

# Vybrané aspekty hodnotenia výsledkov elektronického testovania vedomostí pri štúdiu informatiky

Pavel Horovčák, Beáta Stehlíková

**Abstrakt:** Elektronický test ako výrazne formalizovaný prostriedok hodnotenia priebežnej prípravy a vedomostí študentov má svoje nezastupiteľné miesto v pedagogickom procese, predstavuje rýchlu, modernú a efektívnu formu spätnej väzby medzi študentom a učiteľom. Príspevok je venovaný podmienkam testovania, kvantifikácii kritérií hodnotenia výsledkov testovania a štatistickej analýze faktorov úspešnosti študentov v teste. Faktory úspešnosti v teste rozoberané v príspevku sú: opakovanie testu, dĺžka časového limitu testovania, znalosť otázok zo strany študentov typ (počet) možných odpovedí na otázku, vhodná formulácia otázok. Špecifikuje vybrané aspekty kvantifikácie hodnotenia (kvality) výsledkov elektronického testovania vedomostí.

**Kľúčové slová:** elektronický test, databáza, faktory úspešnosti, štatistická analýza

## Úvod

Masívny nástup informačných a komunikačných technológií (IKT) do denného stereotypu spolu s budovaním informačnej spoločnosti a znalostnej ekonomiky spôsobuje rozšírenie výučby informatiky, informačnej gramotnosti, IKT na všetkých stupňoch škôl ale aj v ostatných oblastiach nášho života. Informačná gramotnosť sa stáva nevyhnutnou súčasťou vedomostného profilu predovšetkým mladej generácie. Rozvoj informačnej spoločnosti zasahuje do sféry ekonomiky, legislatívy, politiky a etiky, predovšetkým ale do sféry pedagogiky. Je preto prioritnou úlohou škôl na všetkých úrovniach venovať tejto problematike primeranú pozornosť a pripraviť tak svojich absolventov na prácu v súčasných aj budúcich podmienkach. IKT sú v súčasnosti nielen nástrojom, ale najmä stimulom pre ekonomický a sociálny rozvoj. Významným pomocníkom pri štandardnom spôsobe vzdelávania ako aj pri samostatnom získavaní vedomostí a informácií je aplikácia nezávislého kontrolného mechanizmu – napr. elektronického testu. Na našom pracovisku bola vyvinutá (a je záujemcom zdarma dostupná) takáto aplikácia v prostredí skriptu PHP a databázového systému MySQL. Využívame ju pri výučbe predmetov súvisiacich s IKT, ale jej využitie je univerzálne pre ľubovoľnú oblasť (napr. knihovníctvo, logistika, história, štatistika, biológia, ekonomika atď.), v ktorej je možné vytvoriť súbor otázok a odpovedí uzavretého typu. Niekoľko poznatkov získaných z doterajšieho používania elektronického testu je predmetom tohto príspevku.

Existujú dva spôsoby objektivizácie hodnotenia výsledkov štúdia. Prvý spôsob spočíva v skúmaní hodnotenia potom, keď už bolo vykonané (procedúry aposteriórne), druhý spôsob zvažuje realizáciu hodnotiaceho aktu pred jeho vlastným priebehom (procedúry apriórne). Väčšina pedagógov v hodnotiacom procese kombinuje tieto dve cesty v snahe dosiahnuť čo najobjektívnejší výsledok hodnotiaceho procesu. Na našom pracovisku sme školskom roku 2003-2004 začali používať aj elektronické testovanie vedomostí. Príspevok je venovaný niekoľkým plánovaným experimentom (apriórna zložka procesu skúmania), realizovanými pri používaní didaktického testu v elektronickej forme a ich výsledkom (aposteriórna procedúra hodnotenia), ktoré môžu byť využité pri ďalšom skvalitňovaní hodnotiaceho procesu (apriórna procedúra nového hodnotenia).

## Charakteristika aplikácie elektronického testu

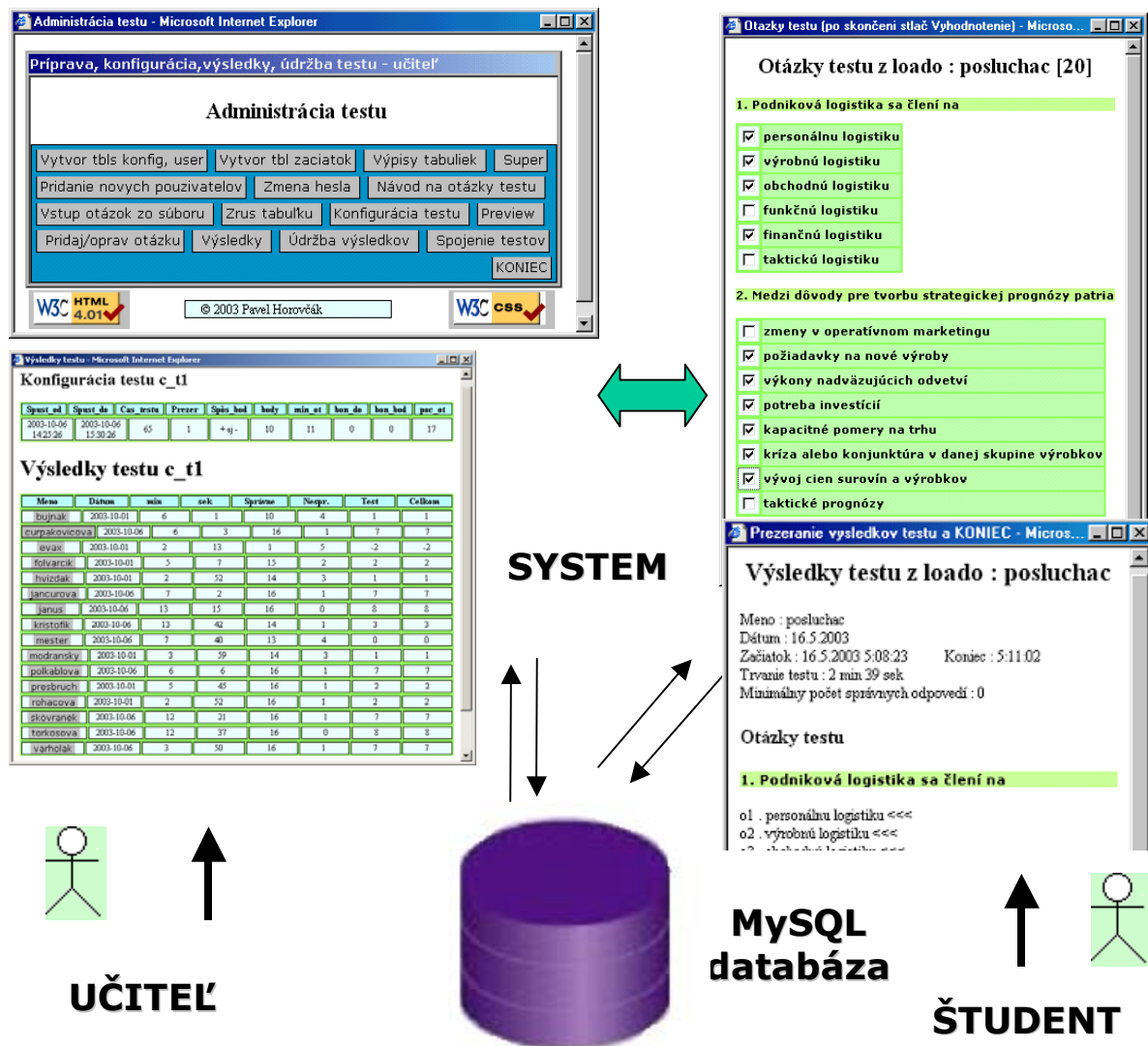
Programové vybavenie elektronického testu podporuje dve skupiny používateľov. Prvou je skupina učiteľ, druhou skupina študent. Používateľ zo skupiny učiteľ môže pracovať na vytvorení skúšobného testu (vstup otázok a odpovedí zo súboru, konfigurácia testu, prezeranie pripraveného testu, zrušenie testu, interaktívne pridanie alebo oprava otázok alebo odpovedí testu, spojenie čiastkových testov do jedného celku) a jeho vyhodnotení (výsledky jednotlivých študentov, celkové hodnotenie jednotlivých otázok za celú skupinu študentov). Okrem toho môže vykonávať celý rad rôznych systémových funkcií (inštalácia programového vybavenia, výpisy obsahov rôznych tabuliek, pridanie – zrušenie používateľov zo skupiny učiteľov, zmena prístupových hesiel a práv). K dispozícii je tiež návod na prípravu otázok testu. Funkcia konfigurácia testu umožňuje zvoliť názov testu (ak ich je viac), stanoviť pre zvolený test čas jeho trvania v rozsahu 1 až 90 minút, nastaviť časový interval platnosti testu (test je možné spúšťať iba v rámci takto zadaného časového intervalu), zvoliť zo zoznamu štýlový súbor pre riadenie zvlhľadu, nastaviť povolenie prezerania výsledkov testu, spôsob bodovania testu (bez bodovania, správne/nesprávne odpovede alebo len správne odpovede), ako aj určiť celkový počet

