

Univerzita Karlova v Praze

Mgr. Hana Cídllová, Dr.

Návrh principu testu dovednosti
práce s vědomostmi a jeho využití
v oblasti obecné a fyzikální chemie

Habilitační práce

Praha, 2006

2 Didaktické testy

2.1 Typy didaktických testů

(zpracováno podle [26], 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. **Standardizací šetření** Půlpán [27] rozumí přesnou formulaci požadavků jak obsahových, tak i realizačních a vyhodnocovacích. Tento termín se liší od pojmu **standardizace testu**, jehož smyslem je podle Chráska [25] vytvoření tzv. testového standardu (testové normy), který umožní zařadit respondenta podle dosaženého počtu bodů do určitého žebříčku (stupnice, škály).

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, jak ukazují následující příklady:

S. Vrána [28] rozděluje testy:

1. **Podle toho, co zjišťují:**
 - a) inteligenční, zaměřené na zjišťování schopností zkoumaných osob.
 - b) didaktické, sloužící ke zjišťování vědomostí a dovedností, které žák získal učením.
2. **Podle způsobu zpracování:**
 - a) informační, sestavené učitelem pro zjištění, jak žáci ovládli učivo.
 - b) 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:**
 - a) zkušební, sloužící k hodnocení vědomostí žáků.
 - b) diagnostické, zaměřené na zjišťování stavu vědomostí žáků a na zjišťování příčin nedostatků.
 - c) kontrolní, sloužící žákům k autokontrolě.
4. **Podle způsobu užití:**
 - a) hromadné, kterými je prověřován větší počet žáků najednou.
 - b) individuální, k ověřování vědomostí a dovedností jednotlivých žáků.

M. Michalička [29] testy klasifikuje podle principů konstrukce na tři skupiny:

1. 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.
2. testy objektivně informační – sestavené školským pracovníkem podle zásad pro přípravu testů standardizovaných. Mají zejména funkci kontrolní a klasifikační.
3. testy standardizované – sestavené podle závazně stanovených a přísně dodržovaných konstrukčních principů. Jsou opatřeny propracovanými tabulkami norem. Zpravidla jde o týmovou práci odborníků.

P. Byčkovský (tuto klasifikaci přejímá i Chráska [25]) třídí didaktické testy podle osmi hledisek:

1. Podle měření charakteristiky výkonu:

- a) testy rychlosti – měří rychlost, kterou je žák schopen provést určitý výkon.
- b) testy úrovně – zjišťují kvalitu (úroveň) výkonu.

2. Podle stupně dokonalosti přípravy testu a jeho příslušenství:

- a) testy standardizované – jsou sestavovány zpravidla týmy odborníků, v nichž je vedle specialistů daného vyučovacího předmětu také statistik, psycholog, metodik, učitel ze školní praxe 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) – jsou připravovány učiteli pro jejich vlastní 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 – částečně standardizované nebo důkladněji provedené nestandardizované testy.

3. Podle povahy činnosti testovaného:

- a) testy kognitivní – měří úroveň (kvalitu) poznávacího procesu žáků.
- b) testy psychomotorické – zjišťují výsledky psychomotorického učení (např. psaní na stroji; v chemii by mohl takový test měřit např. rychlost a přesnost pipetování).

4. Podle míry specifičnosti učení zjišťovaného testem:

- a) testy výsledků výuky – měří to, co se žáci v dané oblasti naučili.
- b) testy studijních předpokladů – měří úroveň obecnějších charakteristik jedince potřebných k dalšímu studiu (např. intelektovou výkonnost).

5. Podle toho, jakým způsobem se interpretuje výkon testovaného:

- a) testy rozlišující – vzhledem k populaci testovaných.
- b) testy ověřující – vzhledem k souboru úloh reprezentujících učivo.

6. Podle časového zařazení do výuky:

- a) vstupní testy – zadávané na začátku výuky určitého tematického celku.
- b) průběžné testy – zadávané v průběhu výuky
- c) výstupní testy – zadávané na konci tematického celku učiva nebo na konci výukového období (pololetní, celoroční).

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 nich není možno stanovit jednoznačný předpis pro skórování.

Mužic v [30] dělí testy používané ve školách na:

- **testy didaktické** (testy školních znalostí):
 - o testy vědomostí v užším smyslu (zjišťují znalost faktů)
 - o testy schopnosti aplikovat získané vědomosti
- **testy schopností** – zjišťují některé předpoklady dosažení úspěchu v určité činnosti:
 - o testy senzorické
 - o testy mentální
 - o testy mechanických schopností
 - o testy motorických dovedností

- **testy osobnosti v užším smyslu** – zkoumají některé složitější vlastnosti jednotlivce, např. jeho postoje, zájmy, charakteristické rysy jeho aktivity a citových projevů.

Lapitka [31] testy klasifikuje podle způsobu tvorby položek na:

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

Existují i jiné klasifikace testů (např. z hlediska oprávnění osob s různým vzděláním testy používat, nebo z hlediska toho, jaké skupině respondentů je daný test určen).

2.2 Typy testových položek

2.2.1 Typy testových položek podle formy

Didaktický test je sestaven z jednotlivých testových položek, které mohou mít různou formu. Následuje několik způsobů klasifikace položek podle formální stránky:

M. Michalička [29] 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 textu 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 (než dvou) možností.
5. Položky tzv. dvoustranného výběru (k určitému souboru výrazů, názvů apod. se přihazuje správná **odpověď z nabízených variant**).

P. Byčkovský (uvedeno v [26]) a podle něj **M. Chráska** [25] dělí položky na:

1. **Otevřené** (úlohy s tvořenou neboli volnou odpovědí)
 - a) otevřené široké úlohy, u kterých se očekává pojednání na určité téma, vyřešení zadaného problému, popis určitého procesu,...
 - b) úlohy se stručnou odpovědí – očekává se doplnění značky, symbolu, případně 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.
 - b) úlohy polytomické (s výběrem odpovědi, multiple-choice). Optimální počet nabídnutých alternativ se uvádí 4-5, praxe se ustálila u čtyř nabízených odpovědí. Nabízené odpovědi mohou mít několik forem:
 - vybírá se jedna správná odpověď
 - vybírá se několik správných odpovědí. Tato forma komplikuje binární skórování výsledků. Respondenti musejí být na tento typ otázek předem upozorněni.

Kromě uvedených základních dvou typů položek (**otevřené**, **uzavřené**) uvádějí někteří autoři (pro dotazníky např. **Lapitka** [32]) také položky **polootevřené**. Takové položky umožňují, aby si respondent volil z připravených variant nebo odpovídal volně, mimo naznačené možnosti.

Turek [24] dělí testové položky na **otevřené** a **uzavřené**. Otevřené pak dále dělí na položky se širokou odpovědí (nestrukturalizované a strukturalizované) a s odpovědí stručnou (položky produkční a doplňovací). Uzavřené položky klasifikuje na dichotomické, položky s výběrem odpovědi, položky přiřazovací a uspořádací).

Byčkovský (uvedeno v [26]) a Chráska [25] považují za **úlohy objektivního typu** takové úlohy, které lze objektivně skórovat, tzn. úlohy otevřené se stručnou odpovědí a úlohy uzavřené.

G. Rötling (Metodické centrum Banská Bystrica) [33] klasifikuje testové položky obdobně jako M. Chráska [25]:

- **Otevřené** (odpověď je respondentem tvořena), tzv. produkční
 - subjektivně skórovatelné (široká odpověď s vymezenou strukturou):
 - Vysvětlete, 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 alternativ)
 - 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 údajům v levém sloupci přiřadit odpovídající údaje z pravého sloupce*
 - výběr uspořádání podle určitého kritéria, např.
 - Seřadit podle velikost vlastností, podle pořadí úkonů, ...*

2.2.2 Typy testových položek podle náročnosti myšlenkových operací

Jednotlivé testové položky je možné klasifikovat nejen podle formy položek, ale také podle náročnosti myšlenkových operací, které jsou od respondenta při řešení vyžadovány. Velmi známá a používaná je **Bloomova taxonomie** učebních úloh, která úlohy člení podle náročnosti kognitivních cílů do šesti kategorií (hierarchických úrovní). Ke každé kategorii jsou přiřazena určitá slovesa, která vymezují, co mají žáci dělat (tab. 1).

Cílová kategorie (úroveň osvojení)	Typická slovesa a jejich vazby používané k vymezení cílů
Zapamatování	definovat, doplnit, napsat, opakovat, pojmenovat, popsat, přiřadit, reprodukovat, seřadit, vybrat, vysvětlit
Pochopení (porozumění)	dokázat jinak formulovat, ilustrovat, interpretovat, objasnit, odhadnout, opravit, převést, vyjádřit vlastními slovy, vysvětlit, vypočítat, zkontrolovat
Aplikace	aplikovat, demonstrovat, diskutovat, interpretovat, načrtnout, navrhnout, použít, prokázat, registrovat, řešit, uvést vztah, uspořádat
Analýza	analyzovat, provést rozbor, rozhodnout, rozlišit, rozčlenit, specifikovat
Syntéza	kategorizovat, klasifikovat, kombinovat, modifikovat, napsat sdělení, organizovat, reorganizovat, shrnout, vytvořit obecné závěry
Hodnotící posouzení	argumentovat, obhájit, ocenit, oponovat, podpořit (názory), porovnat, provést kritiku, posoudit, prověřit, srovnat s normou, uvést klady a zápory, zdůvodnit, zhodnotit

Tab. 1: Bloomova taxonomie a slovník aktivních sloves používaných k vymezení cílů vyučování. Zpracováno podle [34], kde je odvolání na [35].

V rámci jednotlivých kategorií existuje ještě jemnější členění.

D. Tollingerová navrhla podle Bloomova třídění vyučovacích cílů taxonomii učebních úloh, rozdělených podle operací nutných k jejich řešení. Učební úlohy jsou podle Tollingerové rozděleny do pěti kategorií, obsahujících 27 typů učebních úloh. Toto dělení je možno nalézt např. v [35].

Známa je také **Niemierkova taxonomie výukových cílů**, kterou např. Chráska [25] pokládá za výhodnou pro plánování didaktických testů (tab. 2).

A	Zapamatování poznatků Žák je schopen vybavit si určitá fakta (např. termíny, zákony), přičemž je nesmí mezi sebou zaměňovat. <i>Typická aktivní slovesa:</i> 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. <i>Typická aktivní slovesa:</i> 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. <i>Typická aktivní slovesa:</i> 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. <i>Typická aktivní slovesa:</i> rozhodnout, provést rozbor, kombinovat, vyvrátit, obhájit, prověřit, zhodnotit, posoudit.

