení**.
- Testujeme **shodu parametru rozptýlení**
 - o **u velkých výběrů**
 - **Klasický Fisherův-Snedecorův F-test**
(oba výběry jsou nezávislé a pocházejí z normálního rozdělení)
 - **Modifikovaný Fisherův-Snedecorův F-test**
(výběry nepocházejí z normálního rozdělení)
 - **Robustní Jacknife test F_J**
(výběry nepocházejí z normálního rozdělení a jsou v nich odlehlá data)
 - o **u malých výběrů (4-20 prvků) [39]**
- Testujeme **shodu parametru polohy**
 - o **u velkých výběrů:**
 - **Klasický Studentův t-test pro homoskedasticitu** (shodné rozptyly)
 - **Klasický Studentův t-test pro heteroskedasticitu** (odlišné rozptyly)
 - **Modifikovaný Studentův t-test** pro výběry odchýlené od normálního rozdělení
 - **Robustní Jacknife test polohy pro homoskedasticitu** a přítomná odlehlá měření
 - **Robustní Jacknife test polohy pro heteroskedasticitu** a přítomná odlehlá měření
 - o **u malých výběrů (4-20 prvků) [39]**
 - Posoudíme, zda dochází k překryvu obou intervalů spolehlivosti. Pokud se intervaly spolehlivosti překrývají, potom odlišnost středních hodnot obou výběrů na zvolené hladině významnosti pokládáme za nevýznamnou. Pokud se intervaly spolehlivosti nepřekrývají, pak odlišnost středních hodnot obou výběrů na zvolené hladině významnosti pokládáme za významnou.

1.2.2.1 Odhad parametru polohy a parametru rozptýlení

(podle [40])

Odhad těchto parametrů závisí na počtu prvků v souboru. Odlišujeme:

- zvláště malé výběry (2 prvky, 3 prvky)
- malý výběr (4-20 prvků)
- střední výběr (20-30 prvků)
- velký výběr (více než 30 prvků)

U didaktického testování připadají prakticky v úvahu výběry malé, střední a velké. U malých výběrů jsou závěry vždy zatíženy značnou mírou nejistoty, proto je použijeme jen tam, kde nelze získat větší počet dat.

Malý výběr

Použijeme tzv. **Hornův postup** (pro zájemce je popsán např. v [40]). Odhadem *parametru polohy* je tzv. *pivotová polosuma*. Odhadem *parametru rozptýlení* je tzv. *pivotové rozpětí*.

Střední výběr

Odhadem *parametru polohy* je aritmetický průměr:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}, \text{ kde}$$

(1)

\bar{X} aritmetický průměr
n počet měření
i číslo měření
 X_i i-tá naměřená hodnota

Jednou z charakteristik parametru rozptýlení je *výběrová směrodatná odchylka* [49]:

$$s = +\sqrt{s^2}, \text{ kde}$$

(2)

s směrodatná odchylka výběru
 s^2 *rozptyl (variance) výběru*:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \text{ význam symbolů viz výše.}$$

(3)

Pozn.: Pomocí programu Excel lze vypočítat rozptyl výběru pomocí funkce VAR.VÝBĚR a směrodatnou odchylku výběru pomocí funkce SMODCH.VÝBĚR [43].

Velký výběr

Odhadem *parametru polohy* je výběrový průměr (1).

Mírou rozptýlení je rozptyl:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \text{ kde}$$

(4)

σ^2 rozptyl (variance) velkého souboru, ostatní symboly viz výše.

Směrodatná odchylka se vypočte analogicky k (2):

$$\sigma = +\sqrt{\sigma^2}, \text{ kde}$$

(5)

σ směrodatná odchylka základního (velkého) souboru

σ^2 rozptyl (variance) základního (velkého) souboru

Pozn.: Program Excel vypočte rozptyl základního souboru (resp. velkého výběru) pomocí funkce VAR a směrodatnou odchylku základního souboru (velkého výběru) pomocí funkce SMODCH [43].

$$95\% \text{ interval spolehlivosti střední hodnoty je } \bar{X} - 1,96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}}.$$

(6)

Dalšími používanými charakteristikami rozptýlení výsledků jsou např. *varianční rozpětí* [67] R (rozdíl mezi nejvyšší X_{\max} a nejnižší X_{\min} naměřenou hodnotou): $R = X_{\max} - X_{\min}$

(7)

a *varianční koeficient* [67] (udává v procentech, nakolik kolísají naměřené hodnoty kolem aritmetického průměru:

$$V = \frac{\sigma}{\bar{X}} \cdot 100, \text{ kde}$$

(8)

V varianční koeficient

σ směrodatná odchylka

\bar{X} aritmetický průměr naměřených hodnot

1.2.2.1.1 Ověření nezávislosti prvků výběru

V [40] se pro identifikaci závislosti dat na čase nebo pořadí měření se používá **von Neumannovo kritérium**. V případě didaktických testů připadá (spíše než závislost na pořadovém čísle respondenta) v úvahu závislost na jeho umístění v učebně vůči ostatním respondentům (opisování). Objektivní vyhodnocení této závislosti by však bylo velmi složité. Proto se při zpracování dat získaných z didaktických testů většinou test nezávislosti prvků výběru neprovádí, případně prování pouze intuitivně.

1.2.2.1.2 Ověření homogenity výběru

Homogenní výběr znamená, že všechny prvky pocházejí ze stejného rozdělení s konstantním rozptylem. Problém zpracování dat s tzv. odlehlými daty je velmi komplikovaný a starší literatura o zpracování výsledků didaktického testování se jím vůbec nezabývá. Nejjednodušším řešením problému nehomogenity je rozdělit výběr podle nějakých logických kritérií do několika podskupin, pak statisticky vyhodnotit každou skupinu zvlášť a nakonec na základě testů shody středních hodnot v podskupinách rozhodnout, zda toto dělení je významné. V odůvodněných případech (např. identifikovaná hrubá chyba měření) je možno odlehlá data (lze je identifikovat např. pomocí *metody vnitřních hradeb* [39], [40]) ze souboru vyloučit.

1.2.2.1.3 Ověření normality rozdělení výběru

Normalitu rozdělení výběru lze u velkých souborů provést např. využitím **testu χ^2 dobré shody (Pearsonův test** - viz str. 10). U menších datových souborů je tento postup špatně použitelný, protože vyžaduje rozdělení dat do většího počtu tříd s četností alespoň 5. Lze však použít **test kombinace výběrové šikmosti a špičatosti dle Jarque-Berra** popsany např. v [40].

1.2.2.1.4 Testy shody parametru rozptýlení

Klasický Fisherův-Snedecorův F-test

(podle [37], [40])

Vychází se z předpokladu, že oba výběry jsou nezávislé a pocházejí z normálního rozdělení.

Uvažujme dva výběry s rozsahy n_1 , n_2 , se středními hodnotami \bar{X}_1 , \bar{X}_2 a s výběrovými směrodatnými odchylkami s_1 , s_2 – viz

(2),

(3). Testovacím kritériem je veličina

$$F = \frac{s_1^2}{s_2^2} \geq 1.$$

(9)

Tabulky kritických hodnot rozdělení $F_\alpha(f_1, f_2)$ jsou konstruované tak, že do čitatele zlomku dosazujeme vždy větší z rozptylů. Počet stupňů volnosti:

$$f_1 = n_1 - 1, f_2 = n_2 - 1$$

(10), (11)

Ověření významnosti rozdílu mezi dvěma rozptyly:

Pokud vypočtená hodnota F překročí kritickou hodnotu $F_\alpha(f_1, f_2)$, můžeme s pravděpodobností $(100 - \alpha) \%$ předpokládat, tzv. nulová hypotéza H_0 neplatí. Nulová hypotéza je předpoklad, že odlišnosti mezi rozptyly obou srovnávaných tříd jsou dány konkrétní volbou výběru, nikoli vlastnostmi tříd.

Hodnoty $F_\alpha(f_1, f_2)$, můžeme najít v tabulkách, nebo test můžeme provést pomocí vhodného software. Např. Excel [43] tyto hodnoty vrací pomocí funkce FINV ($\alpha; f_1; f_2$). Hodnota α je v tomto vztahu desetinné číslo udávající pravděpodobnost, že oba rozptyly významně neliší.

(12)

Modifikovaný Fisherův-Snedecorův F-test

(podle [40])

Pokud mají výběrová rozdělení jinou špičatost než odpovídá normálnímu rozdělení, je potřeba modifikovat počet stupňů volnosti používaný v F-testu, viz např. [40]. Další postup je stejný jako u klasického F-testu.

Robustní Jacknife test

použijeme, pokud data nepocházejí z normálního rozdělení a jsou-li v nich navíc odlehlé hodnoty. Postup provedení testu popisuje např. [40].

1.2.2.1.5 Testy shody parametru polohy

Klasický Studentův t-test T_1 pro homoskedasticitu (shodné rozptyly)

F-test v případě, že se srovnávají pouze 2 třídy, přechází na tzv. **Studentovo t-kritérium:**

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \text{ kde}$$

(13)

\bar{X}_1 , resp. \bar{X}_2 průměrné skóre v první, resp. druhé třídě

n_1 , resp. n_2 počet respondentů v první, resp. druhé třídě

s směrodatná odchylka vypočtená ze vztahu

(14)

$$s^2 = \frac{1}{f} \left[\sum_{i=1}^{n_1} (X_{i,1} - \bar{X}_1)^2 + \sum_{i=1}^{n_2} (X_{i,2} - \bar{X}_2)^2 \right] \quad (14)$$

$X_{i,1}$, resp. $X_{i,2}$ skóre i-tého respondenta v první, resp, druhé třídě

s směrodatná odchylka

i číslo respondenta

$\bar{X}_1, \bar{X}_2, n_1, n_2$ viz (13)

f počet stupňů volnosti. V daném případě

$$f = n_1 + n_2 - 2$$

(15)

Ověření významnosti pomocí t-testu:

Když vypočtená hodnota t podle vztahu (13) překročí kritickou hodnotu $t_{\alpha}(f)$, můžeme na hladině spolehlivosti α zamítnout nulovou hypotézu. Hodnoty $t_{\alpha}(f)$ můžeme buď najít v tabulkách (např. v [22]), nebo pomocí vhodného software. Nař. Excel [43] vrací tyto hodnoty pomocí funkce $TINV(\alpha;f)$ Hodnota α musí být zadána jako desetinné číslo (nikoli v procentech) udávající tolerovanou nejistotu.

(16)

Klasický Studentův t-test T_2 pro heteroskedasticitu (odlišné rozptyly)

(podle [40])

Pokud oba výběrové soubory mají normální rozdělení, avšak jejich rozptyly se liší, použijeme Studentův t- test pro heteroskedasticitu. Testovací kritérium má tvar:

$$T_2 = \frac{|\bar{X}_2 - \bar{X}_1|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \text{ kde}$$

(17)

T_2 testovací kritérium

\bar{X}_1, \bar{X}_2 výběrové průměry

s_1, s_2 výběrové směrodatné odchylky

n_1, n_2 počet prvků ve výběrech

Počet stupňů volnosti zjistíme pomocí vztahu:

$$v = \frac{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}{\frac{s_1^4}{n_1(n_1-1)} + \frac{s_2^4}{n_2(n_2-1)}}, \text{ význam symbolů viz výše.}$$

(18)

Další postup je stejný jako u Studentova t-testu pro homoskedasticitu.

Modifikovaný Studentův t-test T_3 pro výběry odchýlené od normálního rozdělení

Robustní Jackknife test polohy T_4 pro homoskedasticitu a přítomná odlehlá měření

Robustní Jackknife test polohy T_5 pro heteroskedasticitu a přítomná odlehlá měření

Potřebné výpočetní vztahy k testům T_1, T_2 a T_3 jsou relativně složité. Zájemci viz [40], str. 152-153.

Test χ^2 dobré shody (Pearsonův test)

je starší metodou testování shody parametrů (shoda parametru polohy, shoda parametru rozptylu) porovnávaných dat. Pearsonův test byl s oblibou používán v době malé dostupnosti počítačů. Dnes se již dává přednost metodám využívajícím sice složitější výpočetní vztahy, avšak s možností snadného automatického opakování.

1.3 Korelace

(podle [24] a [40])

Pro statistické zpracování **vícerozměrných** experimentálních dat jsou mimo jiné významné tzv. korelační koeficienty. Hodnoty korelačních koeficientů se pohybují v intervalu $\langle -1;1 \rangle$. Krajních hodnot je dosaženo jedině tehdy, když mezi náhodnými veličinami X , Y je přesně lineární vztah. Pokud jsou náhodné veličiny X , Y zcela nezávislé, je jejich korelační koeficient roven nule. Neplatí to však obráceně: Nulová hodnota korelačního koeficientu nemusí znamenat nezávislost veličin X , Y . Významná párová korelace také není důkazem příčinné souvislosti.

1.3.1 Pearsonův párový korelační koeficient

Tento koeficient vyjadřuje míru vazby mezi složkami x_1 , x_2 dvojrozměrného vektoru X . Pro jeho výpočet se v literatuře uvádí více ekvivalentních vztahů, které však na první pohled vypadají natolik odlišně, že nezkušenému pracovníku jejich totožnost nemusí být zřejmá. Např. Turek [66] uvádí:

$$r_{12} = \frac{n \sum_{i=1}^n (x_{1i} x_{2i}) - \sum_{i=1}^n x_{1i} \sum_{i=1}^n x_{2i}}{\sqrt{\left[n \sum_{i=1}^n (x_{1i}^2) - \left(\sum_{i=1}^n x_{1i} \right)^2 \right] \cdot \left[n \sum_{i=1}^n (x_{2i}^2) - \left(\sum_{i=1}^n x_{2i} \right)^2 \right]}}, \text{ kde} \quad (19)$$

r_{12} Pearsonův párový korelační koeficient

n počet měření

i číslo měření

x_{1i} , resp. x_{2i} hodnota složky x_1 , resp. složky x_2 vektoru X v i -tém měření

Výpočet pomocí vztahu (19) se v programu Excel [43] volá příkazem PEARSON(Pole1;Pole2).

Půlpán [49] a po opravě překlepu také Meloun [40] na str. 554 uvádí:

$$r_{12} = \frac{\sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sqrt{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 \cdot \sum_{i=1}^n (x_{2i} - \bar{x}_2)^2}}, \text{ symboly viz předchozí vztah.} \quad (20)$$

Výpočet pomocí vztahu (20) se v programu Excel [43] volá příkazem CORREL(Pole1;Pole2).

1.3.2 Spearmanův pořadový korelační koeficient

Jednotlivým měřením se přiřadí místo naměřených hodnot pouze pořadová čísla, odpovídající seřazení prvků podle velikosti. Označme pořadí prvků vzhledem ke složce x_1 jako x_{1Si} a pořadí prvků vzhledem ke složce x_2 jako x_{2Si} . Pro Spearmanův pořadový korelační koeficient pak platí:

$$r_S = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (x_{1Si} - x_{2Si})^2, \text{ kde} \quad (21)$$

r_S Spearmanův pořadový korelační koeficient

n počet měření

x_{1Si} a x_{2Si} viz výše.

V praxi se stává, že pro několik prvků vychází stejné pořadí. Pak se Spearmanův pořadový korelační koeficient počítá podle upraveného vztahu. Vztah je velmi složitý, zájemci viz [40].