bodov za test. Okrem toho pri konfigurácii parametrov testu môže učiteľ nastaviť hranicu úspešnosti testu, čiže minimálny počet správnych odpovedí, od ktorého sa začína pridelovanie bodov za test. Tento počet samozrejme nesmie prekročiť počet otázok testu. Tiež je možné formou bonifikácie "odmeniť" študentov za rýchlejšie ukončenie testu. V takom prípade učiteľ môže zadať dĺžku intervalu a počet bodov bonifikácie (štandardné predvolené hodnoty sú 0, 0, to znamená bez časovej bonifikácie).

Používateľ zo skupiny študent dostane po úvodných obrazovkách (so zadaním názvu testu a priezviska študenta) formulár testu s otázkami a odpoveďami, správnosť ktorých vyznačiť je jeho úlohou. Otázky a odpovede testu sú generované výberom z databázových tabuliek v náhodnom poradí, takže každý študent má pri väčšom počte otázok (asi od 10 vyššie) jedinečné individuálne znenie testu. Po vyplnení testu alebo po skončení predpísaného časového intervalu na vykonanie testu sa uskutoční vyhodnotenie správnosti odpovedí a výpočet získanej bodovej hodnoty. Študent má potom možnosť prezrieť si výsledky svojho vypracovania testu.

Výsledky testu sú k dispozícii v dvoch formátoch – o študentovi a o otázkach . Prvý formát je tvorený položkami meno študenta, dátum, doba vypracovania testu, počet správne a nesprávne zodpovedaných aj počet nezodpovedaných otázok, počet získaných bodov za test a celkový počet bodov (pri nulovej časovej bonifikácii sú tieto hodnoty rovnaké). Druhý formát prezentuje úspešnosť riešenia jednotlivých otázok za všetkých študentov vo forme trojice čísel, a to počet správne odpovedajúcich študentov, počet nesprávne odpovedajúcich študentov a počet študentov, ktorý na otázku neodpovedali. Súčet týchto troch hodnôt dáva počet študentov, ktorí sa zúčastnili vypracovania testu. Funkčný model elektronického testu je na Obr. 1.

**Obr. 1 Funkčný model elektronického testu**



## Vybrané faktory úspešnosti

Pri objektivizácii hodnotenia výsledkov elektronického testovania je možné zaoberať sa úspešnosťou študentov a tiež úspešnosťou jednotlivých otázok testov. Úroveň úspešnosti študenta hodnoteného prostredníctvom didaktického testu je dynamická premenná, na ktorú vplyvajú mnohé faktory. Niektoré z nich sme sa pokúsili vytypovať, pomenovať a na základe konkrétnych výsledkov aj kvantifikovať a tak určiť ich opodstatnenosť (či neopodstatnenosť). Opakovanie toho istého testu rovnakou skupinou študentov s určitým časovým odstupom by malo preukázať pôsobenie *faktora učenia sa*. Vplyv na úspešnosť študenta pri teste môže mať dĺžka časového limitu, ktorý je na prácu s testom určený, teda *faktor času*. Existuje viacero možností aktivácie študentov ku zopakovaniu si učiva. Jednou z nich je priebežná kontrola ich pripravenosti na cvičenie. Vyššiu úroveň predstavuje participácia študentov na príprave priebežnej kontroly, ktorá ich núti hľadať podstatu preberaného učiva, aby „mohli klásť otázky“. Spätnou väzbou pre učiteľa sú potom nie len odpovede študentov na zadané otázky, ale aj samotné otázky, pretože už z nich je zrejmé, či skutočne smerujú k podstate. Ak študenti poznajú formuláciu zadaných otázok, aj „správne odpovede“, výsledok testu by mohol byť iný, ako v prípade, že tieto konkrétne informácie nemajú, *faktor „poznania otázok“*. Na výsledok testu môže vplyvať aj variantnosť možných odpovedí. V prípade, že medzi predvolenými odpoveďami je správna len jedna, voľba je jednoduchšia než v prípade, ak správnych odpovedí je jedna a viac, pričom ich počet nie je známy. Ako správne zodpovedaná je takáto otázka vyhodnotená len vtedy, keď sú z ponuky odpovedí vybrané všetky správne odpovede. Faktor sme označili ako *faktor typu (počtu) odpovedí*. Bolo by naivné, keby učiteľ predpokladal, že študent sa potom, ako absolvoval test, nepodelí o svoje poznatky so svojimi kolegami, ktorí ešte test neabsolvovali. Tí potom prichádzajú s cieľom absolvovať skúšanie čo najlepšie, pripravení na konkrétne otázky, ktoré sa v teste už vyskytli, to však nie je zárukou pochopenia učiva a schopnosti orientovať sa v preberanej látke. Nezanedbateľný faktor pôsobiaci na výsledok testu môže byť teda aj schopnosť študentov navzájom komunikovať. Faktor sme označili ako *faktor komunikácie*.