Tab. 2: Niemierkova taxonomie výukových cílů. Převzato z [25].

Forma položek ve vztahu k testovaným výukovým cílům

V Turkově knize [24] se diskutuje vhodnost různých forem položek didaktického testu k měření jednotlivých úrovní učení, jak shrnují tabulky tab. 3 a tab. 4.

forma položky	zapamatování	porozumění	aplikace	analýza	syntéza	hodnocení
široké nestrukturalizované	-	-	+	+	++	++
široké strukturalizované	-	+	++	++	+	++
produkční	++	++	++	+	-	-
doplňovací	++	+	+	-	-	-
dichotomické	++	++	+	-	-	-
s výběrem odpovědi	+	++	++	-	-	-
přiřazovací	++	++	+	+	-	+
uspořadací	+	++	-	-	-	+

Tab. 3: Vhodnost různých forem položek didaktického testu k měření jednotlivých úrovní učení podle taxonomie cílů B. S. Blooma.

- ... nevhodné, + ... středně vhodné, ++ ... velmi vhodné

forma položky	zapamatování	porozumění	specifický transfer	nespecifický transfer
se širokou odpovědí	-	-	+	++
produkční	++	++	+	-
doplňovací	++	+	+	-
dichotomické	++	++	+	-
s výběrem odpovědi	+	++	++	+
přiřazovací	++	++	+	-
uspořadací	+	++	-	-

Tab. 4: Vhodnost různých forem položek didaktického testu k měření jednotlivých úrovní učení podle taxonomie cílů B. Niemierka.

- ... nevhodné, + ... středně vhodné, ++ ... velmi vhodné

Podle tab. 3 by se k měření aplikační dovednosti nejlépe hodily položky široké strukturalizované, produkční a položky s výběrem odpovědi. Podle Niemierka lze aplikace zařadit do úrovně C a D (specifický a nespecifický transfer). Nejvhodnější formou položek testu pro kontrolování schopnosti nespecifického transferu by měly být podle tab. 4 položky se širokou odpovědí a položky s výběrem odpovědi. Položky produkční jsou (na rozdíl od tab. 3) pokládány za nevhodné.

2.3 Základní kroky tvorby didaktického testu

Příprava kvalitního didaktického testu by měla probíhat v několika krocích, jak naznačuje následující přehled:

1. specifikace cílů
2. tvorba položek
3. kontrola položek nezávislymi odborníky
4. administrace pretestu
5. vyhodnocení pretestu, úprava znění položek, zařazení distraktorů, úprava počtu položek
6. kontrola nové verze nezávislymi odborníky
7. nová administrace testu
8. korektury testu
9. administrace testu velkému počtu respondentů, standardizace

2.3.1 Zásady pro navrhování testových položek

Problematika zásad pro navrhování testových úloh je podle názoru autorky habilitační práce výstižně zpracována v učebnicích Chráska [25], Kohoutka [26], Turka [24] a Rötlinga [33], z nichž proto autorka vycházela při zpracování kapitoly č. 3.4.1.

2.3.1.1 Obecné zásady pro navrhování všech druhů testových položek

- Testovat se má důležité učivo, které bylo předmětem výuky.
 - Text položek musí být gramaticky správný.
 - Věnovat dostatečnou pozornost i grafické stránce položek.
 - V textu se nemají vyskytovat nejasná slova, zadání položek musí být jednoznačné.
 - Vyhýbat se položkám, jejichž řešení je dosud předmětem vědecké diskuze.
 - Položky nechat vždy posoudit kompetentními osobami.
 - Vyhýbat se doslovným formulacím z učebnic (podporovalo by to formalismus v získávání vědomostí).
 - Položky nesmějí být sestaveny tak, aby ke správné odpovědi postačoval logický úsudek o formální stránce úkolu.
 - Řešení jedné položky nesmí umožňovat nebo podmiňovat řešení některé z dalších položek.
 - Nikdy nepoužívat *chytáky*, které místo stupně zvládnutí učiva zkouší postřeh, vtíp atd.).
 - Je třeba používat spíše kvantitativních než kvalitativních výrazů.
 - Vyhýbat se zbytečným zdrojům obtížnosti (např. zbytečně nezadávat složité číselné hodnoty).
 - Používat pokud možno jednoduchého (binárního) skórování.
 - Navrhovat položky v dostatečném počtu (pro tzv. předvýzkum je vhodné navrhovat položek více, než kolik má obsahovat konečná podoba testu).
- Hališka [36] 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).

2.3.1.2

Konstrukce otevřených úloh se stručnou odpovědí

- Otevřené položky užívat jen tehdy, lze-li odpovědět velmi stručně (nejlépe jediným údajem).
- Nevyžadovat doslovné opakování textu z učebnice.
- Zvážit předem všechny možné odpovědi, a je-li jich mnoho, raději položku nepoužívat.
- Ponechat v odpovědním archu dostatek místa pro uvedení odpovědi.
- Dávat přednost produkčním položkám před doplňovacími. Při použití doplňovacích položek dodržovat následující doporučení:
 - Vynechat jen důležité údaje.
 - Z neúplné věty musí být jednoznačně patrné, co se má doplnit.
 - Údaj, který se má doplnit, umístit pokud možno na konec věty.

2.3.1.3

Konstrukce testových úloh s výběrem odpovědi

Položky s výběrem odpovědi vyžadují kromě formulace zadání také vytvoření vhodných nabídek špatných odpovědí (tzv. distraktorů). Mělo by 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.

Návrh vhodných distraktorů je velkým problémem. Při jejich navrhování se většinou vychází z logické úvahy nebo zkušeností s nejčastějšími chybami. U pečlivě připravovaných testů se někdy postupuje tak, že se úloha respondentům zadá nejdříve jako otevřená a pak se nejčastějších chyb použije jako distraktorů. Takto postupovala i autorka habilitační práce.

2.3.1.4

Zásady pro konstrukci úloh s výběrem odpovědi

- Položkami s výběrem odpovědi nezkoušet pokud možno zapamatování konkrétních poznatků. K tomu se lépe hodí položky produkční.
- Nejvhodnější počet nabídek je 4-5, praxe se ustálila na čtyřech možnostech (viz dále).
- Soubor nabídek má být homogenní, tj. podobný obsahovým zaměřením i formou.
- Při formulaci položek s výběrem odpovědi dávat přednost otázkám před neúplnými tvrzeními.
- Ve formulaci úlohy se vyhýbat slovům nebo údajům, které by mohly sloužit jako nápověda.
- Pokud se ve formulaci úlohy vyskytuje zápor, je nutno jej zvýraznit např. podtržením.
- V úloze s výběrem nejlepší odpovědi je zapotřebí tento fakt zdůraznit vhodnou formulací.
- Distraktory se nesmějí navzájem překrývat nebo jinou formou vyjadřovat totéž.
- Distraktory musí být jednoznačně nesprávné nebo musí obsahovat nepostačující odpověď.
- Umístění správné odpovědi mezi distraktory se má volit zcela náhodně.
- Navrhovat jen takové distraktory, u nichž je předpoklad, že budou využívány.
- Při používání úloh s vícenásobnou volbou odpovědi a při používání neurčitých odpovědi na tuto skutečnost respondenty upozornit.
- Vyhýbat 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 neměly používat. Odpověď *nic z uvedeného* je však některými jinými autory, např. Turkem, naopak doporučována.

2.3.1.5

Problém náhodné volby správné odpovědi

U uzavřených úloh existuje vždy určitá pravděpodobnost, že žák správnou odpověď zcela náhodně uhodne. 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 úlohu nepřehlednou a také sestavování většího počtu distraktorů 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 jedině tehdy, když odpověď zná. V Chráskově [25] je uveden vztah pro úpravu skóre pomocí korekce na hádání. Z tohoto vztahu však plyne, že se dá použít jen tehdy, pokud všechny položky v testu jsou polytomické se stejným počtem nabídek v každé položce. Jestliže se provádí korekce na hádání, je nutno na tuto skutečnost žáky předem upozornit. Zkušenosti 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 tím, že jako podmínka pro uznání odpovědi je požadován výpočet nebo zdůvodnění odpovědi. Pravděpodobnost uhodnutí správné odpovědi lze v některých případech snížit také užitím tzv. neurčitých odpovědi (*nic z uvedeného*).

2.3.1.6

Některé zásady pro konstrukci přiřazovacích úloh

- navrhovat homogenní soubor nabídek a doplňků,
- nenabízet více než 8-10 nabídek a doplňků,
- poskytnout více nabídek než doplňků (aspoň o jednu); např. Hališka [36] ovšem o takovém požadavku nehovoří,
- doplňky formulovat výstižně a stručně,
- doplňky a nabídky uvádět v oddělených sloupcích,
- doplňky a nabídky značit odlišně (např. jedny značit písmeny a druhé číslovat).

2.3.2 Počet položek v testu

Počet položek v testu se musí odvíjet především od dvou základních faktorů: spolehlivosti (**reliability**) testu a **času** potřebného k vyřešení testu. Také **věk respondentů** a jejich **intelektuální úroveň** jsou činiteli ovlivňujícími počet položek didaktického testu. Počet položek v testu by také měl respektovat způsob, jakým test má být vyhodnocen.

Větší počet položek umožňuje větší spolehlivost (reliabilitu) testu. Podle Smékala [22]: *Na počet položek není mezi různými autory jednotný názor*. Seriozní standardizovaný didaktický test by podle některých autorů neměl mít méně než 40 položek. Mají-li ovšem být respondentům zadány úlohy myšlenkově náročnější a je-li zadavatel testu současně omezen časem, může tento požadavek být nereálný.

Dalším důležitým a často limitujícím faktorem, určujícím počet položek v testu, je čas, který je možno pro testování vymežit. Časově náročnější testy mívají čistý testovací čas 35-40 minut. Rötling [33] s odvoláním na psychohygienické požadavky udává 40 minut jako horní časový limit. Turek [24] jako horní časovou hranici pro žáky 2. stupně základní školy uvádí 40 minut, pro žáky střední školy připouští i delší čas. Mužic [30] žádá,

aby u testů měřících úroveň vědomostí žáků byl limitován čas tak, aby 90 % žáků stačilo testem projít. Řešátko [23] doporučuje určovat optimální čas na vypracování testu jako průměrný čas, který žáci při vypracovávání 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.