1.3.3 Cronbachův korelační koeficient γ spolehlivosti výsledku

Představuje nejrozšířenější kritérium posouzení vnitřní jednotnosti výsledku a vypočte se podle vzorce

$$\gamma = \frac{m}{m-1} \left[1 - \frac{\sum_{i=1}^m \sigma_{ii}}{\sum_{i=1}^m \sum_{j=1}^m \sigma_{ij}} \right], \text{ kde}$$

(22)

m počet proměnných

σ_{ij} kovariance mezi proměnnou i a j

σ_{ii} rozptyl proměnné i

Jsou-li data předem standardizována (odečtením průměru a podělením směrodatnou odchylkou položky), dostaneme standardizovanou verzi Cronbachova koeficientu:

$$\gamma = \frac{m\bar{\rho}}{1 + \bar{\rho}(m-1)}, \text{ kde}$$

(23)

$\bar{\rho}$ průměr všech korelačních koeficientů mezi všemi m proměnnými.

2 Didaktické testy

2.1 Typy didaktických testů

(podle [35], není-li uvedeno jinak)

Pojem didaktický test je sice u různých autorů definován různě, ale tato různá vymezení se shodují v tom, že jde o cílené a standardizované šetření [48]. Standardizací testu zde Půlpán [48] rozumí přesnou formulaci požadavků jak obsahových, tak i realizačních a vyhodnocovacích. Test musí být obsahově správný, stručný a jednoznačně skórovatelný; zejména musí být validní a reliabilní. V pedagogické praxi se můžeme setkat s didaktickými testy různých druhů. V současné české pedagogice a psychologii se užívají např. klasifikace testů podle S. Vrány, M. Michaličky a P. Byčkovského [35]:

S. Vrána rozděluje testy takto (zpracováno podle [23]) :

1. Podle toho, co zjišťují:

- inteligentní, zaměřené na zjišťování schopností zkoumaných osob.
- didaktické, sloužící ke zjišťování vědomostí a dovedností, které žák získal učením.

2. Podle způsobu zpracování:

- informační, které sestavuje učitel, aby zjistil, jak žáci ovládli učivo.
- standardizované, sestavené podle přesně dodržovaných principů konstrukce, ověřené na reprezentativním vzorku, s předem stanoveným klíčem pro jejich hodnocení a normami sloužícími ke správné interpretaci výsledků.

3. Podle účelu:

- zkušební, které slouží k hodnocení vědomostí žáků.
- diagnostické, zaměřené na zjišťování kladů a nedostatků ve vědomostech žáků a na zjišťování příčin nedostatků.
- kontrolní, které si dává žák sám.

4. Podle způsobu užití:

- hromadné, kterými je prověřován větší počet žáků najednou.
- individuální, k ověřování vědomostí a dovedností jednoho žáka.

M. Michalička testy klasifikuje podle principů konstrukce na tři skupiny (podle [23], [35]):

- testy volných odpovědí (nestandardizované). Je to v podstatě zdokonalená forma tradičního písemného zkoušení, kdy žáci podle vlastního uvážení odpovídají na několik stanovených otázek či řeší několik úkolů.
- testy objektivně informační. Sestavuje je učitel, ředitel nebo jiný školský pracovník, a to již podle zásad pro přípravu testů standardizovaných. Mají zejména funkci kontrolní a klasifikační.
- testy standardizované, sestavené podle závazně stanovených a přísně dodržovaných konstrukčních principů. Zpravidla jde o týmovou práci odborníků. Jsou produktem vysokého stupně profesionální dovednosti a jsou opatřeny zpracovanými tabulkami norem.

P. Byčkovský [13] (tuto klasifikaci přejímá i Chráska [28]) třídí didaktické testy podle osmi hledisek:

1. Podle měření charakteristiky výkonu:

- testy rychlosti - jestliže didaktickým testem měříme rychlost, kterou je žák schopen provést určitý výkon.
- testy úrovně - pokud testem zjišťujeme kvalitu (úroveň) výkonu.

2. Podle stupně dokonalosti přípravy testu a jeho příslušenství:

- testy standardizované - jsou sestavovány zpravidla týmy odborníků, v nichž je vedle specialistů daného vyučovacího předmětu také statistik, psycholog, metodik apod. Standardizované didaktické testy jsou ověřovány a normovány na větších souborech či výběrech žáků. Umožňují vyjádřit výkon individuálně testované osoby ve vztahu k výkonům dané populace.

- b) testy nestandardizované (učitelské, neformální) - připravují si je učitelé sami pro svoji potřebu. Tyto testy sice musí vyhovovat základním pravidlům pro konstrukci přesných poznávacích nástrojů, ale často nevyhovují výkonnostním populačním normám a není u nich empiricky zjišťovaná validita.
 - c) Testy kvazistandardizované - tj. částečně standardizované nebo důkladněji provedené nestandardizované testy
- 3. Podle povahy činnosti testovaného:**
- a) test kognitivní - pokud didaktický test měří úroveň (kvalitu) poznávání žáků.
 - b) test psychomotorický - pokud testem zjišťujeme výsledky psychomotorického učení (např. psaní na stroji či počítači).
- 4. Podle míry specifičnosti učení zjišťovaného testem:**
- a) test výsledků výuky - měří to, co se žáci v dané oblasti naučili.
 - b) test studijních předpokladů - měří úroveň obecnějších charakteristik jedince, které jsou potřebné k dalšímu studiu (např. intelektovou výkonnost).
- 5. Podle toho, jakým způsobem interpretujeme výkon testovaného:**
- a) testy rozlišující - výkon jedince se určuje vzhledem k populaci testovaných.
 - b) testy ověřující - výkon se určuje vzhledem k populaci úloh, které učivo reprezentují.
- 6. Podle časového zařazení do výuky:**
- a) vstupní testy - jsou zadávány na začátku výuky určitého celku.
 - b) průběžné testy - zadávají se v průběhu výuky a jejich posláním je poskytovat učiteli informace potřebné k optimálnímu řízení výuky.
 - c) výstupní testy - zadávají se buď na konci výukového období, nebo na konci určitého celku a většinou poskytují informace potřebné pro hodnocení.
- 7. Podle rozsahu obsahového zaměření:**
- a) monotematické testy - prověřují jediné téma učební látky.
 - b) polytematické testy - prověřují učivo několika tematických celků.
- 8. Podle stupně objektivnosti skórování:**
- a) testy objektivně skórovatelné - obsahují úlohy, u nichž lze objektivně rozhodnout, zda byly řešeny správně či nikoliv.
 - b) testy subjektivně skórovatelné - u nichž není možno stanovit předpis pro skórování.

Jinou (účelovou) klasifikaci uvádí **K. Ingenkamp** (podle [22]):

1. Testy vývoje a školní zralosti
2. Postupy k posouzení počátečního čtení a psaní
3. Testy školních výkonů (pro různé vyučovací předměty)
4. Testy inteligence a nadání
5. Testy sociálně psychologické
6. Testy koncentrace
7. Testy a dotazníky o zájmech žáků

Účelovým tříděním je také rozlišování čtyř skupin testů, které použil pro ediční účely **národní podnik Psychodiagnostické a didaktické testy**:

- A Testy mohou používat jen odborníci psychologové
- B Testy může odborný psycholog svěřit jiným pracovníkům
- C Testy mohou používat i zaškolení nepsychologové
- D Testy jsou didaktické, které mohou používat všichni učitelé.

2.2 Charakteristiky didaktického testu

Didaktické testy musí být objektivní, standardní, přiměřeně obtížné, spolehlivé, platné (validní), kvantitativně i kvalitativně interpretovatelné, senzibilní (citlivé), úsporné (ekonomické) a administrované

všem zkoumaným osobám stejně.

2.2.1 Objektivnost

Objektivity (nezávislosti testu na jeho autorovi) didaktického textu dosáhneme, jestliže:

- na všechny položky existuje jednoznačná odpověď,
- na jednotlivé položky odpovídají nezávislí odborníci stejně,
- jsou dána jednoznačná kritéria skórování (a navazující klasifikace)

Lapitka [38] doporučuje provést zkoušku objektivnosti testu tak, že test předloží několika nezávislým odborníkům. Objektivnost pak kvantifikuje podle hodnoty výrazu:

$$O = \frac{\text{celkový poč. bodů získaných odborníky}}{\text{celkový poč. bodů za test} \times \text{poč. odborníků}} \cdot 100 \%$$

(24)

Lapitka se ve stati o objektivnosti testu odkazuje na Ebela [18], který ovšem nehovoří o objektivnosti testu, nýbrž o objektivnosti jednotlivých položek.

Osobně se domnívám, že při přípravě testu je skutečně vhodnější posuzovat přímo jednotlivé položky. Při aplikaci vztahu současně na celý test se totiž může stát, že test bude mít celkově velkou objektivnost, avšak zůstane v něm malé množství položek, v jejichž řešení se odborníci neshodnou (a kteréžto položky by z testu měly být vyřazeny), zatímco na ostatní odpoví zcela shodně – viz následující tabulky:

respondent		test A						test B					
		1	2	3	4	5	6	1	2	3	4	5	6
položka	1	0	0	0	0	0	0	1	0	1	1	1	1
	2	1	1	1	1	1	1	1	1	1	1	1	0
	3	1	1	1	1	1	1	0	1	1	1	1	1
	4	1	1	1	1	1	1	1	1	1	1	1	1
	5	1	1	1	1	1	1	1	1	1	1	1	1
	6	1	1	1	1	1	1	1	1	0	1	1	1
	7	1	1	1	1	1	1	1	1	1	1	1	1
	8	1	1	1	1	1	1	1	1	1	0	1	1
	9	1	1	1	1	1	1	1	1	1	1	0	1
	10	1	1	1	1	1	1	1	1	1	1	1	1
	celkem bodů	9	9	9	9	9	9	9	9	9	9	9	9
	nejvyšší možný počet bodů	10	10	10	10	10	10	10	10	10	10	10	10
							54						54
							60						60

Tab. 1: Příklad kontroly objektivnosti dvou testů.

V Tab. 1 je příklad kontroly objektivnosti dvou testů. Oba mají stejnou objektivnost vypočtenou podle vztahu (24), tj. 90 %, avšak v testu A jde o zřejmě nesprávně zadanou či nesprávně vyhodnocenou položku č. 1 (při 100% objektivnosti zbytku testu), zatímco v testu B je mírná neshoda je v mnoha položkách (může být vyvolána statisticky náhodnými faktory, např. únavou odborníků apod).

2.2.2 Standardnost

Standardností rozumíme požadavek, aby všichni respondenti měli k řešení stejných položek stejné podmínky a také, aby stejné položky byly následně u různých respondentů stejně vyhodnoceny. Dodržení těchto podmínek určuje také míru validity a spolehlivosti metody.

2.2.3 Přiměřená obtížnost

Obtížnost testu úzce souvisí s obtížností jednotlivých položek. O způsobu stanovení obtížnosti položek a o doporučených hodnotách obtížnosti položek pojednává kapitola 2.5.3 Vlastnosti testových položek.

2.2.4 Spolehlivost (reliabilita)

Spolehlivostí (reliabilitou) [48] didaktického testu rozumíme míru jeho stability, tj. schopnosti poskytovat stejné výsledky při opakování za stejných podmínek. Reliabilita je nutnou podmínkou validity. Reliabilitu kvantitativně vyjadřujeme různými způsoby, jejichž volba kromě jiného závisí i na konstrukci a druhu testu. Přehled způsobů kvantifikace reliability udává např. Lapitka [38], Chráška [28] nebo Turek [67].

2.2.4.1 Způsoby stanovení reliability

Spolehlivost lze zkoumat v podstatě třemi způsoby:

- posouzením shody mezi výsledky dosaženými při opakovaném testování stejným testem. U didaktických testů takto obvykle postupovat nelze, protože při krátkém rozestupu mezi testy si žáci mohou velkou část obsahu testu zapamatovat, zatímco dlouhý časový odstup mezi testy umožní změnu skutečných znalostí žáků...
- posouzením shody mezi výsledky dvou paralelních testů (testy obsahují různé položky, ale z téhož učiva). Sestavení paralelních forem testu je velmi obtížné a proto [57] tento postup nedoporučuje. Naopak tuto metodu doporučuje např. [38].

Lapitka v [38] navrhuje použít pro stanovení tzv. **indexu reliability** metodu dvojnásobného testu (dvojice ekvivalentních testů). Postup je založen na myšlence, že dva ekvivalentní testy jsou z hlediska reliability přijatelné, pokud většinu odpovídajících úloh odpoví respondent stejně (buď obě správně, nebo obě nesprávně):

$$I_{\text{rel}} = 1 - \frac{n_{\text{shod}}}{n_{\text{celk}}}, \text{ kde} \quad (25)$$

I_{rel} index reliability

n_{shod} počet shodně zodpovězených dvojic úloh (součet přes všechny položky a všechny respondenty)

n_{celk} počet všech dvojic úloh (součet přes všechny položky a všechny respondenty)

- posouzením shody mezi výsledky dosaženými v různých částech testu (doporučeno v [57]).

Reliabilita testu metodou půlení [28]:

Podmínkou pro použití tohoto způsobu výpočtu je, že test obsahuje sudý počet úloh a jednotlivé úlohy jsou seřazeny podle vzrůstající obtížnosti. Při výpočtu se postupuje tak, že se celý test rozdělí na dvě poloviny tím způsobem, že jednu polovinu tvoří úlohy s lichým pořadovým číslem (L) a druhou polovinu úlohy se sudým pořadovým číslem (S). Výsledky dosažené jednotlivými respondenty v obou polovinách testu se potom navzájem korelují. Z hodnoty vypočítaného korelačního koeficientu se pak vychází při stanovení koeficientu reliability. Samotný výpočet koeficientu reliability metodou půlení se provádí pomocí **Spearmanova-Brownova vzorce**

$$r_{\text{sb}} = \frac{2r_p}{1 + r_p}, \text{ kde} \quad (26)$$

r_{sb} koeficient reliability vypočtený pomocí Spearmanova-Brownova vzorce

r_p koeficient korelace mezi výsledky respondentů v obou polovinách testu

Hniličková [24] uvádí vztah (26) rozšířený pro k částí testu. Říká však, že pomocí Spearmanova-Brownova vzorce je možno vypočítat koeficient reliability jen tehdy, pokud všech k částí testu je rovnocenných s přibližně stejnou reliabilitou.

Půlpán [49] připomíná, že tento vztah platí jen tehdy, pokud rozptyl (směrodatná odchylka) v obou

polovinách testu je stejný. Pokud tomu tak není, vychází hodnota r_{sb} příliš velká.