Podobne úroveň úspešnosti jednotlivých otázok závisí od niekoľkých faktorov. Na základe našich skúseností sme do tejto skupiny zaradili už spomínaný faktor poznania otázok a faktor typu odpovede. Optimistický predpoklad učiteľa pri používaní a hodnotení testov, že obtiažnosť všetkých zadávaných otázok je rovnaká, nie je vždy splnený. Príliš vysoká úspešnosť otázky môže byť známkou dobrého pochopenia, alebo aj triviálnosti otázky, ktorá je v takom prípade v teste zbytočná. Dôvodom nízkej úspešnosti otázky môže byť nepochopenie, či podcenenie významnej časti učiva, ale aj skutočnosť, že otázka, alebo možné odpovede nie sú dostatočne zrozumiteľné. Faktor sme označili ako *faktor kvality*. Analýza kvality otázok ako faktora úspešnosti testu by bola obtiažna, ale hodnotene a zvyšovanie kvality zadávaných otázok prispieva ku zvyšovaniu objektívnosti hodnotiaceho procesu.

Plánovanie experimentov (testy a ich priebeh) bolo uskutočnené na základe predpokladov vplyvu spomenutých faktorov. Realizácia jednotlivých experimentov bola uskutočnená v rámci procesu hodnotenia priebežnej prípravy študentov a počas ich hodnotenia v priebehu skúšky. Študenti o tom, že výsledky „ich testov“ budú použité na štatistické skúmanie neboli informovaní.

## Metodika hodnotenia

Na vyhodnotenie výsledkov plánovaných experimentov boli použité matematicko-štatistické metódy, pretože umožňujú formalizáciu výsledkov skúmania a aplikovaných postupov.

Výsledky testu, ktoré má učiteľ ako jeden z používateľov k dispozícii obsahujú informácie o každom „testovanom“ študentovi a každej v teste zadanej otázke. Analýze, alebo skúmaniu sme podrobili správanie sa (výsledky) študentov aj správanie sa (výsledky) rôznych typov otázok. Z dostupných informácií boli vytvorené a skúmaniu podrobené nasledujúce náhodné premenné:

### pre skúmanie „úspešnosti“ študentov:

- ✓ počet správne zodpovedaných otázok,
- ✓ počet nesprávne zodpovedaných otázok
- ✓ počet nezodpovedaných otázok

### pre skúmanie „úspešnosti“ otázok:

- ✓ počet správnych odpovedí na otázku
- ✓ počet nesprávnych odpovedí na otázku
- ✓ počet, koľkokrát otázka nebola zodpovedaná

Štandardné štatistické testy (pre jedno alebo viac rozmerné výbery) vyžadujú dostatočný počet dát a normalitu dát. Ak výbery obsahujú dostatočný počet hodnôt a nami zvolená náhodná premenná

$x = \text{„početnosť“}$  nespĺňa podmienku normality, môže ju zabezpečiť „odmocninová“ transformácia:  $o = \sqrt{x}$  [4].

V opačnom prípade, na overenie významnosti, či nevýznamnosti uvažovaných faktorov je potrebné použiť testy neparametrické. My sme zvolili znamienkový test a poradovú analýzu rozptylu - Kruskal-Wallisov test.

## Znamienkový test

Tento neparametrický test je možné použiť na overenie hypotézy o zhode dvoch výberov, z hľadiska zmeny pri opakovanom získavaní výberov. Hodnoty výberov musia byť usporiadané v pároch (v našom prípade bol test použitý pri hodnotení úspešnosti študentov v dvoch testoch, pár tvorí premenná  $x$  jedného testu a rovnaká premenná  $x$  porovnávaného testu).

Testom overovaná nulová hypotéza je  $H_0$ : dva výbery sú zhodné, proti alternatívnej hypotéze  $H_A$ : medzi výbermi existuje štatisticky významný rozdiel.

Pre každú dvojicu údajov je určený typ zmeny, nárast alebo pokles (zlepšenie, alebo zhoršenie individuálneho výkonu), z čoho dostaneme určitý počet znamienok „+“ a „-“ (situácie, kedy zmena nenastala vynechávame). Nulovú hypotézu prijímame, ak počet znamienok „+“ ani „-“, neprekročí kritickú hodnotu pre počet dvojíc  $n$  s ktorým pracujeme (tabuľky). Prakticky tento rýchly test vyžaduje, aby existovala rovnaká hodnota pravdepodobnosti výskytu obidvoch typov znamienok  $p(\text{sgn}^+) = p(\text{sgn}^-) = 0,5$ , podľa binomického rozdelenia. V prípade overovania platnosti alternatívnej hypotézy počet znamienok redšie sa vyskytujúcich nesmie kritickú hodnotu prekročiť.

## Kruskal-Wallisov test

Kruskal-Wallisov test je neparametrická analógia analýzy rozptylu. Možnosť použitia tohto testu je viazaná na merateľnosť premennej v rankovej (poradovej) škále (táto požiadavka je pre premennú  $x = \text{„početnosť“}$  splnená).

Testom overovaná nulová hypotéza je  $H_0$ : rozdelenia všetkých základných súborov sú rovnaké, proti alternatívnej hypotéze  $H_A$ : aspoň jedno rozdelenie základného súboru je odlišné.

Všetkých  $n = n_1 + n_2 + \dots + n_k$  vybraných pozorovaní, kde  $n_j$  je rozsah jednotlivých výberov a  $k$  je počet výberov sa zoradí vzostupne v zmiešanom výbere. Každému pozorovaniu je potrebné pridať štandardizované poradie  $p_i$  tak, aby rovnaké hodnoty pozorovaní získali rovnakú hodnotu poradia vypočítanú ako priemer zo zhodných poradí tak, aby platila podmienka  $\sum_{i=1}^n p_i = 1 + 2 + 3 + \dots + n$ . Pre každý

hodnotený výber sa potom vypočíta suma jeho štandardizovaných poradí  $R_j$ . Testovacie kritérium  $H$  je založené na rozptyle štandardizovaných poradí, ktorý je daný vzťahom  $H = \frac{12}{n(n+1)} \cdot \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(n+1)$ .

Kritická hodnota má približne Pearsonovo rozdelenie  $X^2_{(\alpha, k-1)}$ , pre hladinu významnosti  $\alpha$  a  $k-1$  stupňov voľnosti.

Pre prijatie  $H_0$  je potrebné, aby testovacie kritérium bolo menšie, ako kritická hodnota. Pre prijatie hypotézy  $H_A$  je potrebné, aby testovacie kritérium bolo väčšie, ako kritická hodnota.

## Test na vyhľadanie odľahlých pozorovaní vo viacrozmernom výbere

Tento štatistický postup obvykle predchádza vlastnému spracovaniu dát a slúži na vylúčenie možných chýb a omylov z meraní. Výhodiskom je úvaha o tom, že kvôli práci alebo nákladnosti nie je možné preveriť všetky dáta, preto je vhodné zamerať sa na dáta, ktoré sa pri danom rozdelení pravdepodobnosti vyskytujú pomerne vzácnne. Medzi takýmito hodnotami sa relatívne často vyskytujú chyby, ale na druhej strane takéto hodnoty najviac ovplyvňujú výsledky výpočtov.