Meloun a Militký [37] doporučují odlišný způsob statistického zpracování výsledků pro malé soubory do 20 měření (tzv. Hornův postup; rozhovorem s několika pedagogicky-pracovníky i odborníky si však autorka ověřila, že tento postup nepatří ke všeobecně známým a přijímaným metodám statistického vyhodnocení dat). Pro výpočet reliability (kap. 2.4.4.1) je např. pro výpočet pomocí Spearmanova-Brownova vzorce (3) nutný sudý počet položek.

Z výše uvedeného plyne, že chce-li autor test vyhodnotit klasickými (nebo aspoň všeobecně známými) statistickými postupy (např. používat aritmetický průměr) a určit reliability testu metodou půlení, měl by test obsahovat sudý počet položek, a to nejméně 22.

2.4 Charakteristiky didaktického testu

Didaktické testy používané ve výzkumu musí být objektivní, standardní, přiměřeně obtížné, spolehlivé, platné (validní), kvantitativně i kvalitativně interpretovatelné, senzibilní (citlivé), úsporné (ekonomické) a administrované všem zkoumaným osobám (respondentům) stejně.

2.4.1 Objektivnost

Test je objektivní (tj. nezávislý na autorovi), jestliže:

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

Lapitka [31] doporučuje provést zkoušku objektivnosti testu tak, že se test předloží několika nezávislým odborníkům. Objektivnost pak kvantifikuje podle hodnoty výrazu:

$$O = \frac{X_{\text{odb}}}{S_{\text{max}} \cdot n_{\text{odb}}} \cdot 100 \% \quad (1)$$

kde

O objektivnost testu

X_{odb} celkový počet bodů získaný odborníky

S_{max} celkový počet bodů získatelný za test (nejvyšší možné skóre)

n_{odb} počet odborníků, kteří test řešili

Lapitka ve stati o objektivnosti testu cituje Ebela [38], který ovšem v citované části nehovoří o objektivnosti testu, nýbrž o objektivnosti jednotlivých položek.

Autorka habilitační práce se domnívá, že při přípravě testu je vhodnější posuzovat přímo objektivnost jednotlivých položek (o porovnávání řešení poskytnutého nezávislými odborníky u každé jednotlivé položky Lapitka sice hovoří, avšak počítá pouze jedno číslo pro celý test). Při aplikaci vztahu (1) 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 (takové položky by z testu měly být vyřazeny), zatímco na ostatní odpoví zcela shodně.

2.4.2 Standardnost šetření

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

Snadno zaměnitelný termín *standardizace testu* je diskutován v kapitole 2.5.2.

2.4.3 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.

2.4.4 Reliabilita (spolehlivost)

Reliabilitou (spolehlivostí) didaktického testu Půlpán [27] rozumí míru jeho stability, tj. schopnosti poskytovat stejné výsledky při opakování za stejných podmínek. Reliabilita je nutnou podmínkou validity. Testy, které mají vysokou reliabilitu, jsou málo ovlivněny náhodnými vlivy (např. momentální psychosomatickou kondicí, vnějšími podmínkami). Reliabilita se kvantitativně vyjadřuje různými způsoby, jejichž volba, kromě jiného, závisí i na konstrukci a druhu testu. Přehled způsobů kvantifikace reliability udává např. Lapitka [31], Chráska [25] nebo Turek [24].

2.4.4.1 Způsoby stanovení reliability

Reliabilitu testu 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** (obsahujících různé položky, ale z téhož učiva). Tato metoda se nazývá také **metoda dvojnásobného testu** nebo **metoda dvojice ekvivalentních testů**. Sestavení paralelních forem testu s identickými parametry je velmi obtížné a proto Švec [22] tento postup nedoporučuje. Uvedenou metodu naopak doporučuje např. Lapitka [31]. 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). Pro výpočet tzv. **indexu reliability** touto metodou uvádí Lapitka vztah:

$$I_{rel} = 1 - \frac{N_{odl}}{N_{celk}} \quad (2)$$

kde

I_{rel} index reliability

N_{odl} počet neshodně 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** (doporučuje Lapitka [31] a Chráska [25]).

Reliabilita testu metodou půlení:

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 tak, 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 takto určeného korelačního koeficientu se pak vychází při výpočtu koeficientu reliability pomocí **Spearmanova-Brownova vzorce**

$$r_{sb} = \frac{2r_p}{1 + r_p} \quad (3)$$

kde

r_{sb} koeficient reliability vypočtený pomocí Spearmanova-Brownova vzorce
 r_p Pearsonův párový korelační koeficient (4) pro výsledky v obou polovinách testu

Půlpán [39] 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. Hnilíčková [40] uvádí vztah (3) 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).

Koeficient korelace pro dosazení do vztahu (3) se vypočte podle vztahu:

$$r_p = \frac{n \sum_{i=1}^n (X_{iL} X_{iS}) - \sum_{i=1}^n X_{iL} \sum_{i=1}^n X_{iS}}{\sqrt{\left[n \sum_{i=1}^n (X_{iL}^2) - \left(\sum_{i=1}^n X_{iL} \right)^2 \right] \cdot \left[n \sum_{i=1}^n (X_{iS}^2) - \left(\sum_{i=1}^n X_{iS} \right)^2 \right]}} \quad (4)$$

kde

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

Chráska uvádí výpočet koeficientu reliability pomocí tzv. **Kuderova-Richardsonova vzorce 20** (tzv. **KR 20**)

$$k_{kr} = \frac{N}{N-1} \left(1 - \frac{\sum_{j=1}^m p_j (1-p_j)}{s_s^2} \right) \quad (5)$$

kde

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

N počet položek

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

s_s směrodatná odchylka skóre celého testu

Chráska ve shodě s Lapitkou poznamenává, že tento způsob výpočtu se hodí pouze pro testy složené z obsahově homogenních úloh.

Půlpán [39] uvádí i zjednodušení KR 20 na KR 21 pro binárně skórovany test tvořený položkami přibližně stejné obtížnosti:

$$k_{kr} = \frac{N}{N-1} \left(1 - \frac{\bar{S} - \frac{\bar{S}^2}{N}}{s_s^2} \right) \quad (6)$$

kde

\bar{S} průměrné skóre dané skupiny respondentů; ostatní symboly viz výše.

Např. Turek [24] uvádí vztah pro výpočet koeficientu reliability **Cronbachovou metodou**. Vztah je určen pro výpočet koeficientu reliability u testů s položkami skórovanými váženě (složeně) – kap. 2.5.1.

$$r_c = \frac{N}{N-1} \left(1 - \frac{\sum_{j=1}^m s_j^2}{s_s^2} \right) \quad (7)$$

kde

r_c koeficient reliability vypočtený pomocí Cronbachovy metody

N počet položek

s_s směrodatná odchylka skóre celého testu

s_j směrodatná odchylka j-té položky

Internetové stránky [41] uvádějí dnes široce využívané **Cronbachovo alfa**. Je zde řečeno, že pro dichotomické položky je totožné s hodnotou KR-20 (5).

2.4.4.2 Doporučená hodnota reliability

Hodnota koeficientu reliability se pohybuje mezi $0 \leq r \leq 1$. Názory na potřebnou minimální hodnotu reliability se mezi odborníky různí. Hodnota koeficientu reliability podle Řešátka [23] má být u dobrých testů $r \geq 0,9$. Podle Chrásky [25] se pro individuální pedagogickou diagnostiku většinou požaduje koeficient reliability minimálně 0,80. Podle Kohoutka [26] je však u nestandardizovaných testů za postačující považována hodnota koeficientu reliability 0,6, ve Smékalově práci [22] se za uspokojivou hodnotu pokládá koeficient korelace mezi oběma polovinami testu o hodnotě alespoň 0,8. Lapitka [31] doporučuje, aby učitel nepoužíval test s indexem reliability nižším než 0,75.

Obecně platí, že s rostoucím počtem (stejně kvalitních) položek reliability testu roste. Test s nízkou reliability je tedy možno (v určitém rozmezí) upravit přidáním položek tak, že jeho reliability tím vzroste. Potřebný počet položek se dá odhadnout pomocí vztahu uvedeného ve Smékalově práci [22]. Pokud test má $r \leq 0,6$, je koeficient reliability podle Řešátka [23] tak nízký, že nemá smysl se zabývat dalším zlepšováním testu.

2.4.5 Validita

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

Jak uvádí Lapitka [31] s odkazem na Ebela [38], v odborných příručkách se uvádí řada různých druhů validity (Ebel v citované práci uvádí až 10 druhů validity, které klasifi-

kuje do dvou kategorií: **přímá** validita a **odvozená** validita). Turek [24] uvádí např.:
Obsahová validita (typ validity přímé) – většinou se jí rozumí shoda mezi obsahem testu a učební látkou (u testů ověřujících studijní výsledky).

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 výsledkem testu a jiným měřítkem žakových vědomostí. Srovnávacím ukazatelem je často klasifikace (známka).

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

Chráška [25] konstatuje: *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.*

2.4.6 Kvalitativní a kvantitativní interpretovatelnost

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

Možností **kvantifikace** se rozumí, 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).

2.4.7 Senzibilita (citlivost) testu

(zpracováno podle Hniličkové [40], není-li uvedeno jinak)

Pojem *citlivost testu* je nutno rozlišovat od pojmu *citlivost položky* (kap. 2.5.3). Senzibilita (citlivost či selektivnost) didaktického testu charakterizuje, nakolik test umožňuje zjistit rozdíly v rozsahu a kvalitě vědomostí či dovedností respondentů.

Teoretickým modelem rozdělení psychických znaků osobnosti je většinou tzv. normální rozdělení graficky zobrazené Gaussovou křivkou, přičemž citlivějším testům přísluší plošší křivky (větší směrodatná odchylka). Jak však připomíná Lapitka [31], pedagogické jevy nejsou náhodné a proto nemusejí být distribuované normálně. Z těchto důvodů je nutno normalitu rozdělení výsledků didaktického testování před podrobnějším rozбором nejprve vhodnou metodou ověřit. O tom, zda naměřené hodnoty předpokladu normálního rozložení vyhovují, je možno na zvolené hladině významnosti rozhodnout např. pomocí **testu kombinace výběrové šikmosti a špičatosti dle Jarque-Berra** nebo pomocí **Pearsonova testu** neboli **testu χ^2 dobré shody** (obojí kap. 3.3.1).