Koeficient korelace pro dosažení do Spearmanova-Brownova vzorce) se vypočte podle vztahu (19). Jeho úprava na výpočet koeficientu reliability pomocí Spearmanova-Brownova vzorce je následující:

$$r_p = \frac{n \sum_{i=1}^n (x_{iL} x_{iS}) - \sum_{i=1}^n x_{iL} \sum_{i=1}^n x_{iS}}{\sqrt{\left[n \sum_{i=1}^n (x_{iL}^2) - \left(\sum_{i=1}^n x_{iL} \right)^2 \right] \cdot \left[n \sum_{i=1}^n (x_{iS}^2) - \left(\sum_{i=1}^n x_{iS} \right)^2 \right]}}, \text{ kde} \quad (27)$$

r_p koeficient korelace mezi výsledky respondentů v obou polovinách testu

n počet respondentů

i číslo respondenta

x_{iL} počet bodů získaných i-tým respondentem za liché položky

x_{iS} počet bodů získaných i-tým respondentem za sudé položky

V Řešátkovi [57] se uvádí, že za předpokladu přibližně normálního rozdělení testovacích výsledků závisí reliability testu na počtu položek, standardní odchylce a na mediánu. Slabinou postupu je, že předpokládané normální rozložení testovacích výsledků nemá v pedagogice zdůvodnění. Jeden ze způsobů výpočtu koeficientu reliability testu je podle [57] použití tzv. **Kuderova-Richardsonova vzorce***:

$$r = \frac{M_e(m - M_e)}{ms^2}, \text{ kde} \quad (28)$$

r koeficient reliability

M_e medián

m počet položek

s standardní odchylka

V [28] se uvádí výpočet koeficientu reliability pomocí **Kuderova-Richardsonova vzorce 20 (tzv. KR 20)**

$$k_{kr} = \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m p_j(1-p_j)}{s^2} \right), \text{ kde} \quad (29)$$

k_{kr} koeficient reliability vypočtený pomocí Kuderova-Richardsonova vzorce

m počet položek

p_j index obtížnosti j-té položky (vyjádřený jako desetinné číslo, nikoli v procentech).

s směrodatná odchylka pro celkové výsledky (hrubé skóre) žáků v testu

s^2 rozptyl (variance)

V [28] i [38] se současně poznamenává, že tento způsob výpočtu se hodí pouze pro testy složené z obsahově homogenních úloh.

Půlpán [49] uvádí zjednodušení KR 20 na **KR 21** pro binárně skórovaný test tvořený položkami přibližně stejné obtížnosti:

* Je nápadná shoda názvu se vztahem pro koeficient reliability počítaný podle Kuderova-Richardsonova vzorce. Výpočetní vzorec je však úplně jiný.

$$k_{kr} = \frac{n}{n-1} \left(1 - \frac{\overline{X} - \overline{X}^2}{s^2} \right), \text{ kde} \quad (30)$$

\overline{X} průměrné skóre určité skupiny respondentů; ostatní symboly viz výše.

Např. Turek [67] uvádí vztah pro výpočet koeficientu reliability Cronbachovou metodou. Vztah je určen pro výpočet koeficientu reliability u testů s položkami skórovanými váženě (složeně) – viz kap. 2.5.1.

$$r_c = \frac{m}{m-1} \left(1 - \frac{\sum_{j=1}^m s_j^2}{s^2} \right), \text{ kde} \quad (31)$$

r_c koeficient reliability vypočtený pomocí Cronbachovy metody

m počet položek

s^2 rozptyl (variance) celého testu

s_j^2 rozptyl (variance) j-té položky

2.2.4.2 Doporučená hodnota reliability

Testy, které mají vysokou reliabilitu, jsou málo ovlivněny náhodnými vlivy (např. momentální psychosomatickou kondicí, vnějšími podmínkami). Hodnota koeficientu reliability se pohybuje mezi $0 \leq r \leq 1$. Hodnota koeficientu reliability podle [56] má být u dobrých testů $r \geq 0,9$. Podle [27] se pro individuální pedagogickou diagnostiku většinou požaduje koeficient reliability minimálně 0,80. Podle [34] je však u nestandardizovaných testů za postačující považována hodnota koeficientu reliability 0,6, v [57] se za uspokojivou hodnotu pokládá koeficient korelace mezi oběma polovinami testu o hodnotě alespoň 0,8. V [37] se doporučuje, aby učitel nepoužíval test s indexem reliability nižším než 0,75. Pokud má dvojnásobný test reliabilitu vyšší než 0,85, lze jej podle [37] rozdělit na dva testy (aby respondenti sedící vedle sebe neřešili tytéž položky).

Obecně platí, že s rostoucím počtem položek reliabilita testu roste. Test s nízkou reliabilitou je tedy možno (v určitém rozmezí) upravit přidáním položek tak, že jeho reliabilita tím vzroste. Potřebný počet položek se dá odhadnout pomocí vztahu uvedeného v [56]. Pokud test má $r \leq 0,6$, je koeficient reliability podle [56] tak nízký, že nemá smysl se zabývat dalším zlepšováním testu.

$$k = \frac{r_p(1-r_d)}{r_d(1-r_p)}, \text{ kde} \quad (32)$$

k koeficient, kterým je nutno násobit stávající počet položek

r_p požadovaný koeficient reliability

r_d dosažený stávající koeficient reliability

2.2.5 Validita

Validita je nejdůležitější ukazatel kvality didaktického testu. Validitou (platností, adekvátností) rozumíme, nakolik měří to, co měřit má.

Jak uvádí Lapitka [37] s odkazem na Ebela [17], v odborných příručkách se uvádí řada různých druhů

validity (Ebel v citované práci prý uvádí až 10 druhů validity, které klasifikuje do dvou kategorií: **přímá validita** a **odvozená validita**). Turek [66] uvádí např.:

Obsahová validita (typ validity přímé) – většinou se jí rozumí shoda mezi obsahem testu a učební látkou (u testů studijních výsledků)

Kritériová validita – ukazuje míru shody mezi didaktickým testem a nějakým jiným kritériem úspěšnosti.

Zpravidla se odlišují dva typy kritériové validity:

predikční validita (schopnost testu „předpovídat“ míru úspěšnosti žáka při dalším studiu) – u testů studijních předpokladů

souběžná validita (typ validity odvozené) – jde o vztah mezi testem a jiným ukazatelem žákových vědomostí. Za tento srovnávací ukazatel se často přijímá klasifikace (známka).

Pojmová (konstruktová) validita – vyjadřuje rozsah, v němž test měří určitou charakteristiku nebo psychologický konstrukt.

Chráška [27] na rozdíl od Lapitky [37] uvádí: Posouzení stupně validity testu se v praxi většinou ponechává na příslušném odborníkovi (ještě lépe na skupině odborníků). Jen výjimečně je možno posoudit validitu nově vytvořeného testu tak, že se srovnávají v něm dosažené výsledky s výsledky nějakého jiného testu, jehož validita je nesporná.

2.2.6 Kvalitativní a kvantitativní interpretovatelnost a normalizace

Kvalitativní analýza např. zjišťuje, které položky testu byly nejobtížnější, kterých chyb se respondenti nejčastěji dopouštěli a podobně.

Možností kvantifikace rozumíme, že registrační technika dovoluje rozlišování různých stupňů sledované vlastnosti, nejen pouhou klasifikaci na přítomnost nebo nepřítomnost jevu. Teprve tehdy je metoda poměrně exaktním měřením. Měření dovoluje jemnější diferenciaci, než je běžná klasifikace, a usnadňuje srovnávání výsledků. Statistickými početními operacemi s čísly lze získat také další informace o pozorovaných jevech. Jedním z výstupů může být i klasifikace (přiřazení známek).

2.2.7 Senzibilita (citlivost)

(podle Hniličkové [22])

Senzibilita (citlivost či selektivnost) didaktického testu charakterizuje, nakolik test umožňuje zjistit rozdíly v rozsahu a kvalitě vědomostí či dovedností.

Teoretickým modelem na rozdělení psychických znaků osobnosti bývá většinou tzv. normální rozdělení graficky zobrazené Gaussovou křivkou – viz

Obr. 1, přičemž citlivějším testům přísluší plošší křivky (větší směrodatná odchylka). Jak však připomíná Lapitka [37], pedagogické jevy nejsou náhodné a proto nejsou distribuované normálně. Proto ve skutečnosti nelze nic předpokládat o tom, jaké bude rozdělení respondentů v dané (zejména malé) skupině z hlediska úspěšnosti při řešení testu.

Má-li test dobře rozlišovat mezi jednotlivými respondenty, musí hrubé skóre jednotlivých respondentů kolísat v dostatečně širokém intervalu. Nejjednodušší mírou tohoto kolísání je tzv. **rozpětí** (σ), což je rozdíl mezi největší a nejmenší hodnotou sledované náhodné veličiny (v případě testů je touto veličinou hrubé skóre). Citlivost testu častěji zjišťujeme podle rozložení (distribuce) skóre a podle rozptylu neboli variance

(σ^2) nebo variančního koeficientu

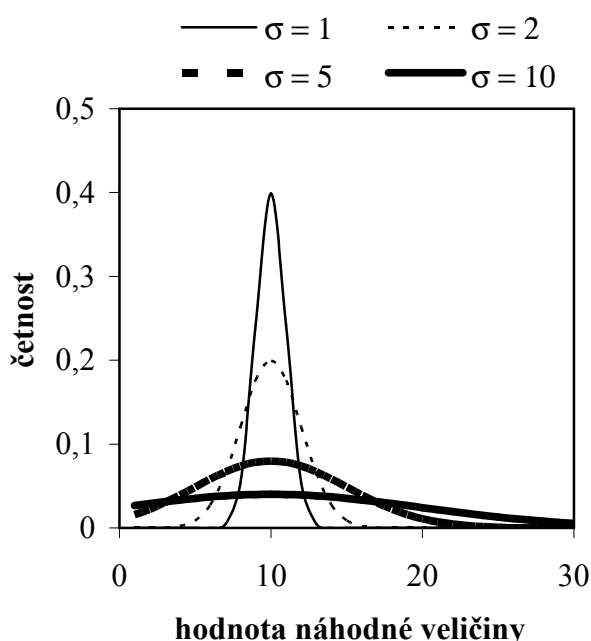
(σ^2 / \bar{x}^2): citlivějšímu testu odpovídá vyšší hodnota variančního koeficientu.

Souhrn četností výskytu jednotlivých hodnot (resp. hodnot v určitých intervalech) náhodné veličiny se nazývá **rozložení** (rozdělení) četností. Pro některé náhodné veličiny se dají z teoretických úvah počtu pravděpodobnosti odvodit určité zákony jejich rozložení. Např. náhodná veličina vznikající jako součet velkého počtu téměř nezávislých sčítanců má zpravidla alespoň přibližně rozložení Gaussovo (tzv. **normální rozložení**). Pokud jde o didaktické testy, je třeba mít na paměti, že didaktické testy měří výsledky výuky (což je cílená činnost), takže skóre testu není veličina zcela náhodná (ovlivnění

vyučujícím, materiálními i jinými podmínkami výuky,...). Také se jen málokdy vyskytuje homogenní test obsahující stejně obtížné položky. Přesto se však při statistickém vyhodnocování didaktických testů normalnost rozložení často předpokládá. O tom, zda naměřené hodnoty předpokladu normálního rozložení vyhovují, je možno na zvolené hladině pravděpodobnosti rozhodnout pomocí **testu kombinace výběrové šikmosti a špičatosti dle Jarque-Berra** nebo pomocí **Pearsonova testu** neboli **testu χ^2 dobré shody**. Gaussovo (normální) rozložení je dokonale popsáno střední hodnotou (průměrem) a rozptylem. Grafickým znázorněním Gaussova rozložení je křivka symetrická kolem svislé osy procházející průměrem \bar{X} a ve vzdálenosti větší než tři standardní odchylky od průměru se velmi těsně přimyká k vodorovné ose. Příklad různých možností Gaussova rozdělení ukazuje

Obr. 1. Hodnoty byly vypočteny pomocí programu Excel, funkce NORMDIST(x, \bar{X} , σ , 0).

**Gaussovo (normální) rozdělení
o střední hodnotě $\mu = 10$
a směrodatné odchylce σ**



Obr. 1: Příklady Gaussova rozložení pro různé hodnoty směrodatné odchylky σ a pro střední hodnotu $\mu = 10$.

Dá se dokázat, že při opakovaných nezávislých pozorováních náhodné veličiny U jen zhruba v 5 % případů překročí pozorovaná hodnota kritickou hodnotu $u_{2,5} = 1,96$. Z toho plyne, že jen asi v 5 % případů se náhodná veličina liší polovina je větší, polovina je menší) od své průměrné hodnoty o více než 2 standardní odchylky ($1,96 \doteq 2$).

2.2.8 Úspornost

Úspornost (resp. ekonomičnost) didaktického testu znamená, že test má být pokud možno časově nenáročný, nenáročný na vyhodnocování a také nenáročný na finanční náklady.

2.2.9 Administrace didaktického testu

(podle Hnilíčkové [23])

Administrace (zejména standardizovaných) testů se týká technických podmínek postupu při testování, vyhodnocování a interpretaci výsledků tsování. Veškeré informace o testu by měly být shromážděny v testové příručce (imatrikulačním listu testu). Zejména by zde měly být uvedeny:

- informace pro volbu testu
 - o cíl a účel, ke kterému byl test sestaven
 - o cílová skupina respondentů (věk, typ školy apod.)

- informace o způsobu standardizace
 - o počet pokusných osob
 - o charakteristika pokusných osob
- informace o vlastnostech testu aplikovaného na danou cílovou skupinu respondentů
 - o druh rozdělení výsledků
 - o průměr skóre
 - o rozptyl a standardní odchylka skóre
 - o reliabilita
- pokyny o využívání testu
 - o všichni respondenti musí dostat stejnou ústní i písemnou instrukci (způsob zaznamenávání správné odpovědi, počet správných odpovědí v polytomických položkách, způsob skórování, způsob řazení položek a strategie řešení testu, zda po vyřešení testu mohou odevzdat dříve, kam si respondenti mohou psát poznámky, jaké mají dovolené pomůcky – např. kalkulačka, tabulky apod.)
 - o stejné znění didaktického testu
 - o stejný záznamový (odpovědní) list.
 - o musí mít také v rámci možností stejné podmínky k vypracování testu (hlučnost místnosti, přiměřené osvětlení, přiměřená teplota, větrání, denní doba,...).
 - o zařazení testu s ohledem na probírané učivo,
 - o denní doba, kdy má být provedeno testování
 - o zda respondenti o testu, případně o jeho obsahu dopředu mají, či nemají vědět
 - o délka trvání testu
 - o kdy lze odpovídat respondentům na dotazy,
 - o
- informace pro vyhodnocení testu
 - o autorské řešení
 - o způsob hodnocení jednotlivých položek
 - o způsob hodnocení celého testu (norma hodnocení, případně klasifikace)

2.3 Základní kroky tvorby didaktického testu

2.3.1 Příprava didaktických testů

- specifikace cílů
- tvorba položek, pro získání distraktorů lze použít více volných odpovědí
- kontrola položek nezávislými odborníky
- administrace pretestu
- vyhodnocení pretestu, úprava znění položek, zařazení distraktorů, úprava počtu položek
- kontrola nové verze nezávislými odborníky
- nová administrace testu
- korektury testu
- administrace testu velkému počtu respondentů, standardizace

2.3.2 Tvorba učitelského ověřovacího didaktického testu

- Vymezení účelu DT
- Vymezení obsahu DT: zde lze využít techniky specifikační tabulky (uvádí jednotky učiva, rozsah každé jednotky učiva (počet hodin výuky nebo počet stran v učebnici), požadovaná úroveň osvojení učiva); já v mé práci specifikační tabulku nepoužívala. Důvodem bylo, že cílem bylo testovat učivo fyzikálníchemie. tento předmět však na základní škole samostatně neexistuje, žáci se s fyzikálně chemickým učivem setkávají nahodile při výkladu jiného učiva. Testové položky byly vybrány tak, že podle nejobvyklejších učebnic chemie pro základní školy [1], [3], [4], [5] byly vybrány pokud možno všechny úlohy pro žáky, týkající se učiva fyzikální chemie, a jejich počet pak byl zredukován tak, aby z každé oblasti učiva zůstala v testu alespoň jedna položka.