V dátach  $p$ -rozmerného výberu s rozsahom  $n$  sa odľahlé pozorovania vyhľadávajú nasledovným spôsobom. Pre každé pozorovanie  $x_i$  sa určí vzdialenosť od výberového centroidu  $\bar{x}$  (výberová Mahalanobisova vzdialenosť)  $D_i^2 = (x_i - \bar{x})^T S^{-1} (x_i - \bar{x})$ , kde  $S$  je kovariančná matica výberu. Funkcia

tejto štatistiky  $F_i = \frac{(n-p) \cdot n}{(n^2-1) \cdot p} \cdot D_i^2$  má Fisherove rozdelenie  $F(p, n-p)$ . V praxi to znamená, že ak hodnota funkcie pri i-tom pozorovaní  $F_i > F(p, n-p)$ , tak dané pozorovanie je považované za odľahlé.

## Test presnosti a správnosti dvoch viacrozmerných výberov

Porovnávanie viacrozmerných výberov je analógiou porovnávania jednorozmerných výberov. Prvým krokom je overenie normality dát (malé odchýlky od normality môžu byť kompenzované rozsahom výberov). Potom nasleduje overenie zhody variability výberov (test zhody kovariančných matic), teda overenie platnosti hypotézy  $H_{0S} : S_1 = S_2 = S$  o rovnakej presnosti - homoskedasticite. Ďalší krok je overenie zhody stredných hodnôt (test zhody vektorov stredných hodnôt), teda test hypotézy  $H_{0\bar{x}} : \bar{x}_1 = \bar{x}_2 = \bar{x}$  o rovnakej polohe – správnosti dát.

Na základe prijatia / zamietnutia hypotéz  $H_{0S}$  a  $H_{0\bar{x}}$  je formulovaný záver o zhode / nezhode týchto dvoch základných charakteristík výberov.

## Výsledky hodnotenia úspešnosti študentov

### Faktor učenia sa

*Podmienky:* test i\_msdos mal 22 otázok, bol zameraný na všeobecné informácie o operačnom systéme MSDOS, experimentu sa zúčastnilo 70 respondentov, ktorí s odstupom jedného týždňa opakovali ten istý test.

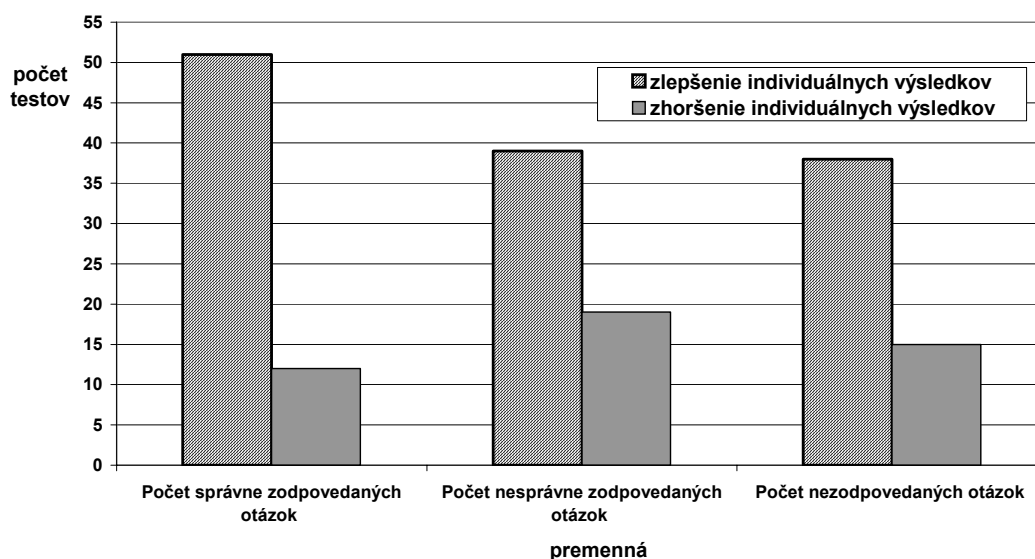
*Očakávaný výsledok:* zlepšenie individuálnych výsledkov respondentov pri opakovanom teste.

*Hodnotenie:* premenné pre skúmanie „úspešnosti“ študentov, znamienkový test. Na hladine významnosti  $\alpha = 0,05$  bola overovaná nulová hypotéza, že opakovaním testu nedošlo k zlepšeniu výsledkov každého študenta oproti alternatívnej hypotéze  $H_1/H_2$ , že opakovaním testu sa výsledky zlepšili / zhoršili. Výsledky štatistického overovania prezentuje Tab. 1.

**Tab. 1 Výsledky experimentu pre analýzu vplyvu faktora učenia sa**

H0:	V opakovanom teste sa výsledky štatisticky významne neodlišujú		
H1:	V opakovanom teste sú výsledky lepšie		
premenná	Počet správne zodpovedaných otázok	Počet nesprávne zodpovedaných otázok	Počet nezodpovedaných otázok
typ očakávanej zmeny	nárast	pokles	pokles
testovacie kritérium 1	51	39	38
kritická hodnota	25		
vyhodnotenie	testovacie kritérium > kritická hodnota, H1 je prijatá		
H2:	V opakovanom teste sú výsledky horšie		
premenná	Počet správne zodpovedaných otázok	Počet nesprávne zodpovedaných otázok	Počet nezodpovedaných otázok
typ očakávanej zmeny	pokles	nárast	nárast
testovacie kritérium 2	12	19	15
kritická hodnota	25		
vyhodnotenie	testovacie kritérium < kritická hodnota, H2 je zamietnutá		

**Obr. 2 Interpretácia výsledkov pre znamienkový test**



*Zhrnutie:* Všetky hodnoty testovacieho kritéria pre  $H_1$  sú väčšie, ako kritická hodnota, všetky hodnoty testovacieho kritéria pre  $H_2$  sú menšie, ako kritická hodnota. Nulová hypotéza  $H_0$  a alternatívna hypotéza  $H_2$  môžu byť zamietnuté v prospech alternatívnej hypotézy  $H_1$ . Znamienkovým testom bol preukázaný štatisticky významný rozdiel (zlepšenie) medzi dvojicami výsledkov študentov v prvom teste a v opakovanom teste.

Početnosti jednotlivých premenných pre prijatie hypotéz  $H_1$  a  $H_2$  sú zobrazené na Obr. 2.

### Faktor dĺžky časového limitu

*Podmienky:* test i\_spol1 mal 50 otázok, bol zameraný na súhrnné informácie z informatiky, experimentu sa zúčastnilo 33 respondentov v dvoch skupinách s počtom 12 a 21. Časový limit prvej skupiny bol 15 minút, časový limit druhej skupiny bol 20 minút.

*Očakávaný výsledok:* lepšie výsledky respondentov pri dlhšom časovom limite.