2.4.8 Úspornost

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

2.4.9 Administrace didaktického testu

Podle Hniličkové [40] se administrace (zejména standardizovaných) testů týká technických podmínek postupu při testování, vyhodnocování a interpretaci výsledků testování. Veškeré informace o testu by měly být shromážděny v testové příručce (imatrikulacním 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 respondentů
 - o charakteristika respondentů
- 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: všichni respondenti musí mít:
 - o stejné instrukce (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 mohou psát poznámky, jaké mají dovolené pomůcky – např. kalkulačka, tabulky apod.)
 - o stejné znění didaktického testu a stejný záznamový (odpovědní) list
 - o pokud možno stejné podmínky k vypracování testu (hlučnost místnosti, přiměřené osvětlení, přiměřená teplota, větrání, denní doba – nejlepší výkon podávají žáci obvykle v době 1.-2. vyučovací hodiny, délka trvání testu...)
 - o stejné zařazení testu s ohledem na probírané učivo
 - o stejná předběžná informace o testování, případně o jeho obsahu
 - o stejný způsob a dobu odpovídání na dotazy respondentů
- informace pro vyhodnocení testu
 - o autorské řešení
 - o způsob hodnocení jednotlivých položek
 - o způsob hodnocení celého testu (norma hodnocení, případně klasifikace)

2.5 Analýza didaktických testů

Při zpracování výsledků didaktického testování se obvykle provádí kvalitativní a kvantitativní analýza. Přehled obou typů analýz je v následujícím textu zpracován dle Lapitky [31].

- **Kvalitativní analýza** je rozbor jednotlivých úloh (rozbor nejčastějších chyb, kterých se respondenti při řešení testu dopouštěli, pokus o odhalení jejich příčin apod.)
- **Kvantitativní analýza** je obvykle 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).
- **Skóre:** Součet všech bodů, které respondent při vypracování testu získal, se nazývá celkové skóre (individuální skóre, hrubé skóre); viz též dále.

2.5.1 Skórování odpovědí

Hodnocení odpovědí na testové položky neboli skórování odpovědí se provádí většinou pomocí bodů. Skórování může být

1. **nevážené (binární):** jeden bod je přidělen 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,
2. **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 habilitační práce si nekladla za cíl studenty hodnotit ani vytvořit test, který by klasifikaci umožňoval, nebude se touto problematikou dále zabývat.

Hrubé skóre lze obecně vypočítat podle vztahu:

$$S_i = \sum_{j=1}^N X_{j,i} \quad (8)$$

kde

S_i hrubé skóre, které získal i -tý respondent

i číslo označující respondenta

j číslo položky

$X_{j,i}$ hodnocení j -té položky řešené i -tým respondentem

V případě binárního skórování:

$X_{j,i} = 1$ znamená položku řešenou správně,

$X_{j,i} = 0$ znamená položku neřešenou nebo řešenou špatně

N počet položek

Průměrné skóre určité skupiny respondentů (úspěšnost skupiny respondentů):

$$\bar{S} = \frac{\sum_{i=1}^n S_i}{n} \quad (9)$$

kde

\bar{S} průměrné skóre určité skupiny respondentů

S_i hrubé skóre, které získal i -tý respondent

i číslo označující respondenta

n počet respondentů

- **Celkové skóre a klasifikace:** Častým cílem didaktických testů je získání podkladů pro klasifikaci žáků. Problematika převodu výsledku testu na známku může být velmi obtížná. Protože předkládaná habilitační práce (a test v ní vytvořený a používaný) si v žádném případě klasifikaci žáků za cíl neklade, neuvádí se na tomto místě podrobnější informace o uvedené problematice.
- **Standardizace didaktických testů:** Lapitka uvádí: O standardizaci didaktických testů se stále živě diskutuje a střetávají se dosti protichůdné názory. Někteří autoři i uživatelé didaktických testů se o standardizaci téměř nezmiňují, jiní jí přikládají velkou důležitost. Je nutno odlišovat dva pojmy:
 - standardizace skóre (soubor matematických operací, pomocí kterých jsou výsledky testování konkrétního respondenta nebo malé skupiny respondentů srovnávány s výsledky celé sledované populace)
 - standardizace testu (určení norem na administraci a vyhodnocení výsledků, používaných následně pro celou sledovanou populaci) – viz též kapitoly 2.4.2 a 2.4.9.

2.5.2 Standardizace skóre didaktického testu

Smyslem standardizace skóre testu [25] je vytvoření standardu (normy), který umožní zařadit respondenta podle dosaženého počtu bodů do určité posloupnosti (stupnice, škály). Dosažený počet bodů (hrubé skóre) v testu sám o sobě nevypovídá nic 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.

Vzorek pro standardizaci skóre testu (reprezentativní, standardizační vzorek) čítá zpravidla stovky respondentů). Škál standardizace skóre je celá řada. Např. Chráška [25] uvádí percentilovou škálu, C-škálu, škálu STANIN, z-škálu, Z-škálu a T-škálu.

Percentilová škála

Ke každému dosaženému počtu bodů (hrubému skóre) se přiřadí tzv. percentilové pořadí, které udává, kolik procent respondentů ve vzorku dosáhlo **horšího výkonu**. Chráška [25] uvádí vzorec pro výpočet percentilového pořadí respondenta pro daný výsledek v testu (před výpočtem se testy seřadí od nejhoršího skóre k nejlepšímu):

$$PR = 100 \frac{c_k - \frac{c_i}{2}}{n} \quad (10)$$

kde

- PR percentilové pořadí respondenta pro daný výsledek v testu,
 c_k kumulativní četnost u daného výsledku (četnost testů s daným skóre a horším),
 c_i četnost daného výsledku,
 n celkový počet respondentů.

Dnes je možno výpočet percentilů provést pomocí počítače. Je-li splněn předpoklad normality rozdělení výsledků a je-li známa střední hodnota a směrodatná odchylka tohoto rozdělení, lze experimentální data aproximovat hladkou křivkou např. za využití programu Microsoft®Excel 2002*, pomocí funkce NORMDIST(x;střed_hodn;sm_odch;1), kde x je skóre daného respondenta, střed_hodn je střední hodnota skóre a sm_odch je směrodatná odchylka daného rozdělení.

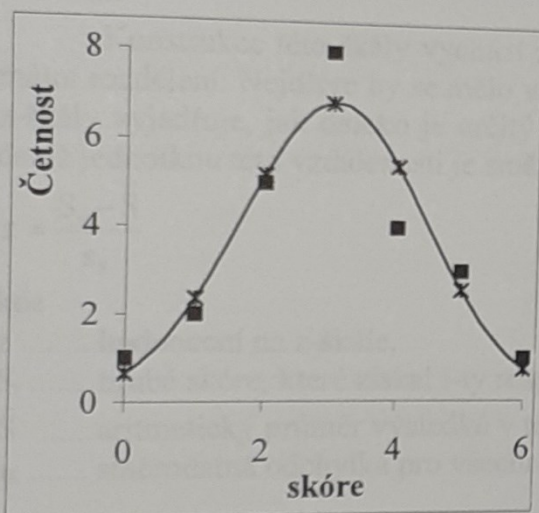
* Všude v habilitační práci je uvažována verze Microsoft®Excel 2002, avšak pro stručnost je v následujícím textu program označován pouze *Excel*.

Procentuální zastoupení respondentů ve vzorku, kteří dosáhli **stejného nebo horšího skóre** než je zvolená hodnota, se pak vypočte ze vztahu $PR' = \text{NORMDIST}(x; \text{střed_hodn}; \text{sm_odch}; 1) 100$

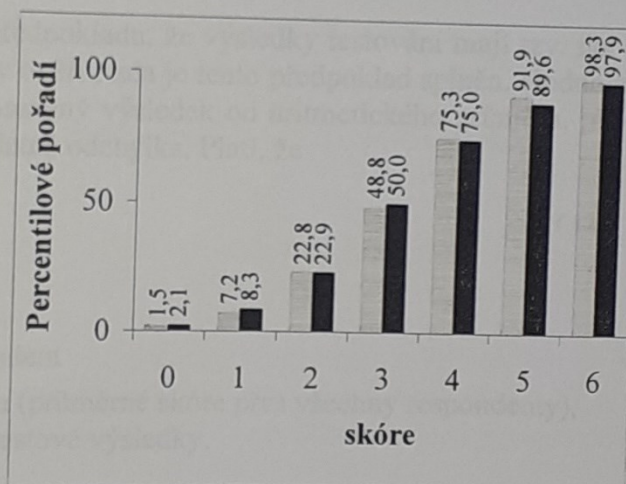
(11)

Autorka si je vědoma skutečnosti, že takto zavedená veličina PR' má poněkud odlišný význam než percentilové pořadí PR definované např. v Chráskovi [25]. Domnívá se však, že veličina PR' je pro standardizaci vhodnější než veličina PR . Důvody jsou podle ní následující:

- 1) V době širokého využití počítačů je výpočet veličiny PR' podstatně rychlejší než výpočet veličiny PR .
- 2) Hodnota PR je počítána z konkrétních experimentálních dat, nikoli z aproximace normálním rozdělením. Znamená to, že pro nízké počty respondentů se mohou výrazněji projevit náhodné odchylky od normálního rozdělení (určené vlastnostmi konkrétního výběru, nemusejí však být vlastností celého základního souboru). Naopak, hodnota PR' méně přenáší náhodné vlastnosti výběru na odhad pro základní soubor.
- 3) Pokud se vlastnosti výběru příliš neliší od normálního rozdělení, jsou hodnoty veličin PR a PR' dosti podobné, jak ukazují následující obrázky:



Obr. 1: Rozložení četností skóre. Experimentální data (čtverce) a „ideální“ normální rozdělení (hvězdičky spojené Gaussovou křivkou).



Obr. 2: Percentilové pořadí. Škála PR (černá), škála PR' (šedá). Experimentální data jsou stejná jako v obr. 1.

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 dalších nejhorších 2,8 % žáků atd. Procento respondentů pro každý bod škály je uvedeno v tab. 5.

Škála STANIN

Devítibodová škála STANIN vznikne spojením prvních dvou a posledních dvou stupňů C-škály (tab.5).

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

Tab. 5: Konstrukce C-škály a škály STANIN.

Je-li splněn předpoklad normality rozdělení výsledků a je-li známa střední hodnota a směrodatná odchylka tohoto rozdělení, lze percentilové pořadí vypočítat např. za využití programu Excel, pomocí funkce NORMINV($p_k; \bar{X}; \sigma$), kde $p_k = P_k/100$, P_k jsou kumulativní procenta (poslední sloupec v tab. 5), \bar{X} je střední hodnota a σ je směrodatná odchylka daného rozdělení [42].

z-škála

Konstrukce této škály vychází z předpokladu, že výsledky testování mají tzv. normální rozdělení. Nejdříve by se mělo vždy ověřit, 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

$$z = \frac{S_i - \bar{S}}{s_s} \quad (12)$$

kde

z hodnocení na z-škále,

S_i hrubé skóre, které získal i -tý respondent

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

s_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 Z- a T-škály.