- Vymezení učebních cílů a úloh – učební cíle viz kapitola 2.4.3
- Návrh testových úloh – vzniknou transformací učebních cílů a úloh
- Posouzení testových úloh (posoudí nezávislí odborníci):
 - soulad testové úlohy a učebního cíle
 - správnost „správné odpovědi“
 - technická kvalita úlohy – viz kapitola 2.4.2
 - důležitost učiva testovaného úlohou
 - náročnost úlohy
- Použití testu
- Analýza testu

2.4 Testové položky

2.4.1 Typy testových položek podle formy

Didaktický test je sestaven z jednotlivých testových úloh (položek). Položky (úkoly) mohou být různým způsobem formulované, přičemž index obtížnosti položky souvisí s formulací (typem) položky. Klasifikací položek podle formální stránky se zabývali např. S. Vrána nebo M. Michalička:

S. Vrána rozlišil šest základních forem položek:

1. Forma jednoslovné odpovědi.
2. Forma doplňovací (vyžaduje se doplnění neúplné věty).
3. Forma správnosti a nesprávnosti (vyžaduje se např. označení správnosti a nesprávnosti tvrzení).
4. Forma několikeré volby (vybírání se správný údaj z několika nabízených možností).
5. Forma vzájemných vztahů (trídí se „co k čemu náleží“).
6. Forma roztříd'ovací, (vyžaduje se roztřídění daných údajů podle určitého kritéria).

M. Michalička uvádí pět základních typů testových položek:

1. Produkční položky (vyžaduje se vytvoření krátké či delší odpovědi. Jde o tzv. otevřené úkoly).
2. Doplňovací položky (vyžaduje se doplnění neúplné věty (výroku) tak, aby dávala smysl).
3. Alternativní položky (vybírání se správná odpověď ze dvou nabízených možností).
4. Položky s výběrem odpovědi z více možností.
5. Položky tzv. oboustranného výběru, (k určitému souboru výrazů, názvů apod. se přiřazuje správná **odpověď z nabízených variant**).

P. Byčkovský [12] a podle něj **M. Chráska** [27] dělí položky na:

1. **Otevřené úlohy** (úlohy s tvořenou odpovědí, úlohy s volnou odpovědí)
 - a) otevřené široké úlohy očekává se pojednání na určité téma, vyřešení zadaného problému, popis určitého procesu,...
 - b) úlohy se stručnou odpovědí (očekává se doplnění značky, symbolu nebo i několika slov).
2. **Uzavřené úlohy** (úlohy s nabízenou odpovědí, úlohy s nucenou volbou odpovědi).
 - a) úlohy dichotomické (true-false item) – respondent vybírá správnou možnost ze dvou navržených. Jejich nevýhodou je velká pravděpodobnost uhodnutí správné odpovědi.
 - b) úlohy polytomické (s výběrem odpovědi, multiple-choice). Pravděpodobnost náhodného uhodnutí správné odpovědi klesá s rostoucím počtem nabízených odpovědí. Jako jejich optimální počet se uvádí 4-5. Praxe se ustálila u čtyř nabízených odpovědí. Pravděpodobnost uhodnutí správné odpovědi lze v některých případech snížit užitím tzv. neurčité odpovědi (správná odpověď v nabídce není uvedena). Nabízené odpovědi mohou mít několik forem:
 - vybírá se jedna odpověď
 - vybírá se několik správných odpovědí. Tato forma komplikuje binární skórování výsledků a někdy odrazuje respondenty. Respondenti musejí být na tento typ otázek předem upozorněni.

Nad uvedené základní dva typy položek (**otevřené**, **uzavřené**) uvádějí někteří autoři (pro dotazníky např. Lapitka [36]) také položky **polootevřené**. Takové položky umožňují, aby si respondent volil z připravených variant nebo odpovídal volně, mimo naznačené možnosti.

Byčkovský [12] a Chráska [27] považují za **úlohy objektivního typu** takové úlohy, které lze objektivně skórovat, tzn. úlohy otevřené se stručnou odpovědí a úlohy uzavřené (viz výše).

V Lapitkovi [37] se testy podle typu položek dělí na:

- **homogenní**: Celá látka se nejprve zpracuje na soubor jednoduchých, rovnocenných faktografických otázek, příkladů nebo úloh, z nichž se pak statisticky náhodně vyberou testové položky.
- **homomorfní**: Takový test vzniká záměrným reprezentativním výběrem prvků z obsahu nějakého tematického celku, zachovává základní strukturní vztahy mezi prvky a jejich významovou hierarchii. Didaktické testy používané na školách jsou většinou tohoto typu. Homomorfní test nevychází z koncepcí platných v psychodiagnostice, což se odráží v platnosti výpočtu testových charakteristik: validity, reliability, senzibility (diskriminativnosti).

G. Rötling (Metodické centrum Banská Bystrica) [54] dělí typy testových položek obdobně jako M. Chráska [27]:

- **Otevřené** (odpověď je respondentem tvořena), tzv. produkční
 - subjektivně skórovatelné (široká odpověď s vymezenou strukturou): *Vysvětlíte, proč...*
 - objektivně skórovatelné → stručná odpověď, krátké řešení:
 - Napište stavovou rovnici ideálního plynu...*
 - Napište vzorec síranu draselného...*
 - doplňovací úloha
 - Doplňte pravou stranu rovnice: $FeS_2 + O_2 \rightarrow \dots$*
- **Uzavřené** (respondent odpověď volí z nabídky)
 - objektivně skórovatelné → alternativní výběr
 - položky dichotomické
 - položky s výběrem jedné správné odpovědi
 - položky s výběrem nejlepší (nejvhodnější) odpovědi
 - položky s výběrem odpovědi s negací v zadání úlohy
 - položky s vícenásobnou správnou odpovědí
 - dvojstupňové výběrové úlohy
 1. stupeň = výběr tvrzení,
 2. stupeň = výběr zdůvodnění
 - přiřazovací výběr
 - K veličinám v levém sloupci přiřad'te jejich jednotky (uvedeny v pravém sloupci)...*
 - výběr uspořádání
 - Do uvedených výrazů dosad'te $a = -2$ a seřad'te je podle jejich číselných hodnot tak, aby výraz s největší číselnou hodnotou byl první...*

2.4.2 Zásady pro navrhování testových úloh

2.4.2.1 Obecné zásady pro navrhování všech druhů testových úloh

(podle Mužiće [41], Byčkovského [12], Chrásky [27] a Turka [66]):

- Úlohou testovat důležité učivo, které bylo předmětem výuky.
- Text úkolů musí být gramaticky správný.
- Věnovat dostatečnou pozornost i grafické stránce úlohy.
- V textu se nemají vyskytovat nejasná slova nebo dvojsmyslnost.

- Je třeba vyhýbat se úkolům, jejichž řešení je dosud předmětem vědecké diskuze.
- Úlohy nechat vždy posoudit kompetentními osobami.
- Vyhýbat se doslovným formulacím z učebnic (podporovalo by to formalismus v získávání vědomostí).
- Úkoly nesmějí být sestaveny tak, aby bylo možno dospět ke správné odpovědi na základě logického úsudku o tom, jak byl úkol formulován nebo uspořádán.
- Řešení jednoho úkolu nesmí umožňovat nebo podmiňovat řešení některého z dalších úkolů.
- Zásadně nepoužívat tzv. „chytáky“ – nezkouší stupeň zvládnutí učiva, ale např. postřeh, vtip atd.).
- Je třeba používat spíše kvantitativních než kvalitativních výrazů.
- Vyhýbat se zbytečným zdrojům obtížnosti (např. zbytečně nezadávat složité číselné hodnoty)
- Aby mohli být žáci navzájem srovnáváni, je nezbytně nutné, aby všichni řešili tytéž úlohy.
- Používat pokud možno jednoduchého (binárního) skórování úloh.
- Navrhovat úlohy v dostatečném počtu (pro tzv. předvýzkum je vhodné navrhovat úloh o něco více, než kolik má obsahovat konečná podoba testu).

Hališka [22] tato pravidla doplňuje:

- střídání různých typů položek zvyšuje provokační sílu testu a tím snižuje únavu respondentů,
- do tvorby testu je možno přiměřeným způsobem zapojit i žáky (motivující forma fixace i aktivizace samostatné tvůrčí práce).

2.4.2.2 Návrh otevřených úloh se stručnou odpovědí

(podle Chrásky [27] a Byčkovského [12])

- 1) Úlohu užívejte jen tehdy, lze-li odpovědět velmi stručně (nejlépe jen jedním údajem).
- 2) Úlohu formulujte zcela jasně a jednoznačně.
- 3) Nevyžadujte doslovné opakování textu z učebnice.
- 4) Uvažte předem všechny možné odpovědi, a je-li jich mnoho, raději úlohu nepoužívejte.
- 5) Ponechte v úlohách vždy dostatek místa pro uvedení odpovědi.
- 6) Dávejte přednost produkčním úlohám před doplňovacími. Chcete-li přece jen použít doplňovací úlohy, dodržujte následující doporučení:
 - Vynechávejte jen důležité údaje.
 - Z neúplné věty musí být patrné, co se má doplnit.
 - Údaj, který se má doplnit, umíst'ujte pokud možno na konec věty.
 - Pokud se má doplnit několik údajů, vynechejte pro doplnění zhruba stejné místo.

Konstrukce testových úloh s výběrem odpovědí

(Podle P. Byčkovského [12], Chrásky [27] a Rötlinga [54])

Sestavování úloh s výběrem odpovědí je obtížnější než příprava úloh produkčních. Zatímco produkční úlohy vyžadují pouze správnou formulaci problému nebo otázky, úlohy s výběrem odpovědí vyžadují navíc vytvoření vhodných nabídek špatných odpovědí (tzv. distraktorů), které by byly pro respondenty stejně přijatelné.

Návrh vhodných distraktorů je velkým problémem. Při navrhování distraktorů většinou vycházíme z logické úvahy anebo ze zkušeností s nejčastěji se vyskytujícími chybami. U pečlivě připravovaných testů se někdy postupuje tak, že se úloha nejdříve zadá respondentům jako otevřená a potom se nejčastěji vyskytující chyb použije jako distraktorů. U testové úlohy, kde distraktory plní svoji funkci, by mělo platit, že respondent (který správnou odpověď nezná) vybírá ze všech předložených nabídek pokud možno náhodně. Přijatelnost distraktorů je však vlastnost značně relativní. Tentýž distraktor se může jevit respondentovi s nižší úrovní vědomostí dostatečně atraktivní, zatímco respondentovi s vyšší úrovní vědomostí jako zcela nepřijatelný.

Zásady pro návrh úloh s výběrem odpovědí

- 1) Úlohami s výběrem odpovědí nezkoušíme pokud možno zapamatování konkrétních poznatků. K tomu se

lépe hodí úlohy produkční.

- 2) Nejvhodnější počet nabídek je 4-5, praxe se ustálila na čtyřech možnostech.
- 3) Soubor nabídek má být homogenní.
- 4) Při formulaci úloh s výběrem odpovědi dáváme přednost otázkám před neúplnými tvrzeními.
- 5) Ve formulaci úlohy se vyhýbáme slovům nebo údajům, které by mohly sloužit jako nápověda.
- 6) Pokud se ve formulaci úlohy vyskytuje zápor, zvýrazníme jej např. podtržením.
- 7) V úloze s výběrem nejlepší odpovědi je nutno tento fakt zdůraznit vhodnou formulací.
- 8) Soubor nabízených odpovědí k jedné úloze by měl být homogenní, tj. podobný obsahovým zaměřením i formou.
- 9) Distraktory se nesmějí navzájem překrývat nebo jinou formou vyjadřovat totéž.
- 10) Distraktory musí být jednoznačně nesprávné nebo musí obsahovat nepostačující odpověď.
- 11) Umístění správné odpovědi mezi distraktory se má volit zcela náhodně.
- 12) Navrhujeme jen takové distraktory, u nichž je předpoklad, že budou využívány.
- 13) Při používání úloh s vícenásobnou volbou odpovědi a při používání neurčitých odpovědi na tuto skutečnost žáky upozorníme.
- 14) Vyhýbáme se příliš dlouhým slovním formulacím. Často lze nahradit dlouhý text vhodným obrázkem, náčrtem nebo grafem.

Nabídky *nic z uvedeného, všechno, nevím* by se podle Rötlinga [54] neměly používat. Odpověď *nic z uvedeného* je však některými jinými autory naopak doporučována.

Problém hádání správných odpovědí u úloh s výběrem odpovědi

(Podle Chrásky [27])

U uzavřených úloh existuje vždy určitá pravděpodobnost, že žák zvolí správnou odpověď zcela náhodně. Toto nebezpečí se zmenšuje s rostoucím počtem nabízených odpovědí. Jako optimální počet předkládaných odpovědí se uvádí 4-5, praxe se ustálila na čtyřech odpovědích. Menší počet odpovědí než čtyři se pro velkou pravděpodobnost uhodnutí správné odpovědi nedoporučuje, více než 5 odpovědí činí zase úlohu nepřehlednou a také sestavování většího počtu přijatelných odpovědí je obtížné.

Pro dvě a tři nabízené odpovědi se někdy doporučuje tzv. korekce na hádání. Při používání korekce na hádání se přisoudí žákovi počet bodů podle toho, kolika chyb se v testu dopustil. Vychází se z toho, že žák, který odpověď hádá, se dopouští častěji chyb než ten, který úlohy skutečně řeší a odpovídá jediné tehdy, když odpověď zná. Korekci dosažených bodových výsledků lze provést podle vzorce

$$s_o = s_n - \frac{n}{y-1}, \text{ kde}$$

(33)

s_o opravené skóre (opravený počet bodů),

s_n neopravené skóre,

n počet nesprávných odpovědí daného žáka v testu,

y počet nabízených odpovědí v jedné úloze.

Jestliže se provádí korekce na hádání, je nutno na tuto skutečnost žáky v úvodní informaci upozornit. Je třeba jim vysvětlit, že ve sporných případech je pro ně výhodnější neodpovídat vůbec. Naopak v testech, kde se korekce na hádání neprovádí, je pro zkoušeného výhodnější zodpovědět všechny úlohy. Zkušenosti však ukazují, že při korekci na hádání dochází např. k poškozování žáků velmi kritických ke své práci.

U některých testových úloh s výběrem odpovědi lze omezit hádání správné odpovědi tím, že požadujeme jako podmínku pro uznání odpovědi výpočet nebo zdůvodnění odpovědi. Pravděpodobnost uhodnutí správné odpovědi lze také v některých případech snížit užitím tzv. neurčité odpovědi.

2.4.2.3 Některé zásady pro návrh přiřazovacích úloh

(podle Rötlinga [54]):

- doplňky formulovat výstižně a stručně,
- navrhovat homogenní soubor nabídek a doplňků,
- nemá se nabízet více než 8-10 nabídek a doplňků,

- nabídek musí být více než doplňků (aspoň o jednu); např. Hališka [22] ovšem o takovém požadavku nehovoří,
- doplňky a nabídky uvádět v oddělených sloupcích,
- doplňky číslovat, nabídky značit písmeny.

2.4.2.4 Některé zásady tvorby úloh s výběrem uspořádání

(podle Rötlinga [54]):

- nabídky musí být jednoznačně uspořadatelné
- navrhovat nanejvýš 6-8 nabídek
- v zadání úlohy je nutno vymežit kritérium pro uspořádání
- podle možností doplňky číslovat, nabídky značit písmeny.
- soubor nabídek i doplňků musí být stejně velký.