*Hodnotenie:* premenné pre skúmanie „úspešnosti“ študentov, Kruskal-Wallisov test. Na hladine významnosti  $\alpha = 0,05$  bola overovaná nulová hypotéza  $H_0$ , že rozdelenia základných súborov pre výsledky testov s rôznymi časovými limitmi sú rovnaké, oproti alternatívnej hypotéze  $H_1$ , že rozdelenia základných súborov pre výsledky testov s rôznymi časovými limitmi sú rôzne. Výsledky štatistického vyhodnotenia výsledkov experimentu zobrazuje Tab. 2.

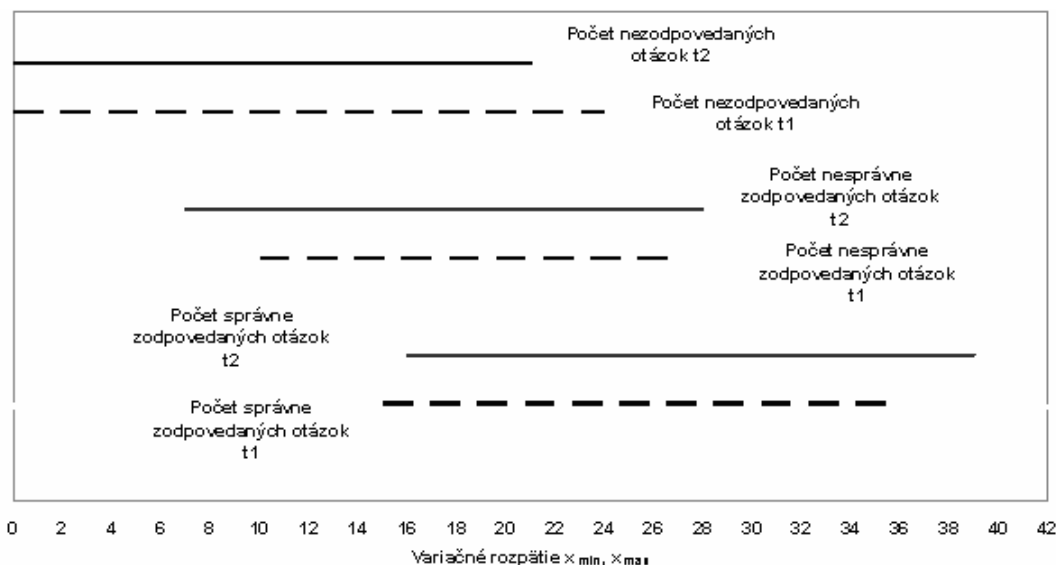
**Tab. 2 Výsledky experimentu pre analýzu vplyvu faktora dĺžky časového limitu**

H0:	Rozdelenia všetkých základných súborov sú rovnaké		
H1:	Aspoň jedno rozdelenie základného súboru je odlišné.		
premenná	Počet správne zodpovedaných otázok	Počet nesprávne zodpovedaných otázok	Počet nezodpovedaných otázok
testovacie kritérium	0,34	0,20	0,04
kritická hodnota	3,84		
vyhodnotenie	testovacie kritérium < kritická hodnota, H0 je prijatá		

*Zhrnutie:*  $H_0$  o zhode rozdelenia výberov nemôže byť zamietnutá. Výsledok testu nepreukázal štatisticky významný vplyv dĺžky časového limitu na úspešnosť študentov v teste. Treba však poznamenať, že časový limit bol v oboch prípadoch stanovený „reálne“, potreba času nebola podhodnotená.

Rozdelenie jednotlivých premenných pre rôzne časové limity je zobrazené na Obr. 3.

**Obr. 3 Hodnotenie vplyvu dĺžky časového limitu na úspešnosť testu**



Z obrázku je zrejmé, že variačné rozpätia  $V_x = x_{\text{maximálne}} - x_{\text{minimálne}}$  zobrazené ako čiara  $x_{\text{minimálne}} \rightarrow x_{\text{maximálne}}$  tej istej premennej v testoch s rôzne zadaným časovým limitom sa líšia málo.

### Faktor poznania otázok

**Podmienky:** test oop9 mal 40 otázok, bol zameraný na tvorbu databázovej aplikácie a základy jazyka SQL, experimentu sa zúčastnilo 13 respondentov. Test pozostával z 20 známych otázok a 20 neznámych otázok. Známe otázky pripravili študenti a odovzdali ich učiteľovi. Tieto otázky, ako aj správne odpovede boli dostupné tvorcom aj ich kolegom. Otázky neznáme sa týkali rovnakého učiva, ale boli pripravené učiteľom a študenti nemali k dispozícii ich znenie, ani správne odpovede.

**Očakávaný výsledok:** lepšie individuálne výsledky respondentov pri známych otázkach.

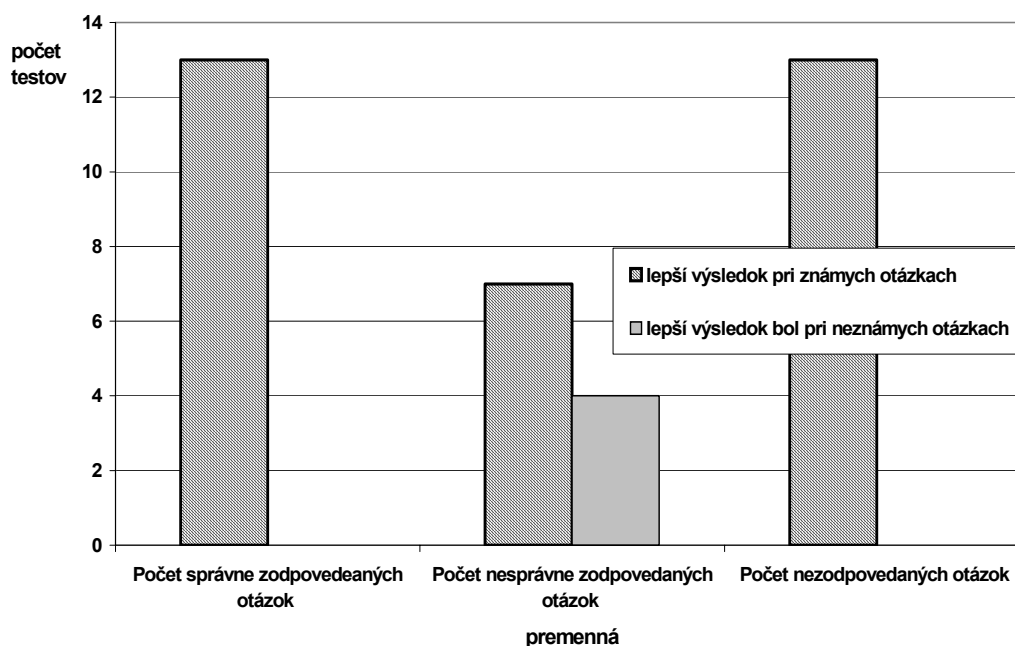
**Hodnotenie:** premenné pre skúmanie „úspešnosti“ študentov, znamienkový test, Kruskal-Wallisov test.

Pre znamienkový test bola na hladine významnosti  $\alpha = 0,05$  overovaná nulová hypotéza  $H_0$ , že poznanie otázok nespôsobil zlepšenie výsledkov každého študenta oproti alternatívnej hypotéze  $H_1 / H_2$ , že pre známe otázky sa výsledky zlepšili / zhoršili. Výsledok experimentu a jeho vyhodnotenie zobrazuje Tab. 3 a Obr. 4.