Z-škála

Tato škála vychází ze z-škály (12) a je definována vztahem

$$Z = 100 + 10z. \quad (13)$$

T-škála

Často se k vyjadřování výsledků ve standardizovaných didaktických testech používá tzv. 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 (12) a je definována vztahem

$$T = 50 + 10z. \quad (14)$$

2.5.3 Vlastnosti testových položek

2.5.3.1 Obtížnost položky

Jak uvádí Lapitka [31], obtížnost testových položek může být interpretována různými způsoby. V naší literatuře převažuje směr, podle kterého se obtížnost úlohy posuzuje podle počtu úspěšných respondentů.

Pro položky se počítá tzv. **index obtížnosti**, nejčastěji v procentech (např. [26]):

$$P_j = \frac{n_{s,j}}{n} \cdot 100 \quad (15)$$

kde

P_j index obtížnosti j -té položky vyjádřený v procentech

$n_{s,j}$ počet respondentů, kteří j -tou položku zodpověděli správně

n celkový počet respondentů

Kromě toho existuje ještě veličina zvaná **hodnota obtížnosti** Q_j , přičemž platí

$$P_j + Q_j = 100 (\%) \quad (16)$$

Za velmi obtížné lze podle Chrásky [25] pokládat takové položky, pro které je $P_j < 20$. Naopak velmi snadné jsou ty položky, pro které $P_j > 80$. Pro položky rozlišujících testů pokládá za nejvhodnější hodnotu $P_j \doteq 50$. Položku extrémně snadnou ($P_j \doteq 100$) z psychologických (uklidňujících) důvodů doporučuje v úvodní části testu. Naopak extrémně těžké položky ($P_j \doteq 0$) v testu používat nedoporučuje.

Existují i jiné způsoby hodnocení obtížnosti položek (viz např. práce Lapitkovy).

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: skupina *lepších* žáků (s nejvyšším počtem dosažených bodů) a skupina *horších* žáků (s nejnižším počtem dosažených bodů). Tyto skupiny musí obsahovat stejný počet respondentů, např. každá obsahuje[†] 50 %, 33 %, nebo 25 % z celkového počtu respondentů (zpravidla se volí 1/3 neboli 33 %).

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

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

Nejjednodušším ukazatelem citlivosti testové úlohy je **koeficient ULI** (upper-lower-index), existují však i jiné ukazatele citlivosti položky. Koeficient ULI je definován vztahem uvedeným např. v [31]:

$$I_d = \frac{n_L - n_H}{n_{ULI}} \quad (17)$$

kde

I_d citlivost položky

n_L počet správných řešení dané položky mezi respondenty z *lepší* skupiny

n_H počet správných řešení dané položky mezi respondenty z *horší* skupiny

n_{ULI} počet respondentů v jedné skupině

U koeficientu ULI se požaduje, 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 *lepších* a *horších* respondentů vytvořené z poloviny celkového počtu respondentů.

2.5.4 Analýza nenormovaných odpovědí

Chráska [25] s odvoláním na Byčkovského 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.

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

Rozborem nesprávných odpovědí u otevřených úloh se zabývá i Reguli [20]. Opakující se nesprávné odpovědi (kromě chybějících odpovědí) klasifikuje na odpovědi typu:

- *špatně* (respondent se pokouší odpovědět, ale odpověď je špatná)
- *mimo* (odpověď neodpovídá na položenou otázku)
- *opis* (respondent pouze jinými slovy opisuje zadání)

2.6 Ověřování a optimalizace didaktického testu

I když plánování a konstrukci didaktického testu je věnována značná pozornost, není bez jeho experimentálního použití nikdy jisté, jaké vlastnosti skutečně má. Díky tomuto ověřování na vzorku respondentů lze případné nevhodné vlastnosti testu odstranit, zmírnit nebo korigovat. U standardizovaných didaktických testů se při ověřování většinou pracuje se vzorky 300-500 respondentů (Chrásková [25]).

2.6.1 Identifikace nevhodných testových položek

Z dříve řečeného plynou typické vlastnosti nevhodné testové položky:

- 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 položek

Nevhodné položky je lépe z testu vyřadit a nahradit je vhodnějšími. Z těchto důvodů se v první fázi přípravy testu připravuje 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 [33] navrhuje i další možnosti tvorby variant testových úloh, avšak podle názoru autorky habilitační práce tyto další možnosti velmi snadno vedou ke vzniku neekvivalentních testů. Proto zde tyto možnosti nejsou uvedeny.

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

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 žákova schopnost samostatné práce s informacemi.

Má-li být úroveň dosažení tohoto cíle kontrolována pomocí didaktických testů, je nutné, aby položky vyžadovaly od žáků logické myšlení, nejen reprodukci poznatků. Tento problém byl různými autory řešen různě, viz další text.

2.7.1 Didaktickologické testy

M. Cipro [43] uvažuje o tzv. didaktickologických testech. Navrhuje využít v didaktických testech podobných principů jako v testech inteligence, např. tzv. **logické úměry**. Jako příklady logické úměry Cipro uvádí např.:

Zadání: $\text{železná ruda} : \text{železo} = \text{bauxit} : ?$

Správná odpověď: *hliník*

Dále Cipro navrhuje např. testy ukládající sestavení nahodile uvedených pojmů do určité logické řady. Mohou to být např. fáze nějakého procesu:

$\text{železná ruda} \rightarrow \text{surové železo} \rightarrow \text{ocel}$

a jiné, např. řazení prvků ve skupině:

$\text{lithium} \rightarrow \text{sodík} \rightarrow \text{draslík} \rightarrow \text{rubidium} \rightarrow \text{cesium} \rightarrow \text{francium}$.

Jiný vhodný typ jsou podle Cipra testy zaměřené na **doplňování klíčových slov**, předpokládající nejen znalost faktů, ale i schopnost postihovat souvislosti mezi fakty. Např. doplněním spojky ve větě: *Voda se čistí chlorem, neobsahovala žádné původce nakažlivých nemocí* má žák prokázat, že chápe finální vztah mezi oběma větami.

Dalším typem jsou testy, prověřující schopnost **logického třídění**. Např. žák má rozdělit do dvou skupin následující pojmy: *Praha, Vltava, Ohře, Plzeň, Brno, Dyje, Ostrava, Labe, Svatka, Liberec*. Cipro k tomu vysvětluje: *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.*

V chemii by mohlo jít např. o třídění prvků do dvou skupin: *rubidium, lithium, baryum, draslík, hořčík, beryllium, sodík, vápník, cesium, stroncium, radium, francium*. Žák musí napřed určit, že jde o prvky 1. a 2. skupiny periodického systému a pak je roztrdit.

2.7.2 Test představivosti

Z. Půlpán si u žáků a studentů vysoce cení představivost. Podle něj intuitivní neformální pochopení souvislostí je základem porozumění a představivost v něm hraje základní roli. Je tedy 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

Učitel takto dostane tři množiny: $M_1 = \{J_3, J_5, J_6, J_{10}\}$, $M_2 = \{J_5, J_{10}\}$, $M_3 = \{J_2, J_5\}$. Jejich průnikem zjistí neosvojené učební jednotky: $N = M_1 \cap M_2 \cap M_3 = \{J_5\}$. Neosvojena je tedy učební jednotka označená J_5 .

Podle názoru autorky je nevýhodou této metody skutečnost, že takto konstruovaný test nerozliší, zda respondent učební jednotku J_5 skutečně nezná, nebo zda ji zná, ale neumí ji použít. Uvedený způsob vyhodnocení také selže, pokud respondent uhodne odpověď na některou z položek P_1, P_4, P_7 .

2.7.4 Rozklad na řadu dílčích operací

V Řešátkově práci [23] se s odkazem na Landu (není však uvedena jeho konkrétní práce) zmiňuje jiný způsob – rozklad na řadu dílčích operací. Jeho princip ilustruje konkrétním příkladem:

... 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:

- Neznají vztah mezi proudem a napětím při konstantním odporu vodiče.
- Dovedou tento vztah sice verbálně vyjádřit jako přímou úměrnost, ale pojem přímé úměrnosti jim je nejasný.
- Neznají značky proudu, napětí a odporu.
- Neovládají příslušný tvar Ohmova zákona, vyjádřený vzorcem $I = \frac{U}{R}$

Prvé dvě příčiny se při vlastním výpočtu vůbec nemusí projevit. Někdy se ovšem ukáží přímo, 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:[‡]

- Co platí o závislosti mezi proudem a napětím při konstantním odporu vodiče?
- Projeví se růst napětí na konstantním odporu změnou proudu v obvodu s odporem?
- Jakou značku má proud?
- Jakou značku má napětí?
- Jakou značku má odpor? ...

Z testu takto konstruovaného je možno určit, které části učiva žáci zvládli a v čem je případná příčina neúspěchu při řešení složitější otázky. Je však třeba pečlivě zvážit, podle jakého klíče test skórovat a případně přiřadit klasifikaci, protože položky testu jsou na sobě závislé. Student také může zvolit jiný, učitelem nepředpokládaný algoritmus řešení, jehož případnou správnost musí učitel uznat.

Autorka habilitační práce se Landovou metodou inspirovala při návrhu své testové metody. Uvědomuje si ovšem, že není možno klást studentům všechny kontrolní otázky na všechny vědomosti, které ke správnému řešení zadané úlohy potřebují (více viz komentář k parametru P, kap. 4.2.4). V návrhu testu se snažila respektovat i názory a rady Turkovy (kap. 2.7.6). Jako příklad zde autorka uvádí rozbor testové položky č. 30 (str. 138):

[‡] Pro zkrácení autorka habilitační práce necituje nabídnuté odpovědi.

Zadání:

Koncentrovaná kyselina dusičná je označována jako „dýmavá“. Důvodem je, že při otevření nádoby s touto kyselinou unikají páry kyseliny a oxidy dusíku vznikající jejím rozkladem. Tyto látky spolu se vzdušnou vlhkostí vytvářejí dojem „dýmu“. Označení směsi „dým“ není správné. Proč?