2.4.3 Typy položek podle náročnosti myšlenkových operací

Jednotlivé testové položky je možné členit ve shodě s **Bloomovou taxonomií** učebních úloh (citace původní práce se v literatuře neshodují, viz např. dále v této kapitole), tedy podle náročnosti myšlenkových operací, které budou od respondenta při jejich řešení vyžadovány. Bloomova taxonomie uvedená v [37] s odkazem na [8]:

- Úlohy vyžadující pouze znalosti, pamětní reprodukci poznatků.
- Úlohy vyžadující navíc porozumění poznatkům.
- Úlohy vyžadující dovednost aplikovat osvojené poznatky.
- Úlohy vyžadující dovednost syntézy poznatků.
- Úlohy vyžadující dovednost hodnotícího posouzení poznatků.

V práci [20] se uvádí:

Snad prvním krokem při formulaci jasných cílů výchovy je systematická analýza různých typů vědomostí a dovedností, jimž lze vyučovat a kterým by se mělo vyučovat. Významným příspěvkem v této oblasti byla práce examinatorů kolejí a univerzit v čele s Benjaminem S. Bloomem. Bloomův výbor nejprve rozdělil výchovné cíle do tří kategorií:

- 1) poznávací oblast
- 2) afektivní oblast
- 3) psychomotorické dovednosti

Po důkladném studiu vypracoval výbor podrobný přehled cílů spadajících do poznávací oblasti. Studie o poznávací oblasti byla publikována v [7]. Poznávací cíle se podle citované práce dělí na (převzato z [20]):
Přehled taxonomie výchovných cílů v poznávací oblasti **podle Blooma**:

- 1.00 Znalosti
- 2.00 Chápání
- 3.00 Použití
- 4.00 Analýza
- 5.00 Syntéza
- 6.00 Hodnocení

V rámci jednotlivých kategorií existuje ještě jemnější členění.

D. Tollingerová navrhla podle Bloomova třídění vyučovacích cílů taxonomii učebních úloh, rozdělených podle jejich operační struktury, tzn. podle operací nutných k jejich řešení.

Učební úlohy jsou podle Tollingerové rozděleny do 5 kategorií, obsahujících 27 typů učebních úloh:

Taxonomie učebních úloh podle D. Tollingerové

(převzato z [64], str. 197).

- 1.0 Úlohy vyžadující pamětní reprodukci poznatků
 - 1.1 úlohy na znovupoznání .
 - 1.2 úlohy na reprodukci jednotlivých faktů, čísel, pojmů
 - 1.3 úlohy na reprodukci definic, norem, pravidel
 - 1.4 úlohy na reprodukci textových celků, básní, tabulek aj.
- 2.0 Úlohy vyžadující jednoduché myšlenkové operace s poznatkami
 - 2.1 úlohy na zjišťování faktů (měření, jednoduché výpočty aj.)
 - 2.2 úlohy na vyjmenování a popis faktů (výčet, soupis aj.)
 - 2.3 úlohy na vyjmenování a popis procesů, způsobů činnosti aj.
 - 2.4 úlohy na rozbor a skladbu (analýza, syntéza)
 - 2.5 úlohy na porovnávání a rozlišování (komparace, diskriminace)
 - 2.6 úlohy na třídění (kategorizace, klasifikace)
 - 2.7 úlohy na zjišťování vztahů (příčina, následek, cíl, prostředek, vliv, funkce, nástroj, způsob aj.)
 - 2.8 úlohy na abstrakci, konkretizaci a zobecňování
 - 2.9 řešení jednoduchých příkladů (s neznámými veličinami).
- 3.0 Úlohy vyžadující složité myšlenkové operace s poznatkami
 - 3.1 úlohy na překlad (translaci, transformaci)
 - 3.2 úlohy na výklad (interpretaci), vysvětlení smyslu a významu, zdůvodnění
 - 3.3 úlohy na vyvozování (indukci)
 - 3.4 úlohy odvozování (dedukci)
 - 3.5 úlohy na dokazování (argumentaci) a ověřování (verifikaci)
 - 3.6 úlohy na hodnocení
- 4.0 Úlohy vyžadující sdělení poznatků
 - 4.1 úlohy na vypracování přehledu, výtahu, obsahu apod.
 - 4.2 úlohy na vypracování zprávy, pojednání, referátu aj.
 - 4.3 samostatná písemná práce, výkresy, projekty apod.
- 5.0 Úlohy vyžadující produktivní myšlení
 - 5.1 úlohy na praktickou aplikaci
 - 5.2 řešení problémových úloh a situací
 - 5.3 kladení otázek a formulace úloh nebo zadání
 - 5.4 úlohy na objevování na základě vlastního pozorování (na senzorické bázi)
 - 5.5 úlohy na objevování na základě vlastních úvah (na racionální bázi)

Kromě uvedených taxonomií se pracuje také s **Niemierkovou taxonomií výukových cílů**, kterou Chráska [27] pokládá za zvláště výhodnou pro plánování didaktických testů (následující třídění je převzato z [27], str. 21):

Niemierkova taxonomie výukových cílů

A Zapamatování poznatků

Žák schopen vybavit si určitá fakta (např. termíny, zákony), přičemž je nesmí mezi sebou zaměňovat.

Typická aktivní slovesa: definovat, napsat, opakovat, pojmenovat, reprodukovat.

B Porozumění poznatkům

Žák je schopen poznatky předložit i v jiné formě než v té, ve které si je zapamatoval, dovede poznatky uspořádat nebo zestručnit.

Typická aktivní slovesa: jinak formulovat, ilustrovat, objasnit, odhadnout, přeložit, převést, vyjádřit vlastními slovy.

C Používání vědomostí v typových situacích (tzv. specifický transfer)

Žák dovede použít vědomostí k řešení situací, které ve výuce již byly řešeny.

Typická aktivní slovesa: aplikovat, použít, prokázat, řešit, diskutovat, načrtnout, vyzkoušet, registrovat, demonstrovat.

D Používání vědomostí v problémových situacích (tzv. nespecifický transfer)

Žák dovede použít vědomostí k řešení problémových situací, které nebyly ve výuce doposud řešeny.

Typická aktivní slovesa: rozhodnout, provést rozbor, kombinovat, vyvrátit, obhájit, prověřit, zhodnotit, posoudit.

2.4.4 Počet položek v testu

Důležitý je také **počet položek v testu**. Větší počet položek umožňuje větší spolehlivost (reliabilitu) testu. Za dolní použitelnou hranici lze považovat deset úloh [56], [27]. Seriozní standardizovaný didaktický test by neměl mít méně než 40 položek. Také **věk** respondentů a jejich **intelektuální** úroveň jsou činiteli ovlivňujícími počet úloh pro didaktický test.

Dalším důležitým a často limitujícím faktorem, určujícím počet úloh v testu, je čas, který je možno testování dát. Časově náročnější testy mívají čistý testovací čas 35-40 minut. Rötling [54] s odvoláním na psychohygienické požadavky udává 40 minut jako horní časový limit. Turek [66] jako horní časovou hranici pro žáky 2. stupně základní školy uvádí 40 minut, pro žáky střední školy až 80 minut. Mužic [41] žádá, aby u testů, které měří úroveň vědomostí žáků, byl limitován čas tak, aby 90 % žáků stačilo testem projít. Education Testing Service, vydávající didaktické testy, uvádí adekvátní čas, ve kterém projde 80 % testovaných celým testem. Řešátko [56] doporučuje určovat optimální čas na vypracování testu jako průměrný čas, který žáci při vypracování testu potřebovali, zvětšený o 25-50 %. Jako vhodnou dobu potřebnou k vyřešení jedné položky uvádí Řešátko přibližně 0,5-1 min., nedoporučuje používat položky, jejichž řešení by vyžadovalo více než 5 minut.

Počet položek v testu by také měl respektovat způsob, jakým test zamýšlíme vyhodnotit. Byčkovský [12] doporučuje odlišný způsob skórování pro testy s více než 20 položkami (lze skórovat binárně) a pro testy s méně než 20 položkami (doporučuje skórování složené). Skórování viz kap. 2.5.1. Kromě toho Meloun a Militký [39] doporučují odlišný způsob statistického zpracování výsledků pro malé soubory do 20 měření (tzv. Hornův postup, str. 6). Pro výpočet reliability je např. pro výpočet pomocí Spearmanova-Brownova vzorce (26) nutný sudý počet položek.

2.5 Analýza didaktických testů

2.5.1 Kvalitativní a kvantitativní analýza

(podle Lapitky [37])

- **Kvalitativní analýza** – rozbor jednotlivých úloh (rozbor nejčastějších chyb, pokus o odhalení jejich příčin apod.)
- **Kvantitativní analýza** – obvykle je rychlejší a přehlednější než analýza kvalitativní. Výstupem kvantitativní analýzy je:
 - o **Relativní úspěšnost:**
 - relativní úspěšnost vztažená na jednotlivé **položky** (tj. kolik procent respondentů danou položku správně řešilo)
 - relativní úspěšnost vztažená na jednotlivé **respondenty** (tj. kolik procent položek daný respondent správně řešil). Průměrem těchto údajů přes všechny respondenty získáme relativní úspěšnost testované skupiny (např. třídy).
 - o **Skóre:** Součet všech bodů, které respondent při vypracování testu získal, se nazývá celkové skóre (nebo individuální skóre nebo hrubé skóre).
 - Podstatně jednodušší je statistické vyhodnocování testů v případě, že každá správně vyřešená úloha je hodnocena jedním bodem, špatná nebo chybějící odpověď je hodnocena nula body (binární skórování).
 - Lapitka připomíná nehomogenitu didaktických testů a z toho plynoucí nutnost hodnocení různých úloh různým počtem bodů (**vážené skórování**).

Hrubé skóre [47]:

$$X_k = \sum_{i=1}^j x_{i,k}, \text{ kde}$$

X_k hrubé skóre, které získal jeden respondent označený k
 k číslo označující respondenta
 i číslo položky
 $x_{i,k}$ hodnocení i -té položky řešené k -tým respondentem
 $x_{i,k} = 1$ znamená položku řešenou správně,
 $x_{i,k} = 0$ znamená položku neřešenou nebo řešenou špatně
 j počet položek

Průměrné skóre určité skupiny respondentů (úspěšnost skupiny respondentů):

$$\bar{X} = \frac{\sum_{k=1}^n X_k}{n}, \text{ kde}$$

(35)

\bar{X} průměrné skóre určité skupiny respondentů
 X_k hrubé skóre, které získal jeden respondent označený k
 k číslo označující respondenta
 n počet respondentů

V [47] se připomíná, že \bar{X} má smysl počítat jen za předpokladu skupiny rovnocenných respondentů. Otázku, kdy lze skupinu respondentů určité skupiny pokládat za rovnocenné, však [47] neřeší.

- **Celkové skóre a klasifikace:** Obvyklým častým cílem didaktických testů je získání podkladů pro klasifikaci žáků. Problematika převodu výsledku testu na známku je velmi obtížná. Protože předkládaná habilitační práce (a test v ní vytvořený a používaný) si v žádném případě klasifikaci žáků za cíl neklade, neuvádím ani na tomto místě podrobnější informace o uvedené problematice.
- **Standardizace didaktických testů:** Lapitka uvádí: O standardizaci didaktických testů se stále živě diskutuje, protože v teorii, ale i v polemikách na veřejnosti, se střetávají dosti protichůdné názory. Někteří autoři i uživatelé didaktických testů se o standardizaci ani nezmiňují, jiní ji pokládají za velmi důležitou. Je nutno odlišovat dva pojmy:
 - o standardizace skóre (soubor matematických operací, pomocí kterého jsou výsledky testování konkrétního respondenta nebo malé skupiny respondentů srovnávány s výsledky celé sledované populace)
 - o standardizace testu (určení norem na administraci a vyhodnocení výsledků, používaných následně pro celou sledovanou populaci) – viz též kapitoly 2.2.2 a 2.2.9.

2.5.2 Skórování odpovědí

Hodnocení odpovědí na testové položky neboli **skórování odpovědí** se provádí většinou pomocí bodů.

Skórování může být

- **nevážené (binární):** jeden bod za správnou odpověď, žádná nebo chybná odpověď má nulu. Toto skórování se používá velmi často z důvodu snadného statistického vyhodnocování výsledků testování. Na druhou stranu ovšem je nutno mít na paměti, že tento způsob skórování by neměl být podkladem pro klasifikaci v případě, že testové úlohy nebyly rovnocenné,
- **vážené:** položkám je podle obtížnosti a důležitosti přiřazen odlišný počet bodů. Konkrétní způsob přiřazení různého počtu bodů za různé úlohy je však předmětem diskuse (lze bodově hodnotit významnost testovaného učiva, obtížnost položky, úroveň myšlenkových operací,...).

Součtem bodů za správné řešení jednotlivých položek v testu získáme tzv. **hrubé skóre**, které se někdy nazývá také **prosté skóre**.

Výsledky testování je možno využít ke klasifikaci žáků. Vzhledem k tomu, že předkládaná habilitační práce si nekladla za cíl studenty hodnotit ani vytvořit test, který by klasifikaci umožňoval, nebudu se touto problematikou dále zabývat.

Medián neboli střední hodnota je roven [36]:

1. je-li lichý počet členů: M_e je roven hodnotě prostředního členu (při seřazení podle velikosti),
2. je-li sudý počet členů: M_e je roven průměru dvou středních členů (při seřazení podle velikosti).

Modus je nejčastěji se vyskytující hodnota v daném souboru.

2.5.3 Vlastnosti testových položek

2.5.3.1 Obtížnost položky

Jak uvádí Lapitka v [37], obtížnost testových položek je znak, který může být interpretován mnoha různými způsoby. V naší literatuře převažuje směr, podle kterého se obtížnost úlohy posuzuje podle počtu úspěšných respondentů.

Pro položky se počítá tzv. **index obtížnosti P**, nejčastěji v procentech (např. [27]):

$$P = \frac{n_s}{n} \cdot 100, \text{ kde}$$

(36)

P index obtížnosti

n_s počet respondentů, kteří danou položku zodpověděli správně

n celkový počet respondentů

Kromě toho existuje ještě položka zvaná **hodnota obtížnosti Q**, přičemž platí $P + Q = 100$ (%).

Za velmi obtížné lze podle Chrásky [27] pokládat takové položky, pro které je $P < 20$. Naopak velmi snadné jsou ty položky, pro které $P > 80$. Pro položky rozlišujících testů pokládá za nevhodnější hodnotu $P \doteq 50$. Položku extrémně snadnou ($P \doteq 100$) z psychologických (uklidňujících) důvodů doporučuje v úvodní části testu. Naopak extrémně těžké položky ($P \doteq 0$) v testu nedoporučuje používat.

Proti hodnocení obtížnosti položek pomocí výše definovaného indexu obtížnosti vznáší Lapitka [37] následující námitku: Úloha, kterou dokáže správně vyřešit většina respondentů, je lehká. Opačná úvaha však může selhat, protože rozdíly v úspěšnosti respondentů závisí nejen na náročnosti položky, ale i na obsahu a kvalitě vyučovacího procesu. Lapitka proto popisuje podstatně odlišnou metodu určení obtížnosti položek, vycházející z důležitosti učeného prvku a náročnosti poznávací operace nebo zručnosti. S odkazem na Ebela [17] uvádí ještě jinou metodu, založenou na tzv. očekávaném náhodném skóre.