**Tab. 3 Výsledky experimentu pre analýzu vplyvu faktora poznania otázok, znamienkový test**

H0:	Výsledky pri otázkach známych a neznámych sa štatisticky významne nelíšia		
H1:	Výsledky pri známych otázkach sú lepšie		
premenná	Počet správne zodpovedaných otázok	Počet nesprávne zodpovedaných otázok	Počet nezodpovedaných otázok
typ očakávanej zmeny	nárast	pokles	pokles
testovacie kritérium	13	7	13
kritická hodnota	3		
vyhodnotenie	testovacie kritérium > kritická hodnota, H1 je prijatá		testovacie kritérium > kritická hodnota, H1 je prijatá
H2:	Výsledky pri známych otázkach sú horšie		
premenná	Počet správne zodpovedaných otázok	Počet nesprávne zodpovedaných otázok	Počet nezodpovedaných otázok
typ očakávanej zmeny	pokles	nárast	nárast
testovacie kritérium	0	4	0
kritická hodnota	3		
vyhodnotenie	testovacie kritérium < kritická hodnota, H2 je zamietnutá		testovacie kritérium < kritická hodnota, H2 je zamietnutá

**Obr. 4 Výsledok hodnotenia vplyvu faktora poznania otázok znamienkovým testom**



*Zhrnutie:* Výsledky testov boli vyhodnotené najprv znamienkovým testom. Pre premenné počet správne zodpovedaných otázok a počet nezodpovedaných otázok môže byť prijatá hypotéza  $H_1$ , avšak pri premennej počet nesprávne zodpovedaných otázok test neposkytol odpoveď, na základe ktorej by mohla byť  $H_0$  zamietnutá, kritickú hodnotu prekročil počet znamienok „+“ aj počet znamienok „-“.

Preto bola platnosť hypotézy na hladine významnosti  $\alpha = 0,05$   $H_0$ , že rozdelenia základných súborov pre známe a neznáme otázky sú rovnaké, oproti alternatívnej hypotéze  $H_1$ , že rozdelenia základných súborov pre známe a neznáme otázky sú rôzne overená Kruskal-Wallisovým testom. Štatistické testovanie je zobrazené v Tab. 4.

**Tab. 4 Výsledky experimentu pre analýzu vplyvu faktora poznania otázok Kruskal-Wallisov test**

$H_0$ :	Výsledky pri otázkach známych a neznámych sa štatisticky významne		
$H_1$ :	Výsledky pri otázkach známych a neznámych sa štatisticky významne líšia		
premenná	Počet správne zodpovedaných otázok	Počet nesprávne zodpovedaných otázok	Počet nezodpovedaných otázok
testovacie kritérium	18,12	2,29	18,78
kritická hodnota	5,99		
vyhodnotenie	testovacie kritérium > kritická hodnota, $H_1$ je prijatá	testovacie kritérium < kritická hodnota, $H_0$ je prijatá	testovacie kritérium > kritická hodnota, $H_1$ je prijatá

*Zhrnutie:* Pre premenné počet správne zodpovedaných otázok a počet nezodpovedaných otázok môže byť prijatá hypotéza  $H_1$ . Pri premennej počet nesprávne zodpovedaných otázok  $H_0$  nemôže byť zamietnutá.

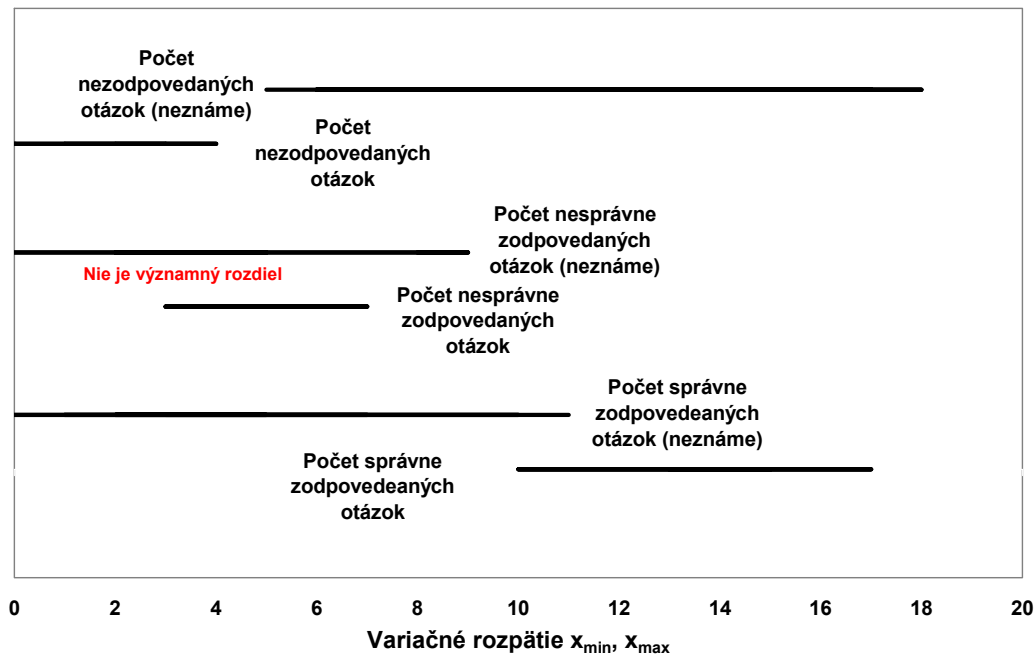
Rozdelenia premennej počet nesprávne zodpovedaných otázok pri otázkach známych a otázkach neznámych sa štatisticky významne nelíšia.

Výsledky študentov boli pri známych otázkach lepšie vzhľadom na počet správne zodpovedaných otázok (viac) a vzhľadom na počet nezodpovedaných otázok (menej).

Obr. 5 názorne interpretuje výsledok Kruskal Wallisovho testu, znázorňuje variačné rozpätie rovnakých premenných. Test so známymi otázkami a s neznámymi otázkami sa v premennej počet nesprávne zodpovedaných otázok neodlišujú.