Autorka odhaduje, že k řešení této úlohy potřebuje respondent následující znalosti:

1. vědět, co v terminologii koloidní chemie znamená termín *dým*
2. znát skupenství oxidů dusíku za obvyklých podmínek panujících v laboratoři
3. znát skupenství, které lze označit termínem *pára*
4. znát skupenství vzdušné vlhkosti za běžných laboratorních podmínek
5. znát význam termínu *směs*

Autorka předpokládá, že na otázky č. 2-5 umí správně odpovědět téměř všichni respondenti (otázka č. 2 sice není úplně jednoduchá, avšak odpověď na ni se dá odhadnout z popisu chování oxidů dusíku, které je součástí zadání úlohy). V testu se proto kontroluje jen znalost odpovědi na otázku č. 1 (co je to dým) – kontrolní otázka č. 3 II, str. 97.

2.7.5 Metoda sdělení základních informací

Jáchim [45] v testu žákům sděluje potřebné vztahy (výpočetní vzorce) a v testu žádá jejich vysvětlení. Tímto postupem eliminuje složku pamětní (reprodukční) a soustředí se pouze na vyšší úroveň (porozumění, aplikace,...).

2.7.6 Cílené využití různých forem položek

Turek [24] uvádí vhodnost jednotlivých typů úloh k měření jednotlivých úrovní učení podle taxonomie cílů (jak Bloomovy, tak Niemiřkovy) – tab. 3 a tab. 4. Na Niemiřkovu taxonomii a doporučené formy testovacích položek v tab. 4 se autorka habilitační práce odvolává při zdůvodnění, proč do testu mezi aplikační úlohy A zařadila i položky subjektivně skórovatelné.

3 Použité statistické metody

Autorka rozsah a obsah této kapitoly pečlivě zvažovala a rozhodla se prezentovat pouze ty metody, které sama použila v habilitační práci. Důvodem je zejména malá autorčina znalost problematiky statistiky a obava, že autorčiny případné chyby nebo nepřesnosti by mohly být zdrojem chyb práce dalších pracovníků. Zbývající část statistických výpočtů provedl RNDr. J. Polách pomocí programu Statistica 7.0.

3.1 Výběr vzorku

Kapitola je zpracována podle prací Hniličkové [40], Hájka [46] a Hendla [47].

Je-li zapotřebí určit vlastnosti určitého základního souboru (složeného z velmi rozsáhlého množství prvků), není z praktického hlediska možné zjistit je přímým měřením vlastností všech jednotlivých prvků. V takovém případě je nutno provést **výběr** tzv. **reprezentativního vzorku**.

Hendl popisuje čtyři základní způsoby výběru:

- výběr **na základě dobrovolnosti**,
- výběr **na základě dostupnosti**,
- **kvótní výběr** (předem je určen počet respondentů v několika různých kategoriích; volba respondentů však může být ovlivněna zvyky a vlastnostmi tazatele),
- **náhodný výběr** – je pro statistické šetření nejlepší. Představuje však ideál, který je v praxi často jen obtížně uskutečnitelný.

Ideální by bylo provést tzv. **prostý náhodný výběr**, při němž má každá jednotka základního souboru (např. respondent) stejnou naději, že bude do vzorku pojata. Prostý náhodný výběr někdy neuspokojuje. V případě nehomogenního základního souboru může být užitečné nejprve rozdělit základní soubor na dílčí soubory (např. podle typů škol) a teprve pak v nich provádět náhodný výběr. Lze 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, je zapotřebí výběr z nich proti proporcionálnímu způsobu zvýhodnit. Existují i složitější metody výběru, jejichž podstatu i statistické zpracování vysvětluje např. Hájek.

V habilitační práci byl použit výběr na základě dostupnosti. Autorka byla omezena především ochotou škol spolupracovat na testování.

3.2 Statistická hypotéza a hladina významnosti

Podle Hájka [46] spočívá první krok při statistickém testování ve formulování nulové hypotézy H_0 , resp. alternativní hypotézy H_1 .

Nulová hypotéza H_0 je tvrzení, které obvykle deklaruje *žádný rozdíl* (tj. nalezený rozdíl lze přičíst přirozené variabilitě dat). Dokud není dokázán opak, předpokládá se, že platí nulová hypotéza. Při statistickém testování se volí tzv. **hladina významnosti** α , což je pravděpodobnost, že se zamítne nulová hypotéza, ačkoliv ona platí. Hladina α se volí velmi malá, obvykle 0,05 nebo 0,01 (v habilitační práci bylo voleno $\alpha = 5\%$, neboli $\alpha = 0,05$).

Alternativní hypotéza H_1 znamená situaci, kdy nulová hypotéza H_0 neplatí. Obvykle se vyjadřuje jako *existence odlišnosti mezi skupinami* nebo *existence závislosti mezi proměnnými*.

3.3 Statistická analýza jednorozměrných dat

Po prostudování dostupné statistické literatury dospěla autorka práce k názoru, že nejpřehlednější výklad statistické analýzy experimentálních dat podávají autoři Hendl, Meloun a Militký v knihách [47] a [37], avšak především, že názory různých autorů na správné statistické zpracování experimentálních dat se poměrně různí. Není-li uvedeno jinak, je text této kapitoly zpracován právě podle [47] a [37].

Autorka se zde zaměřuje především na postup testování statistické hypotézy o dvou srovnávaných výběrech (v habilitační práci jsou porovnávány výsledky testování u dvou skupin respondentů – studenti gymnázia a vysokoškolští studenti učitelství chemie).

3.3.1 Stručný přehled postupu při zpracování dat

Při testování statistické hypotézy o dvou výběrech je postupně nutno:

- **Ověřit předpoklady o datech**, tj. ověřit, zda oba srovnávané výběry jsou:
 - **nezávislé.**
 - **homogenní.**
 - **mají normální rozdělení.**
- Provést **odhad parametru polohy** a **odhad parametru rozptýlení**. Způsob vyhodnocení závisí na počtu prvků ve výběru.
- Testovat **shodu parametru rozptýlení** (volba testu závisí na rozsahu výběru, typu rozdělení výběru a na přítomnosti odlehlých hodnot)
- Testovat **shodu parametru polohy** (volba testu závisí na rozsahu výběru, shodě či neshodě parametru rozptýlení, typu rozdělení výběru a na přítomnosti odlehlých hodnot)

3.3.2 Ověření předpokladů o datech

Ověření nezávislosti prvků výběru

Pro identifikaci závislosti dat na čase nebo pořadí měření doporučuje Meloun **von Neumannovo kritérium**. V případě didaktických testů však podle názoru autorky habilitační práce připadá v úvahu spíše závislost výsledku respondentovy práce na jeho umístění v učebně (možnost vzájemných konzultací respondentů nebo opisování od jiných respondentů, využití pomůcek umístěných např. na stěnách učebny,...), než závislost na čase nebo pořadovém čísle respondenta. Vyhodnocení této závislosti je však jen velmi obtížně proveditelné pomocí von Neumannova kritéria. Proto nezávislost prvků nebyla testována a autorka prostě předpokládala, že zadavatel testu naznačeným situacím zabránil.

Ověření homogenity výběru

Homogenní výběr znamená, že všechny prvky pocházejí ze stejného rozdělení s konstantním rozptylem. Problém zpracování dat s odlehlými prvky je velmi komplikovaný a starší literatura o zpracování výsledků didaktického testování se jím vůbec nezabývá.

Nejjednodušším řešením problému nehomogenity je rozdělit výběr podle logicky zvolených 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 odlehlou hodnotu ze souboru vyloučit.

Hendl uvádí, že např. aritmetický průměr je velmi citlivý vůči krajním hodnotám. Při podezření, že výběr obsahuje odlehlé hodnoty, navrhuje použití jiného výpočtu (medián, percentilový průměr, seříznutý průměr, useknutý průměr). Uvádí také alternativní metody výpočtu míry rozptýlenosti. Dále pak říká: *Existují statistické testy, kterými odlehlé*

hodnoty posuzujeme. Jejich užitečnost je sporná. Pak uvádí princip několika metod identifikace odlehlých hodnot.

Meloun se k problému odlehlých hodnot staví méně skepticky. Pro jejich identifikaci podrobně rozebírá např. **metodu vnitřních hradeb** (tuto metodu použila autorka v habilitační práci). Je založena na tvrzení, že prvky výběru ležící mimo hradby $[B_D^*; B_H^*]$ se považují za odlehlé. Vnitřní hradby B_D^*, B_H^* lze určit pomocí vztahů:

Modifikovaná dolní hradba B_D^* :

$$B_D^* = \bar{\bar{X}}_{0,25} - L(\bar{\bar{X}}_{0,75} - \bar{\bar{X}}_{0,25})$$

kde

$\bar{\bar{X}}_{0,25}$ první kvartil (lze jej určit např. v programu Excel funkcí QUARTIL(pole;1)

$\bar{\bar{X}}_{0,75}$ třetí kvartil (lze jej určit např. v programu Excel funkcí QUARTIL(pole;3)

L parametr zvolený tak, aby pravděpodobnost, že z výběru velikosti n a pocházejícího z normálního rozdělení nebude žádný prvek mimo vnitřní hradby, byla dostatečně vysoká, např. 0,95. Při volbě pravděpodobnosti 0,95 platí pro výběr o 8-100 prvcích vztah:

$$L \doteq 2,25 - \frac{3,6}{n} \quad (20)$$

Modifikovaná horní hradba B_H^* :

$$B_H^* = \bar{\bar{X}}_{0,75} + L(\bar{\bar{X}}_{0,75} - \bar{\bar{X}}_{0,25}) \quad (18)$$

(19)

Podobná je také **metoda krabicového grafu** (použita RNDr. Poláchem ke zpracování dat habilitační práce), avšak koeficient L má pro hladinu významnosti 5 % konstantní hodnotu $L = 1,5$.

Autoři učebnice [48] přehledně zpracovávají problematiku statistického zpracování výsledků naměřených v chemicko-analytické laboratoři. Pokud lze na odlehlé hodnoty měření nazírat jako na data zatížená hrubou chybou, je podle uvedené učebnice možno je identifikovat a vyloučit také pomocí Grubbsova nebo Dean-Dixonova testu. Oba tyto testy však pracují pouze s malým počtem měření (zhruba do 20).