Bývá slušným zvykem seřadit položky tak, aby počáteční otázky byly lehčí než otázky další, aby žák získal sebedůvěru.

2.5.3.2 Citlivost položky

Citlivost položky (index diskriminace, rozlišovací hodnota, diskriminační hodnota, rozlišovací ostrost, rozlišovací schopnost, selektivnost) je veličina, která kvantifikuje schopnost testové položky rozlišovat respondenty podle jejich výkonnosti ve vztahu k celkovému výsledku testu. Metody výpočtu citlivosti položky vycházejí z myšlenky, že položka je dostatečně citlivá tehdy, jestliže ji správně zodpoví velký počet respondentů, kteří v celkovém testu dosáhli vynikajícího výsledku, a naopak ji nezvládne většina respondentů z nejméně výkonné skupiny.

Při posuzování citlivosti položek se většinou nejprve vyberou podle celkového počtu dosažených bodů (hrubého skóre) dvě skupiny respondentů: „lepší“ skupina (s nejvyšším počtem dosažených bodů) a „horší“ skupina (s nejnižším počtem dosažených bodů). Tyto skupiny musí obsahovat stejný počet respondentů, např. každá obsahuje[†] 50 %, 33 %, nebo 25 % z celkového počtu respondentů. Obě skupiny musí obsahovat stejný počet respondentů.

Citlivost položky se dá exaktně posoudit pomocí výpočtu některého z koeficientů citlivosti, kterých byla navržena celá řada. Všechny tyto koeficienty mohou nabývat hodnot od -1 přes 0 do 1 . Čím větší hodnotu koeficient má, tím lépe úloha rozlišuje mezi respondenty s lepšími vědomostmi a respondenty s horšími vědomostmi. Pokud má koeficient hodnotu 0 , mezi oběma skupinami

[†] Údaje o procentech jsou přibližné. Musejí se podříditi elementárnímu požadavku, že počet respondentů v každé ze skupin H, D musí být celé číslo.

respondentů nerozlišuje. Záporná hodnota vypovídá o tom, že položka zvýhodňuje respondenty, kteří mají v testu celkově horší výsledky.

Nejjednodušším ukazatelem citlivosti testové úlohy je **koeficient ULI** (upper-lower-index). Je definován vztahem uvedeným např. v [37] nebo v [27]:

$$I_d = \frac{L - H}{N}, \text{ kde} \quad (37)$$

I_d citlivost položky

Lpočet správných řešení dané položky mezi respondenty z „lepší“ skupiny

H počet správných řešení dané položky mezi respondenty z „horší“ skupiny

N počet respondentů v jedné skupině

U koeficientu ULI se požaduje [27], aby v případě úloh s hodnotou obtížnosti 0,3-0,7 bylo I_d alespoň 0,25 a u úloh s obtížností 0,2-0,3 a 0,7-0,8 alespoň 0,15. To platí pro skupiny L, H vytvořené z 50 % respondentů.

Poněkud pracnější, ale podle [27] většinou spolehlivější metodou výpočtu koeficientu citlivosti je výpočet tzv. **tetrachorického koeficientu citlivosti**:

Pro výpočet tohoto koeficientu je potřeba pro každou položku sestavit tzv. tetrachorickou tabulku, která uvádí počty respondentů ze skupin L („lepší“) a H („horší“), kteří na položku odpověděli správně (+), nebo špatně, resp. neodpověděli (-).

Tetrachorický koeficient citlivosti úlohy se vypočítá ze vztahu:

$$r_{tet} = \cos \left(180 \frac{\sqrt{bc}}{\sqrt{bc} + \sqrt{ad}} \right), \text{ kde} \quad (38)$$

apočet respondentů ze skupiny L, kteří odpověděli správně

b počet respondentů ze skupiny L, kteří odpověděli špatně nebo neodpověděli

cpočet respondentů ze skupiny H, kteří odpověděli správně

d počet respondentů ze skupiny H, kteří odpověděli špatně nebo neodpověděli

cos goniometrická funkce (zpracovávající argument uvedený v úhlových stupních)

Většinou se uvádí, že tetrachorický koeficient citlivosti by neměl být u vyhovujících testových úloh nižší než 0,15. Tato hodnota platí v případě, pokud skupiny L a H byly vytvořeny z 50 % respondentů. Pokud k vytvoření skupin L a H použijeme menší počet respondentů, pak při posuzování vypočítaných hodnot r_{tet} musíme uplatnit přísnější měřítko.

Někdy se podle [27] citlivost testových úloh posuzuje také pomocí tzv. **bodově biseriálního koeficientu**. Vzorec pro výpočet je:

$${}_b b_{bis} = \frac{\bar{X}_s - \bar{X}_n}{s_x} \sqrt{p(1-p)}, \text{ kde} \quad (39)$$

${}_b r_{bis}$ bodově biseriální koeficient citlivosti položky

\bar{X}_s průměrný počet bodů v testu u respondentů, kteří danou úlohu řešili správně

\bar{X}_n průměrný počet bodů v testu u respondentů, kteří danou úlohu řešili nesprávně

s_x směrodatná odchylka, vypočítaná ze všech testových výsledků

p index obtížnosti položky (vyjádřený jako desetinné číslo, nikoli v procentech).

Zřejmě tedy $p = 0,01P$.

Vyhovující testová úloha by měla vykazovat bodově biseriální koeficient citlivosti minimálně 0,2.

2.5.3.3 Analýza nenormovaných odpovědí

Chráska [27] s odvoláním na Byčkovského [12] rozebírá i analýzu tzv. nenormovaných (tj. vynechaných a nesprávných) odpovědí:

Vynechání odpovědi může znamenat vedle neznalosti učiva také nepochopení formulace úlohy, nedostatek času k vypracování odpovědi atd. Zvýšenou pozornost je třeba věnovat otevřeným úlohám, kde odpověď vynechalo více než 30-40 % respondentů. U uzavřených úloh je tato hranice již na 20 % respondentů.

Rozebírání nesprávných odpovědí: U *úloh s výběrem odpovědí* postačí překontrolovat, zda všechny nabídnuté distraktory jsou pro respondenty dostatečně atraktivní. Distraktor, který téměř nikdo z respondentů nevolí, by měl být pokud možno nahrazen atraktivnějším distraktorem, případně odstraněn.

U *otevřených úloh* se doporučuje veškeré chyby žáků v určité testové úloze rozdělit na tzv. základní chyby (způsobené skutečným nezvládnutím učiva) a vedlejší chyby (způsobené různými náhodnými vlivy, např. přehlédnutím, numerickou chybou ve výpočtu, nepřesností, špatnou čitelností textu atd.). V dobré testové úloze by počet hlavních chyb měl být vždy větší než počet chyb vedlejších.

2.5.3.4 Vliv jedné položky na celkové skóre testu

Tuto hodnotu ve své práci kvantifikuje např. Půlpán [47], [48]. V jeho pracech je např. odvozen výpočetní vztah pro určení položky, která nejméně ovlivňuje skóre celého testu. Pomocí tohoto vztahu lze zredukovat celkový počet testových položek tak, aby se celkové výsledky upraveného testu co nejméně lišily od testu původního.

2.6 Ověřování a optimalizace didaktického testu

(podle [27])

I když plánování a konstrukci didaktického testu věnujeme značnou pozornost, nemůžeme si být bez jeho experimentálního použití nikdy jisti tím, jaké vlastnosti skutečně má. Relativně definitivní představu o jeho vlastnostech můžeme získat až po ověření testu na vzorku respondentů. Díky tomuto ověřování lze případné nevhodné vlastnosti testu odstranit, zmírnit nebo korigovat. U standardizovaných didaktických testů se při ověřování většinou pracuje se vzorky 300-500 respondentů.

2.6.1 Identifikace nevhodných testových položek

Z dříve řečeného plyne, že nevhodná testová položka má zejména následující vlastnosti:

- Je příliš obtížná nebo příliš snadná.
- Je málo citlivá.
- Příliš mnoho respondentů vynechá odpověď.
- Počet vedlejších chyb v úloze převažuje nad počtem hlavních chyb (u položek otevřených).
- Některé distraktory jsou pro respondenty podstatně méně přijatelné než jiné (u položek s výběrem odpovědí).

2.6.2 Vytvoření definitivní podoby didaktického testu

Nevhodné položky je lépe z testu vyřadit a nahradit je vhodnějšími. Proto je vhodné navrhovat v první fázi přípravy testu položek více. Jestliže se určitá položka jeví jako problematická, ale zkouší přitom důležitou část učiva, je možno se pokusit o její úpravu: přesnější (srozumitelnější) formulace úlohy, změna některého z distraktorů apod.

Pokud se v testu užívá položek více typů, doporučuje se položky stejného druhu soustředit do jedné části testu. Položky v takto vzniklých částech testu se seřadí podle vzrůstající obtížnosti.

Jestliže má být didaktický test použit jako seriózní prostředek měření, je třeba zabezpečit podmínky pro samostatnou práci respondentů. Tento požadavek je zvláště naléhavý u testů, ve kterých jsou používány úlohy s výběrem odpovědí, protože kód správných odpovědí se velmi snadno napovídá. Nesamostatné práci respondentů lze zabránit jednak důsledným dozorem, ale především tím, že vytvoříme dvě nebo více ekvivalentních forem testu.

Ekvivalentní formu testu lze vytvořit několika způsoby:

- změnou pořadí položek v testu (položky však mají stále mít celkovou tendenci vzrůstající obtížnosti).

- změnou pořadí nabídek odpovědí (u úloh s výběrem odpovědí); vyhodnocování odpovědí v testu se tím však poněkud zkomplikuje.
- ve výpočetních příkladech změnou číselného zadání.
- různými kombinacemi výše uvedených možností

Rötling [54] navrhuje i další možnosti tvorby variant testových úloh, avšak podle mého názoru tyto další možnosti velmi snadno vedou ke vzniku neekvivalentních testů. Proto zde tyto možnosti neuvádím.

2.6.3 Standardizace didaktického testu

(podle Chrásky [27])

Smyslem standardizace testu [27] je vytvoření testového standardu (testové normy), který umožní zařadit respondenta podle dosaženého počtu bodů do určitého žebříčku (stupnice, škály). Dosažený počet bodů (hrubé skóre) v testu sám o sobě neříká nic neříká o tom, zda výkon respondenta je průměrný, dobrý či slabý. Teprve na základě srovnání dosaženého výkonu s výkony ostatních respondentů lze daného jedince adekvátně posoudit.

U tzv. standardizovaných didaktických testů se výkon jednotlivých respondentů porovnává s reprezentativním (tzv. standardizačním) vzorkem (zpravidla se jedná o stovky respondentů). Postup, kterým se toto srovnávání realizuje, se nazývá standardizace testu. Způsobů provedení standardizace je celá řada. Chráska [27] uvádí percentilovou škálu, C-škálu, škálu STANIN, z-škálu, Z-škálu a T-škálu.

2.6.3.1 Percentilová škála

Ke každému dosaženému počtu bodů (hrubému skóre) se přiřadí tzv. percentilové pořadí, které udává, kolik procent respondentů ve vzorku dosáhlo horšího výkonu. Před výpočtem seřadíme testy od nejhoršího k nejlepšímu. Percentilové pořadí pro určitý výsledek v testu lze vypočítat podle vzorce

$$PR = 100 \frac{n_k - \frac{n_i}{2}}{n}, \text{ kde} \quad (40)$$

PR percentilové pořadí respondenta pro daný výsledek v testu,

n_k kumulativní četnost u daného výsledku (četnost testů s daným skóre a horším),

n_i četnost daného výsledku,

n celkový počet respondentů.

Jestliže se jsou zachyceny jen četnosti výsledků v určitých bodových intervalech, provádí se výpočet podle vztahu

$$PR = 100 \frac{n_k - \frac{d_L n_i}{h}}{n}, \text{ kde} \quad (41)$$

PR, n ... viz (40),

d_L rozdíl mezi daným výsledkem a dolní hranicí intervalu, v němž se tento výsledek nachází,

n_k kumulativní četnost v intervalu, v němž se daný výsledek nachází,

n_i četnost v intervalu, v němž se výsledek nachází,

h hloubka intervalu (rozdíl mezi nejlepším a nejhorším skóre určujícím daný interval).

2.6.3.2 C-škála

Celý standardizační vzorek respondentů (uspořádaný podle rostoucího počtu dosažených bodů) se rozdělí do 11 skupin (stupňů škály) tak, že do první skupiny (bod škály 0) se umístí 1,2 % nejhorších respondentů, do druhé skupiny (bod škály 1) se umístí 2,8 % žáků atd. Procento respondentů pro každý bod škály je uvedeno v následující tabulce. Procentuální rozdělení k jednotlivým bodům škály je symetrické vzhledem ke střednímu bodu škály. Tento střední bod škály obsahuje také největší procento žáků (19,8 %).

Tab. 2: Konstrukce C-škály.

Body C-škály	Procenta případů	Kumulativní procenta
0	1,2	1,2
1	2,8	4,0
2	6,6	10,6
3	12,1	22,7
4	17,4	40,1
5	19,8	59,9
6	17,4	77,3
7	12,1	89,4
8	6,0	96,0
9	2,8	98,8
10	1,2	100,0

Při konstrukci C-škály se nejdříve vypočítají kumulativní četnosti. Ty se pak převedou na tzv. relativní kumulativní četnosti (v procentech) pomocí vztahu

$$n_{kr} = 100 \frac{n_k}{n}, \text{ kde}$$

(42)

n_{kr} relativní kumulativní četnost u daného výsledku

n_k kumulativní četnost u daného výsledku (tj. součet četností testů s daným skóre a horším),

n celkový počet respondentů.

K relativním kumulativním četnostem vypočítaným podle (42) se potom vyhledají stejné nebo nejbližší nižší hodnoty kumulativních procent z výše uvedené tabulky. Tak se získají intervaly hodnot skóre pro jednotlivé (stupně) C-škály.

2.6.3.3 Škála STANIN

Devítibodová škála STANIN vznikne spojením prvních dvou a posledních dvou stupňů C-škály. První a poslední bod této škály tedy obsahují 1,2 % + 2,8 % = 4 % případů (respondentů)

2.6.3.4 z-škála

Konstrukce této škály vychází z předpokladu, že výsledky testování mají tzv. normální rozdělení. Nejdříve by se mělo vždy ověřit (pomocí testu dobré shody χ^2 postupem popsáným např. v [22]), zda je tento předpoklad splněn. Hodnota z-škály vyjadřuje, jak daleko je určitý dosažený výsledek od aritmetického průměru, přičemž jednotkou této vzdálenosti je směrodatná odchylka. Platí, že

$$z = \frac{x - \bar{x}}{s}, \text{ kde}$$

(43)

z hodnocení na z-škále,

x určitý testový výsledek (hrubé skóre),

\bar{x} aritmetický průměr výsledků v testu (průměrné skóre přes všechny respondenty),

s směrodatná odchylka pro všechny testové výsledky.

Hodnoty z-škály se pohybují zpravidla v intervalu od -3 do $+3$, průměrná hodnota je $z = 0$.

Samotná z-škála se při standardizaci didaktických testů užívá jen zřídka, užívá se jí však při konstrukci jiných standardních škál.