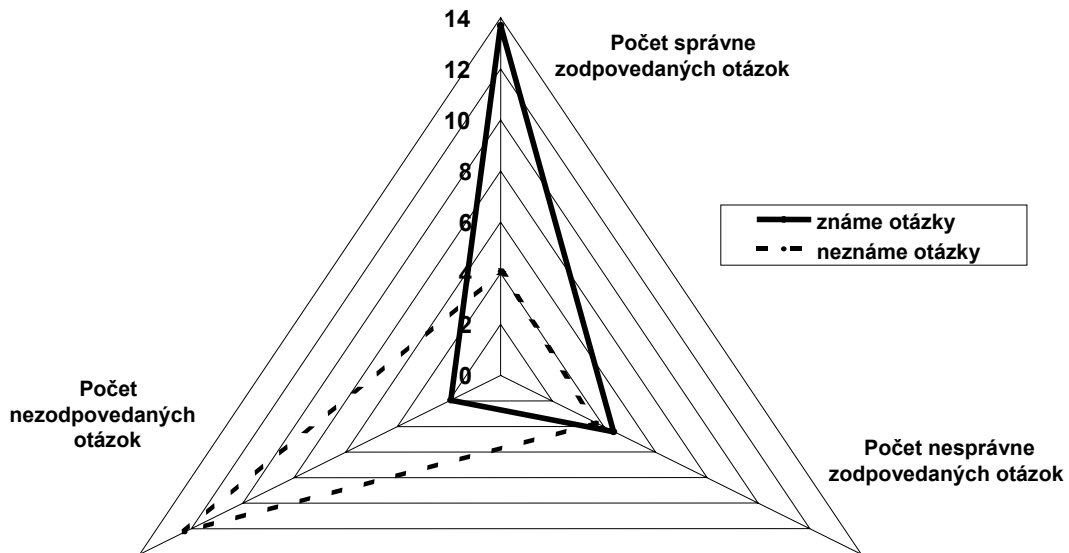


**Obr. 5 Porovnanie rozdelení premenných pre hodnotenie úspešnosti študentov pre známe a neznáme otázky**



Na Obr. 6 sú výsledky pre známe a neznáme otázky zobrazené ako trojuholníkové plochy. Vrcholy trojuholníkov sú stredné hodnoty premenných pre známe a neznáme otázky. V súlade s výsledkom štatistického testovania, rozdiel v polohe je zapríčinený rozdielom premenných Počet správne zodpovedaných otázok a počet nezodpovedaných otázok.

**Obr. 6 Rozdiel v polohe stredných hodnot pre známe a neznáme otázky**



Celkový výsledok hodnotenia vplyvu poznania otázok hodnotený rozdielom bodového zisku respondenta za známe a neznáme otázky (bodový zisk sa vypočíta ako rozdiel počet správne zodpovedaných otázok - počet nesprávne zodpovedaných otázok): Výsledok u všetkých respondentov je jednoznačne lepší pre známe otázky.

### Faktor komunikácie

*Podmienky:* test i\_spol1 mal 50 otázok, bol zameraný na oblasť informatiky, experimentu sa zúčastnilo 73 respondentov. Test bol vykonaný štyri krát.

Očakávaný výsledok: postupné zlepšovanie výsledkov.

Hodnotenie: Pre štyri trojrozmerné výbery bola overená zhoda kovariančných matíc (Boxov test) a následne analýzou rozptylu overená hypotéza o zhode vektorov stredných hodnôt (Tab. 5).

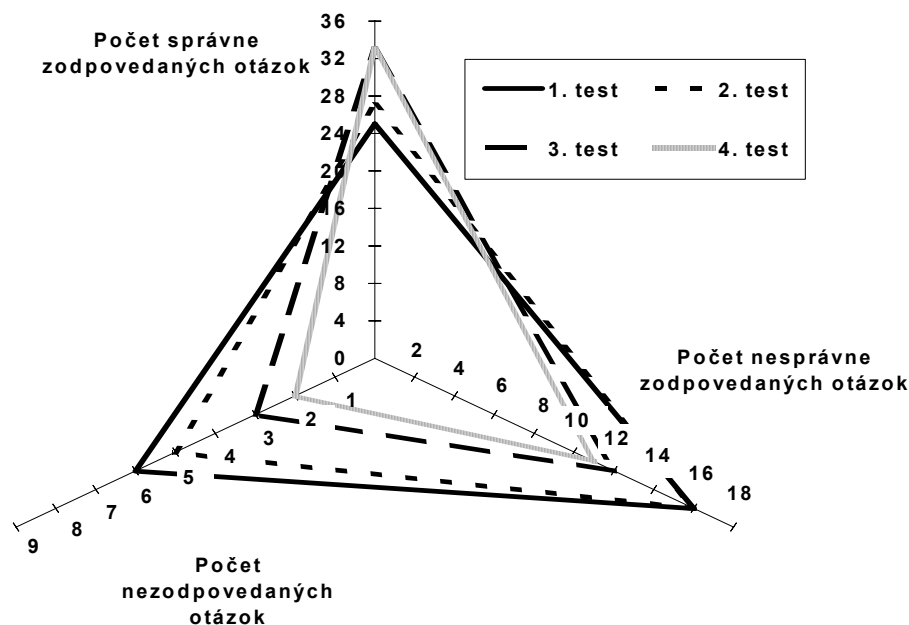
**Tab. 5 Výsledky experimentu pre analýzu vplyvu faktora komunikácie, Boxov test a analýza rozptylu**

H0:	Kovariančné matice sú zhodné, výberové rozptyly sú zhodné $S_1=S_2$
H1:	Medzi výberovými rozptylmi je štatisticky významný rozdiel
testovacie kritérium	25,09
kritická hodnota	28,87
vyhodnotenie	testovacie kritérium < kritická hodnota, H0 je prijatá
H0:	Medzi vektormi stredných hodnôt nie je štatisticky významný rozdiel
H1:	Medzi vektormi stredných hodnôt je štatisticky významný rozdiel, aspoň jeden
testovacie kritérium	2,62
kritická hodnota	2,17
vyhodnotenie	testovacie kritérium > kritická hodnota, H0 je zamietnutá, H1 je prijatá

Zhrnutie:  $H_0$  o zhode kovariančných matíc môže byť prijatá. Medzi výbermi bola potvrdená zhoda vo variabilite.

Pri analýze rozptylu  $H_0$  o zhode vektorov stredných hodnôt môže byť zamietnutá. Analýza rozptylu potvrdila rozdiel medzi výsledkami testov v strednej hodnote premenných. Bol potvrdený štatisticky významný rozdiel aspoň jednej skupiny výsledkov (zlepšenie výsledkov). Pomerne výrazné priebežné zlepšovanie čiastkových aj celkových výsledkov (prítom vždy nových) respondentov naznačuje, že treba opatrne voliť opakované používanie rovnakého testu. Za vhodnejšie pokladáme striedanie dvoch, prípadne viacerých testov.

**Obr. 7 Posun priemerných hodnôt výsledkov testov vplyvom komunikácie**



Obr. 7 a Tab. 6 poukazujú na rozdiel medzi priemernými výsledkami študentov v testoch. Zreteľný je rozdiel medzi výsledkami v prvom týždni skúšobného obdobia (1. a 2. test) a v druhom týždni (3. a 4. test), nastalo zlepšenie.

**Tab. 6 Priemerné hodnoty premenných pre hodnotenie úspešnosti študentov**

dátum	počet respondentov	priemerná hodnota premennej		
		počet správne zodpovedaných otázok	počet nesprávne zodpovedaných otázok	počet nezodpovedaných otázok
19.5.2004	12	25	16	6
25.5.2004	21	27	16	5
27.5.2004	21	33	12	3
1.6.2004	19	33	11	2

## Výsledky hodnotenia úspešnosti otázok

Premenné definované pre hodnotenie úspešnosti otázok boli transformované (odmocninová transformácia). Pre transformované premenné z premenných týkajúcich sa otázok bol overený a splnený predpoklad normality a to pre jednorozmerný výber každej premennej aj pre trojrozmerný výber.