Ověření normality rozdělení výběru

Normalitu rozdělení výběru lze u velkých souborů ověřit např. využitím **testu χ^2 dobré shody (Pearsonův test)**. U menších datových souborů je tento postup špatně použitelný z toho důvodu, že jen obtížně se data rozdělí do většího počtu tříd s dostatečnou četností. V habilitační práci je autorkou používán **test kombinace výběrové šikmosti a špičatosti dle Jarque-Berra** popsáný např. v [37]:

$$\chi_{\text{exp}}^2 = \frac{g_1^2}{D(g_1)} + \frac{[g_2 - E(g_2)]^2}{D(g_2)} \quad (21)$$

kde

g_1 výběrová šikmost
 $D(g_1)$ rozptyl výběrové šikmosti
 g_2 výběrová špičatost
 $E(g_2)$ střední hodnota výběrové špičatosti
 $D(g_2)$ rozptyl výběrové špičatosti

Jejich hodnoty se vypočtou podle vztahů:

$$g_1 = \frac{\sqrt{n} \sum_{i=1}^n (X_i - \bar{X})^3}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 \right]^{\frac{3}{2}}} \quad (22)$$

$$D(g_1) = \frac{(n-2)}{(n+1)(n+3)} \quad (23)$$

$$g_2 = \frac{n \sum_{i=1}^n (X_i - \bar{X})^4}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 \right]^2} \quad (24)$$

$$E(g_2) = 3 - \frac{6}{n+1} \quad (25)$$

$$D(g_2) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)} \quad (26)$$

Význam symbolů n , X_i , \bar{X} viz (27).

Ověření normality rozdělení výběru pomocí testu kombinace výběrové šikmosti a špičatosti dle Jarque-Berra:

Pokud hodnota χ_{exp}^2 vypočtená podle (21) překročí kritickou hodnotu $\chi_{1-\alpha}^2(2)$, pak data se spolehlivostí $100(1-\alpha)\%$ podporují zamítnutí nulové hypotézy H_0 . Nulová hypotéza je v dané situaci předpoklad, že rozdělení výběru je normální.

Kritické hodnoty $\chi_{1-\alpha}^2(2)$ lze vyhledat v tabulkách (např. [47]); program Excel vrací tyto hodnoty pomocí funkce CHIINV(prst;volnost). Hodnota hladiny významnosti α (v příkaze CHIINV označena *prst*) musí být zadána jako desetinné číslo (nikoli v procentech) udávající tolerovanou nejistotu. Parametrem *volnost* v příkaze CHIINV se rozumí počet stupňů volnosti. Pro uvažovaný test má hodnotu 2.

Např. $\chi_{1-0,05}^2(2) = 5,9918$ podle [47]

CHIINV(0,05;2) = 5,99148

Šikmost a špičatost by měly vracet i funkce SKEW a KURT z programu Excel. Tyto funkce však dávají jiné číselné hodnoty, než vychází podle odpovídajících vzorců uvedených v [37] i [49]. Proto funkce SKEW a KURT autorka nepoužívala a vycházela přímo ze vztahů (21)-(26).

Při nezávislé kontrole výsledků použil RNDr. Polách pro ověření normality rozdělení dat test **Shapiro-Wilks**, který standardním vybavením programu Statistica 7.0.

Odhad parametru polohy a parametru rozptýlení

Odhad těchto parametrů závisí na počtu prvků v souboru. Podle toho se odlišují:

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

Při didaktickém testování připadají prakticky v úvahu výběry malé, střední a velké. U malých výběrů jsou závěry vždy zatíženy značnou mírou nejistoty. Proto se malých výběrů používá jen tam, kde není možné získat větší počet dat.

V habilitační práci byl použit výběr **velký** (ve skupině gymnazistů bylo 85 respondentů, ve skupině vysokoškoláků bylo 98 respondentů) a výběr **malý** (počet skórovaných testových položek byl 18).

Meloun a Militký [37] doporučují tzv. **Hornův postup** statistického zpracování výsledků pro **malé výběry** do 20 měření. Rozhovorem s několika pedagogickými pracovníky i odbornými matematiky si však autorka ověřila, že Hornův postup nepatří ke všeobecně známým a přijímaným metodám statistického vyhodnocení dat. Aby autorka předešla problémům spojeným s nesprávnou volbou metody, rozhodla se statistické zpracování malých souborů svěřit odborníkovi[§]. Data byla zpracována statistickým balíčkem Statistica 7.0.

Vztahy pro velký výběr lze s přijatelnou přesností ztotožnit vztahy pro základní soubor.

Odhadem *parametru polohy* je **aritmetický průměr** (pro výběr značený \bar{X} , pro základní soubor μ):

$$\mu = \frac{\sum_{i=1}^n X_i}{n} \quad (27)$$

kde

μaritmetický průměr (pro výběr značený \bar{X})

X_ii-tá naměřená hodnota

npočet měření

ičíslo měření

Mírou rozptýlení je **rozptyl**:

směrodat. odch. = $\sqrt{\text{rozptyl}}$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \quad (28)$$

kde

σ^2 rozptyl (variance) velkého výběru, ostatní symboly viz (27).

Směrodatná odchylka se vypočte ze vztahu

$$\sigma = +\sqrt{\sigma^2} \quad (29)$$

kde

σ směrodatná odchylka základního (velkého) souboru

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

Programem Excel lze vypočítat rozptyl základního souboru (resp. velkého výběru) pomocí funkce VAR a směrodatnou odchylku základního souboru (velkého výběru) pomocí funkce SMODCH.

[§] RNDr. Polách Jiří (člen České statistické společnosti).

Směrodatná chyba (směrodatná odchylka výběrového rozdělení průměrů, směrodatná chyba průměru, střední chyba průměru): Matematicky lze dokázat, že směrodatná odchylka rozdělení výběrových průměrů je rovna směrodatné odchylce původního rozdělení vydělené druhou odmocninou rozsahu výběru [50].

Někteří autoři (např. Hendl) počítají rozptyl s^2 (a s ním související směrodatnou odchylku) i pro velký výběr ze vztahu (hodnota s bývá nazývána *směrodatná odchylka výběru*; hodnota s^2 je *rozptyl výběru*):

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

význam symbolů \bar{X} , X_i , n , i viz (27).

rozptyl

$$s = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$$

směrodatná odch.
(30)

Pomocí programu Excel lze vypočítat rozptyl výběru pomocí funkce VAR.VÝBĚR a směrodatnou odchylku výběru pomocí funkce SMODCH.VÝBĚR.

Pro velké počty respondentů však poskytují vztahy (28) a (30) velmi podobné hodnoty.

Např. při počtu respondentů 31 je poměr $\frac{s^2}{\sigma^2} = \frac{1}{\frac{30}{31}} = 1,03$.

Medián

dělí řadu podle velikosti seřazených výsledků na dvě stejně početné poloviny.

Pro sudý počet prvků:

$$\tilde{X} = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$$

Pro lichý počet prvků:

$$\tilde{X} = X_{\frac{n+1}{2}} \quad (31)$$

kde

\tilde{X} medián pro výběr

X_i hodnota konkrétního prvku

n počet prvků

i index označující pořadové číslo prvku, jsou-li seřazeny podle velikosti

Modus neboli modální hodnota

je hodnota, která se v datech vyskytuje nejčastěji. Pokud data obsahují několik modálních hodnot, udávají se všechny. Více informací lze nalézt např. v Hendlovi [47].

Variační rozpětí

je rozdíl mezi nejvyšší a nejnižší naměřenou hodnotou:

$$R = X_{\max} - X_{\min}$$

kde

(33)

R variační rozpětí

X_{\max} ... nejvyšší naměřená hodnota

X_{\min} ... nejnižší naměřená hodnota

Variační koeficient

udává (v procentech), nakolik kolísají naměřené hodnoty kolem aritmetického průměru:

$$V = \frac{\sigma}{\mu} \cdot 100, \text{ resp. } V = \frac{s}{\bar{X}} \cdot 100$$

(34)

kde

V variační koeficient

σ , resp. s směrodatná odchylka velkého, resp. středního výběru

μ , resp. \bar{X} aritmetický průměr naměřených hodnot pro velký, resp. střední výběr

Testy shody parametru rozptýlení a parametru polohy

Potřebné výpočty provedl RNDr. J. Polách, pro parametr rozptýlení metodami F-test a test Leveneův, pro parametr polohy využil t-test (data s normálním rozdělením) a test Mann-Whitneyův (data s rozdělením výrazně odlišným od normálního).

3.4 Vícerozměrná data

V práci je autorkou využít **Pearsonův párový korelační koeficient**, který vyjadřuje míru vazby mezi složkami x_1, x_2 dvojrozměrného vektoru X . Jeho výpočet se v programu Excel volá příkazem PEARSON(Pole1;Pole2).

Existují i jiné korelační koeficienty, např. **Spearmanův pořadový korelační koeficient**, hodnotící vztah mezi pořadím prvků ve dvou souborech (např. vztah mezi umístěním stejných závodníků v běhu na 5 km a na 10 km).