2.6.3.5 Z-škála

Tato škála vychází ze z-škály (43) a je definována vztahem $Z = 100 + 10z$

(44)

2.6.3.6 T-ŠKÁLA

Velmi často se k vyjadřování výsledků ve standardizovaných didaktických testech používá stupnice nazývaná T-škála. Její hodnoty jsou zpravidla v intervalu od 20 do 80 bodů a průměrná hodnota je 50. I tato škála vychází ze z-škály (43) a je definována vztahem

$$Z = 50 + 10z$$

(45)

2.7 Využití didaktických testů k posouzení úrovně zvládnutí různých úrovní myšlenkových operací s učivem

Podle Chrásky [27] je pojem didaktický test u různých autorů definován různě. Tato různá vymezení se ale shodují v tom, že jde o zkoušku, která se orientuje na objektivní zjišťování úrovně zvládnutí učiva u určité skupiny osob. Chráska uvádí také Byčkovského [12] definici: Didaktický test je nástroj systematického zjišťování (měření) výsledků výuky. Řešátko [56] připomíná, že didaktický test se stal také významným pomocníkem pedagogické vědy (zejména didaktiky), již pomáhá při objektivním ověřování různých hypotéz.

S ohledem na informační explozi a z ní plynoucí rychle rostoucí množství informací, které by se (při stávajícím způsobu výuky) měly stát předmětem učiva, se čím dále tím více zdůrazňuje požadavek, aby žáci nejen vstřebali určité množství informací, ale především aby se naučili tyto informace samostatně vyhledávat a zpracovávat. Důležitým výukovým cílem se tedy stává především žákovská schopnost samostatné práce s informacemi. Podle Niemierkovy klasifikace výukových cílů se jedná o úroveň C (používání vědomostí v typových situacích) a D (používání vědomostí v problémových situacích) – viz též kapitola 2.4.3.

Má-li být úroveň dosažení tohoto cíle kontrolována pomocí didaktických testů, je nutné, aby položky vyžadovaly od žáků logické myšlení [57], tedy nejen mechanickou reprodukci poznatků. Někteří autoři se tímto problémem zabývali s různou úspěšností již dříve. Slibná se podle [57] ukazuje cesta využívání metod psychometrie k vyhodnocování didaktických testů.

Přesto, že jeho práce jsou již poměrně staré (od 70. let 20. století), neobjevuje se v současné české literatuře, zabývající se didaktickými testy, o této kategorii testů mnoho zmínek. Využití některých principů testů inteligence

2.7.1 Didaktickologické testy

M. Cipro [14] uvažuje o tzv. didaktickologických testech. Navrhuje využít v didaktických testech podobných principů jako v testech inteligence, např. tzv. logické úměry. Jako příklad uvádí: „Za jednu z účinných forem testu inteligence se považuje doplňování analogických členů tzv. **logické úměry** typu:

nahore : vysoko = dole : (nízko) .

Jeden z těchto čtyř členů (zpravidla poslední) je vynechán a žák jej má doplnit. Autoři podobných psychologických testů užívají zpravidla úmyslně materiál didakticky irelevantního.“

Cipro navrhuje využití tohoto principu i v didaktických testech, např. ve vlastivědě:

„letadlo : vzduch = loď : (voda)
železná ruda : železo = bauxit : (hliník)

Při správném řešení takového testu musí žáci prokázat jak věcné znalosti, tak i vtip, inteligenci, prostě schopnost těchto věcných poznatků logicky využít. Musí především pochopit podstatu vztahu mezi prvními dvěma členy a pak hledat analogický vztah, určený třetím členem.“

Dále Cipro navrhuje: „Didakticko-logické testy mají mnoho možností, jak kombinovat kontrolu izolovaných věcných znalostí zároveň s kontrolou rozvoje myšlení – schopnosti uvádět tyto znalosti do vyšších myšlenkových struktur. K tomu účelu se mohou dobře osvědčovat např. testy ukládající **sestavení nahdile uvedených pojmů do určité logické řady**. Mohou to být např. fáze nějakého procesu (vajíčko – housenka – kukla – motýl), (železná ruda – surové železo – ocel) a jiné.“

Jiný vhodný typ jsou podle Cipra testy zaměřené na **doplňování klíčových slov**, předpokládající ne pouze znalost faktů, ale schopnost postihnout souvislosti mezi fakty. „Např. ve větě: Voda se čistí chlorem, aby neobsahovala žádné původce nakažlivých nemocí, se vypustí spojka aby. Jejím doplněním má žák prokázat, že chápe finální vztah mezi oběma větami.“

„Užitečné jsou také didakticko-logické testy, prověřující schopnost **logického třídění**. Např. žák má rozdělit do dvou skupin následující pojmy: Praha, Vltava, Ohře, Plzeň, Brno, Dyje, Ostrava, Labe, Svratka, Liberec. Kdyby úkol zněl „Odděl názvy měst od názvů řek“, byl by to vlastně běžný didaktický test, vyjadřující pouze faktické znalosti. Proto je lépe formulovat úkol obecněji, tak aby žák musel srovnáním významu jednotlivých slov přijít na to, o které dvě skupiny se jedná.“

Využití tohoto typu položek v didaktických testech je podle [68] málo vhodné, protože je v nich potřebné prokázat také tu složku inteligence, která je člověku vrozená (tedy nikoli získatelná vzděláváním).

2.7.2 Test představitosti

Z. Půlpán [49] vysoce cení představitost. Podle něj intuitivní neformální pochopení souvislostí je základem porozumění a představitost v něm hraje základní roli. Je tdy významnou složkou vzdělávacího procesu. jako matematik připomíná, že vidění v matematice je nutno se učit a nelze je oddělit od vědění. Podle jeho slov se představitost výrazně uplatňuje v procesu poznání, a to zejména při řešení úloh. Proto spojuje výzkum představitosti s testem, kdy měli respondenti řešit 12 úloh z oblasti geometrie. Odpovědi respondentů hodnotí u každé položky na stupnici 0, 0,1, 0,2, ... 0,9, 1. Test uvedený v [49] má podle mého názoru velmi mnoho znaků společných s testy psychologickými. Hodnoceny jsou totiž i takové části odpovědi, na které zadání testu explicitně neupozorňovalo. To podle mého názoru je odůvodnitelné při zkoumání psychických vlastností respondentů, ne však při zkoumání jejich znalostí.

Příklad:

„Úloha 1:

Nakreslete dva čtverce tak, aby jejich průnikem (společnou částí) byl

- a) čtverec.
- b) trojúhelník
- c) pětiúhelník.“

Výňatek z klíče hodnocení: „...Ačkoli v úloze nebyla formulována výzva k diskusi o možnostech konstrukce různých výsledků, rozlišujeme kvalitu řešení podle toho, zda je uvedeno určité „speciální“ řešení nebo řešení „obecné“. Řešení každé dílčí úlohy hodnotíme stupněm 0,2; obecnější řešení a názny nebo provedení diskuse o možných výsledcích hodnotíme postupně stupni 0,7,...,1.“

Podle mého názoru je možno tímto způsobem testovat učením získanou složku představitosti. Je však nutno v zadání explicitně uvést veškeré požadavky. V opačném případě se testují (mimo rozsah pokrytý zadáním) především vlastnosti vrozené.

Publikace [49] dále obsahuje poměrně složitý matematický aparát umožňující statistické vyhodnocení testových výsledků.

Využití diagnostických tabulek

V [57] je uvedena práce A. A. Čencova a příspěvek V. Švece, využívající tzv. **diagnostických tabulek**. Princip je přibližně ten, že autor testu při jeho přípravě rozdělí učivo na malé učební jednotky, jako jsou pojmy, číselná a slovní data, definice apod. Tyto jednotky označme např. J_1, J_2, \dots, J_{10} . Nyní autor testu sestaví např. sedm položek (složitějších otázek či úkolů) testu (označme je P_1, P_2, \dots, P_7) tak, že správné řešení položky P_1 závisí např. na osvojených učebních jednotkách J_3, J_5, J_6, J_{10} , správné řešení položky P_2 závisí na osvojení učebních jednotek J_1, J_4, J_8, J_9 apod. Tyto vztahy zaznačí do tzv. diagnostické tabulky:

Tab. 3: Příklad diagnostické tabulky. Symbol 1 značí závislost správného řešení dané položky na osvojení příslušných poznatků, 0 značí nezávislost.

Položky testu	Prověřované učební jednotky									
	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8	J_9	J_{10}
P_1	0	0	1	0	1	1	0	0	0	1

P ₂	1	0	0	1	0	0	0	1	1	0
P ₃	0	0	0	1	1	0	0	0	1	1
P ₄	0	0	0	0	1	0	0	0	0	1
P ₅	0	1	1	0	0	0	0	0	0	0
P ₆	0	1	0	1	0	1	0	0	0	1
P ₇	0	1	0	0	1	0	0	0	0	0

Předpokládejme, že analýzou výsledků testu učitel zjistí, že žák neodpověděl správně na položky P₁, P₄ a P₇. Z uvedené tabulky zjistí, že chybné řešení těchto položek je způsobeno nezvládnutím těchto učebních jednotek:

u položky P₁ na jednotce J₃, J₅, J₆, J₁₀,

u položky P₄ na jednotce J₅, J₁₀,

u položky P₇ na jednotce J₂, J₅.

Učitel takto dostane tři množiny: M₁ = {J₃, J₅, J₆, J₁₀}, M₂ = {J₅, J₁₀}, M₃ = {J₂, J₅}. Jejich průnikem zjistí neosvojené učební jednotky: N = M₁ ∩ M₂ ∩ M₃ = {J₅}. Neosvojena je tedy učební jednotka označená J₅.

Nevýhodou této metody je podle mého názoru skutečnost, že takto konstruovaný test nerozliší, zda respondent učební jednotku J₅ skutečně nezná, nebo zda ji zná, ale neumí ji použít. Uvedený způsob vyhodnocení také selže, pokud respondent uhodne odpověď na některou z položek P₁, P₄, P₇.

Rozklad na řadu dílčích operací

V Řešátkově práci [56] se s odkazem na **Landu** ([56] však neuvádí konkrétní citaci) zmiňuje **jiný způsob – rozklad na řadu dílčích operací**. Na str. 71 je jeho princip ilustrován konkrétním příkladem:

„Složitější je situace při zjišťování příčiny chyb v komplikovanějších úlohách, kde řešení může probíhat řadou logických postupů. Především je nutno odhadnout postup, kterým žák řeší úlohu. Tento postup se mnohdy liší od postupu dospělých. Podle Landy je vhodné rozdělit celý problém na řadu dílčích operací. Postupně se pak zjišťuje, která operace je prvou, v níž žák nemůže postoupit dále, protože v ní soustavně chybí, popř. ji neřeší.

Učitel fyziky např. zjistí, že žáci nedovedou vypočítat proud, který prochází rezistorem s konstantním odporem při daném napětí. Příčinou může být:

d) Neznají vztah mezi proudem a napětím při konstantním odporu vodiče.

e) Dovedou tento vztah sice verbálně vyjádřit jako přímou úměrnost, ale pojem přímé úměrnosti jim je nejasný.

f) Neznají značky proudu, napětí a odporu.

g) Neovládají příslušný tvar Ohmova zákona, vyjádřený vzorcem $I = \frac{U}{R}$.

h) Nedovedou dosadit do vzorce.

i) Dělalí numerické chyby.

j) Dopouštějí se chyb z nepozornosti, např. zamění číslice, opíšou chybně číslici atd.

k) Neuvědomují si, co vypočítali.

Prvé dvě příčiny se při vlastním výpočtu vůbec nemusí projevit. někdy se ovšem ukáží nepřímou, např. tím, že žák si vůbec neuvědomí chybu, když u ve dvou po sobě následujících příkladech při zvětšení napětí vychází na stejném odporu menší proud.

Dobře sestavený didaktický test však musí odhalit i chyby tohoto rázu. Do takového testu zařadíme např. položky:[‡]

1. Co platí o závislosti mezi proudem a napětím při konstantním odporu vodiče?

2. Projeví se růst napětí na konstantním odporu změnou proudu v obvodu s odporem?

3. Jakou značku má proud?

4. Jakou značku má napětí?

5. Jakou značku má odpor?

6. Co vyjadřuje vzorec $I = \frac{U}{R}$?

[‡] Pro zkrácení necituji nabídnuté odpovědi.

7. Je některý z uvedených vzorců nesprávný pro vyjadřování Ohmova zákona?
8. Jaký proud prochází rezistorem o odporu 20Ω při napětí 10 V ?
9. Žárovka je připojena na napětí 220 V . Její vlákno má odpor 2000Ω . Jaký proud jí prochází?

Uvedené pořadí položek odpovídá provedenému rozboru. Při sestavení testu je možno pořadí položek upravit, na př. na začátek přesunout položky č. 3 a 5, u nichž očekáváme vysoké procento správných odpovědí.“

Princip Landova testování nemá nevýhodu, kterou jsem zmínila u metody diagnostických tabulek. Z testu konstruovaného podle Landy je možno určit, které části učiva žáci zvládli a v čem je případná příčina neúspěchu při řešení složitější otázky. Pokud však má být takový test použit jako podklad např. pro klasifikaci, mělo by se pečlivě zvážit, podle jakého klíče klasifikaci přiřadit, protože položky testu jsou na sobě závislé. Student také může zvolit jiný, učitelem nepředpokládaný algoritmus řešení.

2.7.3 Cílené využití různých forem položek

Turek [66] uvádí vhodnost jednotlivých typů úloh k měření jednotlivých úrovní učení podle taxonomie cílů (jak Bloomovy, tak Niemierykové). Následující tabulky jsou převzaty z [66], str. 31.

Tab. 4: Vhodnost jednotlivých forem úloh didaktického testu k měření jednotlivých úrovní učení podle taxonomie cílů B. S. Blooma.

Forma úloh	1	2	3	4	5	6
otevřené se širokou odpovědí nestrukturalizované	–	–	+	+	++	++
otevřené se širokou odpovědí strukturalizované	–	+	++	++	+	++
otevřené se stručnou odpovědí – produkční	++	++	++	+	–	–
otevřené se stručnou odpovědí – doplňovací	++	+	+	–	–	–
uzavřené dichotomické	++	++	+	–	–	–
uzavřené polytomické	+	++	++	–	–	+
uzavřené s přiřazovacím výběrem	++	++	+	+	–	–
uzavřené s výběrem uspořádání	+	++	–	–	–	+

Vysvětlivky: ++ velmi vhodná forma úlohy didaktického testu
 + vhodná forma úlohy didaktického testu
 – nevhodná forma úlohy didaktického testu

Tab. 5: Vhodnost jednotlivých forem úloh didaktického testu k měření jednotlivých úrovní učení podle taxonomie cílů B. S. Blooma.

Forma úloh	A	B	C	D
otevřené se širokou odpovědí	–	–	+	++
otevřené se stručnou odpovědí – produkční	++	++	+	–
otevřené se stručnou odpovědí – doplňovací	++	+	+	–
uzavřené dichotomické	++	++	+	–
uzavřené polytomické	+	++	++	+
uzavřené s přiřazovacím výběrem	++	++	+	–
uzavřené s výběrem uspořádání	+	++	–	–

Vysvětlivky viz

Tab. 4.

3 Použitá literatura

- [1] BANÝR J.: Pokus o formulaci standardu kvalifikace učitele chemie. In Eseje na téma Standard kvalifikace učitele chemie, s. 4-7. Podpora rozvoje učitelských vzdělávacích programů a jiných vzdělávacích aktivit 2001. Univerzita Karlova v Praze, Pedagogická fakulta. Praha 2001.
- [2] BENEŠ, Pavel. – PUMPR, Václav – BANÝR, Jiří. *Základy chemie 1 pro 2. stupeň základní školy, nižší ročníky víceletých gymnázií a střední školy*. 3. vyd. – dotisk. Praha: Fortuna, 2004. ISBN 80-7168-720-0.
- [3] BENEŠ, Pavel. – PUMPR, Václav – BANÝR, Jiří. *Autorská řešení. Základy chemie 1. Základy chemie 2. Pracovní sešity*. Fortuna, Praha 2004.
- [4] BENEŠ, Pavel. – PUMPR, Václav – BANÝR, Jiří. *Základy chemie 1 – pracovní sešit*. Praha: Fortuna, 1996.
- [5] BENEŠ, Pavel. – PUMPR, Václav – BANÝR, Jiří. *Základy chemie 2 – pracovní sešit*. Praha: Fortuna, 1997.
- [6] BENEŠ, Pavel. – PUMPR, Václav – BANÝR, Jiří. *Základy chemie 2 pro 2. stupeň základní školy, nižší ročníky víceletých gymnázií a střední školy*. 3. vyd. – dotisk. Praha: Fortuna, 2003. ISBN 80-7168-748-0.
- [7] BENEŠ, Pavel. Může dobře připravený učitel chemie přispět k rozvoji společnosti? Eseje na téma Standard kvalifikace učitele chemie, s. 8-10. Podpora rozvoje učitelských vzdělávacích programů a jiných vzdělávacích aktivit 2001. Univerzita Karlova v Praze, Pedagogická fakulta. Praha 2001.
- [8] BLOOM, B. S. – KRATHWOHL, D. R. *Taxonomy of Educational Objectives*. Longmans, Green & Co., Inc., New York, 1956. ???
- [9] BLOOM, B.S. – *Taxonomy of Educational Objectives. Handbook I. Cognitive Domain*. New York 1956.
- [10] BOLDIŠ, Petr. *Bibliografické citace dokumentů podle ČSN ISO 690 a ČSN ISO 690-2: Část 2 – Modely a příklady citací u jednotlivých typů dokumentů*. Verze 3.0 (2004), poslední aktualizace 11. 11. 2004. URL: <<http://www.boldis.cz/citace/citace2.ps>>. <<http://www.boldis.cz/citace/citace2.pdf>>
- [11] BRDIČKA, Rudolf – DVOŘÁK, Jiří. *Základy fyzikální chemie*. 2. vyd. Praha: Academia, nakladatelství Československé akademie věd, 1977.
- [12] BUDIŠ, Josef. Standard učitele chemie (aneb stručný neesejový přehled návrhů a zkušeností). Eseje na téma Standard kvalifikace učitele chemie, s. 15-18. Podpora rozvoje učitelských vzdělávacích programů a jiných vzdělávacích aktivit 2001. Univerzita Karlova v Praze, Pedagogická fakulta. Praha 2001.
- [13] BYČKOVSKÝ, P.: *Základy měření výsledků výuky. Tvorba didaktického testu*. Praha, ČVUT 1982.
- [14] CÍDLOVÁ, Hana – PLUCKOVÁ, Irena. Postavení a význam fyzikální chemie v procesu pregraduální přípravy učitelů. In *Pregraduální příprava a postgraduální vzdělávání učitelů chemie*. Ostrava : Ostravská univerzita, 2003. od s. 179-182, 4 s. ISBN 80-7042-960-7.
- [15] CIPRO, Miroslav. *Modernizace základního vzdělání*. 1. vyd. Praha: Státní pedagogické nakladatelství, 1968. Kapitola Didakticko-psychologické testy, s. 113-121.
- [16] DELORS J. (překlad J. Kotásek): Učení je skryté bohatství. Zpráva mezinárodní komise UNESCO „Vzdělávání pro 21. století“. Praha, ÚIV 1997.
- [17] DUŠEK, Bohuslav. Poznámky ke standardu kvalifikace učitele chemie. Eseje na téma Standard kvalifikace učitele chemie, s. 19-22. Podpora rozvoje učitelských vzdělávacích programů a jiných vzdělávacích aktivit 2001. Univerzita Karlova v Praze, Pedagogická fakulta. Praha 2001.
- [18] EBEL, R.L. – *Essential of Education Measurement*. Englewood Cliffs, N.J., Prentice Hall 1972.
- [19] FIKROVÁ, Lenka. *Příspěvek k možnostem integrace biologie a chemie v přípravě učitele chemie a ve výuce na základních školách*. Brno, 2003. Diplomová práce na Pedagogické fakultě Masarykovy univerzity v Brně na katedře chemie. Vedoucí diplomové práce Hana Cídllová.
- [20] FISHER R.A. – YATES F. *Statistical Tables for Use in Biological, Agricultural and Medical Research*. Oliver and Boyd, Edinburgh. 1938.

- [21] FRY, Edward B. *Vyučovací stroje a programované vyučování. Uvedení do problému*. 1. vyd. Praha: Státní pedagogické nakladatelství, 1966. Kapitola 5, Výchovné cíle a programování, s. 75-84.
- [22] HÁJEK, Jaroslav. *Teorie pravděpodobnostního výběru s aplikacemi na výběrová šetření*. 1. vyd. Praha: Nakladatelství Československé akademie věd, 1960.
- [23] HALIŠKA, Jaromír. *Jak testy sestavit a pracovat s nimi*. 2. vyd. Brno: Středisko služeb školám, 1999.
- [24] HNILIČKOVÁ, Jitka – JOSÍFKO, Marcel – TUČEK, Alexandr. *Didaktické testy a jejich statistické zpracování*. 1. vyd. Praha: Státní pedagogické nakladatelství, 1972.
- [25] HOLADA, Karel. Standard kvalifikace učitele chemie. Eseje na téma Standard kvalifikace učitele chemie, s. 23-26. Podpora rozvoje učitelských vzdělávacích programů a jiných vzdělávacích aktivit 2001. Univerzita Karlova v Praze, Pedagogická fakulta. Praha 2001.
- [26] http://astrokurz.wz.cz/19_Meteorologie.htm, 7. 11. 2004.
- [27] <http://measure.feld.cvut.cz/groups/edu/sz/pdfs/uloha11a.pdf>, 7. 11. 2004.
- [28] CHRÁSKA, Miroslav. *Didaktické testy. Příručka pro učitele a studenty učitelství*. 1. vyd. Brno: Paido, 1999. ISBN 80-85931-68-0.
- [29] Jedlová, Táňa. Kvalifikační standard učitele chemie. Eseje na téma Standard kvalifikace učitele chemie, s. 27-30. Podpora rozvoje učitelských vzdělávacích programů a jiných vzdělávacích aktivit 2001. Univerzita Karlova v Praze, Pedagogická fakulta. Praha 2001.
- [30] KADYROV M. *Tablice slučajnych čísel*. Taškent: 1936.
- [31] KENDALL M. G. – BABINGTON SMITH B. *Tables of Random Sampling Numbers*. Tracts of Computers No 24. 1940.
- [32] KING, Alexander – SCHNEIDER, Bertrand. *První globální revoluce*. 1. vyd. Bratislava: Bradlo, 1991. ISBN: 80-7127-048-2.
- [33] KLOUDA, P.: *Fyzikální chemie*. Pavel Klouda, Ostrava 1997.
- [34] KMEŤOVÁ, Jarmila. Učebnica a učiteľ chémie. In *Pregraduálna príprava a postgraduálna vzdelávanie učiteľov chémie. Sborník prednášok z mezinárodnej konferencie konanej 20. – 22. května 2003 v Rožnově pod Radhoštěm*. Ostrava: Ostravská univerzita v Ostravě. Přírodovědecká fakulta. Katedra chemie, 2003. s. 81-86. ISBN 80-7042-960-7
- [35] KOHOUTEK R.: *Didaktické testy*. Akademické nakladatelství CERM, Brno 1996.
- [36] Kolektiv: *Eseje na téma Standard kvalifikace učitele chemie*. Pedagogická fakulta Univerzity Karlovy v Praze, Praha 2001.
- [37] LAPITKA, Marián – LEZO, Horymír – PÍŠŠOVÁ, Mária – TEJ, Ján. *Základy metodológie pedagogického výskumu. Študijný text pre špecializačné štúdium učiteľov výskumníkov*. 1. vyd. Bratislava: Slovenské pedagogické nakladateľstvo, 1985.
- [38] LAPITKA, Marián. *Tvorba a použitie didaktických testov*. 1. vyd. Bratislava: Slovenské pedagogické nakladateľstvo, 1990. ISBN 80-08-00782-6.
- [39] Prof. RNDr. Milan MELOUN, DrSc., Katedra analytické chemie, Univerzita Pardubice: Sdělení v e-mailu. 15. 4. 2005.
- [40] MELOUN, Milan – MILITKÝ, Jiří: *Kompendium statistického zpracování dat. Metody a řešené úlohy včetně CD*. 1. vyd. Praha: Academia, 2002. ISBN 80-200-1008-4
- [41] Ministerstvo školství, mládeže a tělovýchovy: *Národní program rozvoje vzdělávání v České republice*. Bílá kniha. Tauris, Praha 2001. online: <http://www.skolam.cz/INKomensky/BKniha/BKniha.htm>, cit. 5. 5. 2005.
- [42] MUŽÍČ V.: *Testy vědomostí*. SPN, Praha 1971.
- [43] *Nápověda pro Microsoft Excel*. Microsoft®Excel 2002 (10.2701.2625). Korektor českého pravopisu © 1995-2000 Lingea s.r.o.
- [44] PACHMANN Eduard. *Obecná didaktika chemie IV*. Univerzita Karlova v Praze, fakulta pedagogická. Praha: SPN 1981.
- [45] PODROUŽEK L.: *Prvouka a přírodověda s didaktikou IV (neživá příroda)*. Západočeská univerzita v Plzni, Plzeň 2001.
- [46] PROKŠA, Miroslav. *Manuálně a intelektové zručnosti, návyky a spôsobilosti podmieňujúce organizáciu experimentálnej činnosti vo vyučovaní chémie*. Eseje na téma Standard kvalifikace

- učitele chemie, s. 31-35. Podpora rozvoje učitel'ských vzdělávacích programů a jiných vzdělávacích aktivit 2001. Univerzita Karlova v Praze, Pedagogická fakulta. Praha 2001.
- [47] PRŮCHA, J.: Pedagogická evaluace. MU v Brně, Brno 1996. ISBN 80-210-1333-8.
- [48] PŮLPÁN, Zdeněk. K problematice zpracování empirických šetření v humanitních vědách. Praha: Academia, nakladatelství Akademie věd České republiky, 2004. ISBN 80-200-1221-4
- [49] PŮLPÁN, Zdeněk. *Základy sestavování a klasického vyhodnocování didaktických testů*. Hradec Králové: Kotva, 1991. ISBN 80-900254-4-7.
- [50] PŮLPÁN, Zdeněk – KUŘINA, František – KEBZA, Vladimír. O představitosti a její roli v matematice. 1. vyd. Praha: Academia, nakladatelství Československé akademie věd, 1992. ISBN 80-200-0444-0.
- [51] REGULI, Ján. *Fyzikálna chémia v príprave učiteľou chémie*. Profil učitele chemie II, 87 - 90. Gaudeamus, Hradec Králové 2002.
- [52] REGULI, Ján. *Fyzikálna chémia v učive základnej a strednej školy*. Šoltésove dni, Zborník MCMB, Bratislava 2002.
- [53] REGULI, Ján. *Neformálne vzdelávanie v oblasti chémie*. 1. vyd. Bratislava: Slovenská technická univerzita v Bratislave, 2001. ISBN: 80-227-1553-0.
- [54] REKTORYS, Karel, et. al. *Přehled užité matematiky*. 3., nezměněné vyd. Praha: SNTL – Nakladatelství technické literatury, 1973. 33. Počet pravděpodobnosti, s. 1025-1056.
- [55] RÖTLING, Gustáv. *Metodika tvorby učitel'ského didaktického testu*. Metodické centrum Banská Bystrica, 1996. ISBN 80-8041-110-7.
- [56] RYCHTERA, Jiří. Požadavky „pedagogické praxe“ jako základ kvalifikačního standardu učitele chemie? Eseje na téma Standard kvalifikace učitele chemie, s. 36-41. Podpora rozvoje učitel'ských vzdělávacích programů a jiných vzdělávacích aktivit 2001. Univerzita Karlova v Praze, Pedagogická fakulta. Praha 2001.
- [57] ŘEŠÁTKO, Miloš. *Didaktické testy ve školní praxi*. Praha: SNTL-Nakladatelství technické literatury, 1975.
- [58] SMĚKAL, Vladimír – ŠVEC, Vlastimil – ZAJAC, Jaroslav. *Didaktické testy a jejich vyhodnocování*. Brno: Středisko pro výzkum učebních metod a prostředků, 1973.
- [59] STEINHAUS H. *Tablica liczb przetasowanych czterocyfrowych*. Rozprawy Matematyczne VI. 1954.
- [60] SVOBODOVÁ Miroslava. Standard kvalifikace učitele chemie. Eseje na téma Standard kvalifikace učitele chemie, s. 42-44. Podpora rozvoje učitel'ských vzdělávacích programů a jiných vzdělávacích aktivit 2001. Univerzita Karlova v Praze, Pedagogická fakulta. Praha 2001.
- [61] ŠKRÁŠEK, Josef. *Úvod do počtu pravděpodobnosti a matematické statistiky*. 2. přeprac. vyd. Brno: VA AZ, 1971.
- [62] ŠVEC, Vlastimil – FILOVÁ, Hana – ŠIMONÍK, Oldřich. *Praktikum didaktických dovedností*. Brno: Pedagogická fakulta Masarykovy univerzity v Brně, 1996. ISBN 80-210-1365-6. Cvičení 2, Výukové cíle, s. 22-31.
- [63] ŠVEC, Vlastimil – FILOVÁ, Hana – ŠIMONÍK, Oldřich. *Praktikum didaktických dovedností*. Brno: Pedagogická fakulta Masarykovy univerzity v Brně, 1996. ISBN 80-210-1365-6. Cvičení 6, Učební úlohy a otázky, s. 53-61.
- [64] TIPPETT L.H.C. *Random Sampling Numbers*. Tracts of Computers XV, Cambridge University Press, 1927.
- [65] TOLLINGEROVÁ D.: K teorii učebních činností. SPN, Praha 1986.
- [66] TOMEČEK, Otto. Štandard Kvalifikácie učiteľa chémie. Eseje na téma Standard kvalifikace učitele chemie, s. 36-41. Podpora rozvoje učitel'ských vzdělávacích programů a jiných vzdělávacích aktivit 2001. Univerzita Karlova v Praze, Pedagogická fakulta. Praha 2001.
- [67] TUREK I.: *Učiteľ a didaktické testy*. Bratislava: Metodické centrum mesta Bratislavy, 1996. 1. vyd. ISBN 80-7164-139-1.
- [68] VAŠUTOVÁ J.: Být učitelem. Co by měl učitel vědět o své profesi. PedF UK v Praze, Praha 2002. ISBN 80-7290-077-3.
- [69] Doc. PhDr. Jiří DAN, CSc. Ústní sdělení. 15. 12. 2004.