6 Použitá a citovaná literatura

- [1] *Národní program rozvoje vzdělávání v České republice: Bílá kniha* [online]. Praha : Tauris, 2001. ISBN 80-211-0372-8. [cit. 2006-08-30]. Dostupné z World Wide Web: <<http://www.msmt.cz/files/pdf/bilakniha.pdf>>.
- [2] KING, A., SCHNEIDER, B. *První globální revoluce*. 1. vyd. Bratislava : Bradlo, 1991. ISBN 80-7127-048-2.
- [3] REGULI, Ján. *Neformálne vzdelávanie v oblasti chémie*. 1. vyd. Bratislava : Slovenská technická univerzita v Bratislave, 2001. ISBN 80-227-1553-0.
- [4] *Vzdělávání v ČR : Delorsův koncept čtyř pilířů* [online]. [cit. 2006-02-05]. Dostupné z World Wide Web: <http://www.nuov.cz/index.php?page=vzd_v_cr&s=40&idclanku=186>.
- [5] VAŠUTOVÁ, Jaroslava. *Být učitelem : co by měl učitel vědět o své profesi*. Praha : Karlova univerzita, 2002. ISBN 80-7290-077-3.
- [6] HOLADA, Karel. *Eseje na téma Standard kvalifikace učitele chemie : podpora rozvoje učitelských vzdělávacích programů a jiných vzdělávacích aktivit*. Praha : Univerzita Karlova v Praze, 2001.
- [7] BANÝR, Jiří. Pokus o formulaci standardu kvalifikace učitele chemie. In *Eseje na téma Standard kvalifikace učitele chemie : podpora rozvoje učitelských vzdělávacích programů a jiných vzdělávacích aktivit*. Praha : Univerzita Karlova v Praze, 2001, s. 4-7.
- [8] BUDIŠ, Josef. Standard učitele chemie (aneb stručný neesejový přehled návrhů a zkušeností). In *Eseje na téma Standard kvalifikace učitele chemie : podpora rozvoje učitelských vzdělávacích programů a jiných vzdělávacích aktivit*. Praha : Univerzita Karlova v Praze, 2001, s. 15-18.
- [9] SVOBODOVÁ, Miroslava. Standard kvalifikace učitele chemie. In *Eseje na téma Standard kvalifikace učitele chemie : podpora rozvoje učitelských vzdělávacích programů a jiných vzdělávacích aktivit*. Praha : Univerzita Karlova v Praze, 2001, s. 42-44.
- [10] TOMEČEK, Otto. Štandard Kvalifikácie učiteľa chémie. In *Eseje na téma Standard kvalifikace učitele chemie : podpora rozvoje učitelských vzdělávacích programů a jiných vzdělávacích aktivit*. Praha : Univerzita Karlova v Praze, 2001, s. 36-41.
- [11] HOLADA, Karel. *Eseje na téma Standard kvalifikace učitele chemie : podpora rozvoje učitelských vzdělávacích programů a jiných vzdělávacích aktivit*. Praha : Univerzita Karlova v Praze, 2001, s. 23-26.
- [12] PROKŠA, Miroslav. Manuálne a intelektové zručnosti, návyky a spôsobilosti podmieniajúce organizáciu experimentálnej činnosti vo vyučovaní chémie. In *Eseje na téma Standard kvalifikace učitele chemie : podpora rozvoje učitelských vzdělávacích programů a jiných vzdělávacích aktivit*. Praha : Univerzita Karlova v Praze, 2001.
- [13] BENEŠ, Pavel. Může dobře připravený učitel chemie přispět k rozvoji společnosti? In *Eseje na téma Standard kvalifikace učitele chemie : podpora rozvoje učitelských vzdělávacích programů a jiných vzdělávacích aktivit*. Praha : Univerzita Karlova v Praze, 2001, s. 8-10.
- [14] RYCHTERA, Jiří. Požadavky „pedagogické praxe“ jako základ kvalifikačního standardu učitele chemie? In *Eseje na téma Standard kvalifikace učitele chemie :*

- podpora rozvoje učitel'ských vzdělávacích programů a jiných vzdělávacích aktivit.*
 Praha : Univerzita Karlova v Praze, 2001, s. 36-41.
- [15] JEDLOVÁ, Táňa. Kvalifikační standard učitele chemie.
 In *Eseje na téma Standard kvalifikace učitele chemie : podpora rozvoje učitel'ských vzdělávacích programů a jiných vzdělávacích aktivit.*
 Praha : Univerzita Karlova v Praze, 2001, s. 27-30.
- [16] DUŠEK, Bohuslav. Poznámky ke standardu kvalifikace učitele chemie.
 In *Eseje na téma Standard kvalifikace učitele chemie : podpora rozvoje učitel'ských vzdělávacích programů a jiných vzdělávacích aktivit.*
 Praha : Univerzita Karlova v Praze, 2001, s. 19-22.
- [17] BRDIČKA, Rudolf. *Základy fyzikální chemie*. 1. vyd. Praha : Přírodovědecké vydavatelství, 1952.
- [18] KLOUDA, Pavel. *Fyzikální chemie*. 1. vyd. Ostrava : Pavel Klouda, 1997. ISBN 80-902155-2-1.
- [19] REGULI, Ján. Fyzikálna chémia v príprave učiteľov chémie.
 In *Profil učitele chemie II*. Hradec Králové : Gaudeamus, 2002, s. 87-90.
- [20] REGULI, Ján. Schopnosť študentov aplikovať vedomosti na reálne situácie.
Chemické listy, 2004, roč. 98, č. 4, s. 201-204. ISSN 0009-2770.
- [21] PACHMANN, Eduard. *Obecná didaktika chemie. Díl IV*. 1. vyd. Praha : Státní pedagogické nakladatelství, 1981.
- [22] SMÉKAL, V., ŠVEC, V., ZAJAC, J. *Didaktické testy a jejich vyhodnocování*. Brno : Středisko pro výzkum učebních metod a prostředků, 1973.
- [23] ŘEŠÁTKO, Miloš. *Didaktické testy ve školní praxi*.
 Praha : SNTL-Nakladatelství technické literatury, 1975.
- [24] TUREK, Ivan. *Učitel a didaktické testy*. 1. vyd. Bratislava : Metodické centrum mesta Bratislavy, 1996. ISBN 80-7164-139-1.
- [25] CHRÁSKA, Miroslav. *Didaktické testy. Příručka pro učitele a studenty učitelství*.
 1. vyd. Brno : Paido, 1999. ISBN 80-85931-68-0.
- [26] KOHOUTEK, Rudolf. *Didaktické testy*.
 Brno : Akademické nakladatelství CERM, 1996. ISBN 80-7204-018-9.
- [27] PŮLPÁN, Zdeněk.
K problematice zpracování empirických šetření v humanitních vědách.
 1. vyd. Praha : Academia, nakladatelství Akademie věd České republiky, 2004. ISBN 80-200-1221-4.
- [28] VRÁNA, Stanislav. *Zkoušení a známkování*. 1. vyd. Brno : Komenium, 1948.
- [29] MICHALIČKA, Miroslav. *Metody pedagogické diagnostiky*. Praha : Pedagogický ústav J. A. Komenského ČSAV, 1968. počet stran
- [30] MUŽIČ, Vladimír. *Testy vědomostí*. 1. vyd. Praha : Státní pedagogické nakladatelství, 1971.
- [31] LAPITKA, Marián. *Tvorba a použitie didaktických testov*. 1. vyd. Bratislava : Slovenské pedagogické nakladateľstvo, 1990. ISBN 80-08-00782-6.
- [32] LAPITKA, M., LEZO, H., PÍŠŠOVÁ, M., aj.
Základy metodológie pedagogického výskumu. Študijný text pre špecializačné štúdium učiteľov výskumníkov.
 1. vyd. Bratislava : Slovenské pedagogické nakladateľstvo, 1985.
- [33] RÖTLING, Gustáv. *Metodika tvorby učiteľského didaktického testu*.
 Banská Bystrica : Metodické centrum, 1996. ISBN 80-8041-110-7.
- [34] BYDŽOVSKÁ, J., MARINKOVÁ, H., SOUČKOVÁ, J.
Tvoříme vzdělávací program vyšší odborné školy [online]. 17. října 2002

- [cit. 2006-01-22]. Dostupné z World Wide Web:
<https://www.eco.cz/kraj/heslo_file/projekt_VOS.doc>.
- [35] ŠVEC, V., FILOVÁ, H., ŠIMONÍK, O. *Praktikum didaktických dovedností*. 1. vyd. Brno : Masarykova univerzita, 1996. ISBN 80-210-1365-6.
- [36] HALIŠKA, Jaromír. *Jak testy sestavit a pracovat s nimi*. 2. vyd. Brno : Středisko služeb školám, 1999.
- [37] MELOUN, M., MILITKÝ, J.
Kompendium statistického zpracování dat : metody a řešené úlohy včetně CD. 1. vyd. Praha : Academia, 2002. ISBN 80-200-1008-4.
- [38] EBEL, R. L. – *Essential of Education Measurement*. Englewood Cliffs, N.J., Prentice Hall 1972.
- [39] PŮLPÁN, Zdeněk. *Základy sestavování a klasického vyhodnocování didaktických testů*. Hradec Králové : Kotva, 1991. ISBN 80-900254-4-7.
- [40] HNILIČKOVÁ, J., JOSÍFKO, M., TUČEK, A.
Didaktické testy a jejich statistické zpracování. 1. vyd. Praha : Státní pedagogické nakladatelství, 1972.
- [41] *Cronbach's Coefficient Alpha* [online]. SAS Institute Inc., Cary, NC, USA c2003 [cit. 2006-08-08]. Dostupné z World Wide Web:
<http://support.sas.com/91doc/getDoc/procstat.hlp/corr_sect18.htm>.
- [42] Náповěda pro Microsoft Excel. Microsoft®Excel 2002 (10.2701.2625). Korektor českého pravopisu © 1995-2000 Lingea s.r.o.
- [43] CIPRO, Miroslav. *Modernizace základního vzdělání*. 1. vyd. Praha : Státní pedagogické nakladatelství, 1968.
- [44] PŮLPÁN, Z., KUŘINA, F., KEBZA, V. *O představivosti a její roli v matematice*. 1. vyd. Praha : Academia, nakladatelství Československé akademie věd, 1992. ISBN 80-200-0444-0.
- [45] JÁCHIM, František. *Některé metody diagnostiky vědomostí a dovedností v předmětu fyzika na základní škole*. České Budějovice : Krajský pedagogický ústav, 1984.
- [46] HÁJEK, Jaroslav. *Teorie pravděpodobnostního výběru s aplikacemi na výběrová šetření*. 1. vyd. Praha : Nakladatelství Československé akademie věd, 1960.
- [47] HENDL, Jan. *Přehled statistických metod zpracování dat : analýza a metaanalýza dat*. 1. vyd. Praha : Portál, 2004. ISBN 80-7178-820-1.
- [48] KOTOUČEK, M., SKOPALOVÁ, J., ADAMOVSKÝ, P.
Příklady z analytické chemie [online]. [cit. 2006-03-31].
Dostupné z World Wide Web: <<http://ach.upol.cz/ucebnice/index.html>>.
- [49] ŠKRÁŠEK, Josef. *Úvod do počtu pravděpodobnosti a matematické statistiky*. 2. přeprac. vyd. Brno : Vojenská akademie Antonína Zápotockého, 1971.
- [50] ZVÁROVÁ, Jana. *Základy statistiky pro biomedicínské obory : 6.1 Rozdělení výběrového průměru* [online]. 5. ledna 1999 [cit. 2006-08-29].
Dostupné z World Wide Web:
<<http://new.euromise.org/czech/tajne/ucebnice/html/html/node8.html>>.
- [51] KMEŤOVÁ, Jarmila. *Učebnica a učitel' chemie*. In *Pregraduální příprava a postgraduální vzdělávání učitelů chemie : sborník přednášek z mezinárodní konference konané 20.-22. května 2003 v Rožnově pod Radhoštěm*. 1. vyd. Ostrava : Ostravská univerzita v Ostravě, 2003, s. 81-86. ISBN 80-7042-960-7.
- [52] FIKROVÁ, Lenka. *Příspěvek k možnostem integrace biologie a chemie v přípravě učitele chemie a ve výuce na základních školách : diplomová práce*. Brno : Masarykova univerzita, Pedagogická fakulta, 2003. Vedoucí diplomové práce Hana Cídllová.