### Faktor poznania otázok

*Podmienky:* test oop9 mal 40 otázok, bol zameraný na tvorbu databázovej aplikácie a základy jazyka SQL, experimentu sa zúčastnilo 13 respondentov. Test pozostával z 20 známych otázok a 20 neznámych otázok. Známe otázky pripravili študenti a odovzdali ich učiteľovi. Tieto otázky, ako aj správne odpovede boli dostupné tvorcom aj ich kolegom. Otázky neznáme sa týkali rovnakého učiva, ale boli pripravené učiteľom a študenti nemali k dispozícii ich znenie, ani správne odpovede.

*Očakávaný výsledok:* lepšie výsledky známych otázok.

*Hodnotenie:* Premenné definované pre hodnotenie úspešnosti otázok boli transformované (odmocninová transformácia). Pre transformované premenné z premenných týkajúcich sa otázok bol overený a splnený predpoklad normality pre trojrozmerný výber. Ďalej boli výbery známych a neznámych otázok navzájom porovnané. Boxov test, test zhody vektorov stredných hodnôt pre heteroskedasticitu. Postup a výsledky obsahuje Tab. 7.

**Tab. 7 Výsledky experimentu pre analýzu vplyvu faktora poznania otázok Boxov test, test zhody vektorov stredných hodnôt**

H0:	Kovariančné matice sú zhodné, výberové rozptyly sú zhodné $S_1=S_2$		
H1:	Medzi výberovými rozptylmi je štatisticky významný rozdiel		
testovacie kritérium	25,75		
kritická hodnota	12,59		
vyhodnotenie	testovacie kritérium > kritická hodnota, H0 je zamietnutá, H1 je prijatá		
H0:	Medzi vektormi stredných hodnôt nie je štatisticky významný rozdiel		
H1:	Medzi výberovými rozptylmi je štatisticky významný rozdiel		
testovacie kritérium	70,94		
kritická hodnota	7,81		
vyhodnotenie	testovacie kritérium > kritická hodnota, H0 je zamietnutá, H1 je prijatá		
simultánne úsudky o premenných	Počet správnych odpovedí na otázku	Počet nesprávnych odpovedí na otázku	Počet, koľkokrát otázka nebola zodpovedaná
testovacie kritérium	15,74	0,06	34,35
kritická hodnota	7,81		
vyhodnotenie	testovacie kritérium > kritická hodnota, H0 je zamietnutá v prospech H1	testovacie kritérium < kritická hodnota, H0 je prijatá	testovacie kritérium > kritická hodnota, H0 je zamietnutá v prospech H1

*Zhrnutie:* Výsledkom porovnania dvoch trojrozmerných výberov je zistenie nezahody kovariančných matíc. Výbery otázok známych a neznámych sa nezhodujú vo variabilite ( $H_0$  môže byť zamietnutá  $S_1 \neq S_2$ ). V teste zhody vektorov stredných hodnôt  $H_0$  tiež môže byť zamietnutá v prospech  $H_1$ . Rozdiel v strednej hodnote je zapríčinený premennými počet správnych odpovedí na otázku a počet,

koľkokrát otázka nebola zodpovedaná. Z celkovo dosiahnutých priemerných výsledkov vyplýva, že úspešnosť zodpovedania otázok je jednoznačne lepšia (niekoľkonásobne) pre známe otázky.

## Faktor typu odpovede

*Podmienky:* test kontrol1 mal 34 otázok, bol zameraný na oblasť informatiky, experimentu sa zúčastnilo 113 respondentov. Test pozostával z 25 otázok, s možnosťou voľby len jednej správnej odpovede (radiobox) a 9 otázok s možnosťou voľby ľubovoľného počtu správnych odpovedí (checkbox).

*Očakávaný výsledok:* lepšie výsledky jednodpovedových otázok.

*Hodnotenie:* Vzhľadom na nízky počet otázok typu checkbox (9) bol na vyhodnotenie vplyvu typu odpovedí na úspešnosť otázky použitý neparametrický Kruskal-Wallisov test. Na hladine významnosti  $\alpha = 0,05$  bola overovaná platnosť hypotézy  $H_0$  oproti  $H_1$ , o rozdelení poradí premenných pre otázky typu radiobox a otázky typu checkbox (Tab. 8).

**Tab. 8 Výsledky experimentu pre analýzu vplyvu faktora typu odpovede Kruskal Wallisov test**

H0:	Rozdelenia všetkých základných súborov sú rovnaké		
H1:	Aspoň jedno rozdelenie základného súboru je odlišné.		
premenná	Počet správnych odpovedí na otázku	Počet nesprávnych odpovedí na otázku	Počet, koľkokrát otázka nebola zodpovedaná
testovacie kritérium	11,26	11,67	4,17
kritická hodnota	3,84		
vyhodnotenie	testovacie kritérium > kritická hodnota, H1 je prijatá		

*Zhrnutie:* Všetky hodnoty testovacieho kritéria pre  $H_1$  sú väčšie, ako kritická hodnota. Nulová hypotéza  $H_0$  môže byť zamietnutá v prospech alternatívnej hypotézy  $H_1$ . Pre všetky premenné bol vplyv typu odpovedí vyhodnotený ako štatisticky významný, lepšie výsledky pri jednodpovedových otázkach boli potvrdené.

## Hodnotenie kvality otázok

Optimistický predpoklad učiteľa pri používaní a hodnotení testov, že obtiažnosť všetkých zadávaných otázok je rovnaká, nie je vždy splnený. Hodnotenie kvality otázok bolo realizované na niekoľkých z aplikovaných testov. Pri hodnotení kvality si pozornosť zaslúžia otázky, pri ktorých je usporiadaná n-tica {počet správnych odpovedí; počet nesprávnych odpovedí; počet koľkokrát otázka nebola zodpovedaná} odlišná. Príliš vysoká úspešnosť otázky môže byť známkou dobrého pochopenia, alebo aj triviálnosti otázky. Dôvodom nízkej úspešnosti otázky môže byť nepochopenie, či podcenenie významnej časti učiva, ale aj skutočnosť, že otázka, alebo možné odpovede nie sú dostatočne zrozumiteľné.

## Faktor kvality

*Podmienky:* test i\_spol1 mal 50 otázok, bol zameraný na oblasť informatiky, experimentu sa zúčastnilo 50 respondentov.

*Očakávaný výsledok:* Identifikácia extrémne vysoko úspešných, alebo extrémne nízko úspešných otázok.

*Hodnotenie:* Premenné definované pre hodnotenie úspešnosti otázok boli transformované (odmocninová transformácia). Pre transformované premenné z premenných týkajúcich sa otázok bol overený a splnený predpoklad normality pre trojrozmerný výber. V trojrozmernom súbore boli vyhladané odľahlé pozorovania.

**Tab. 9 Výsledok experimentu hodnotenia kvality otázok**

	ot.11	ot.12	ot.13	ot.14	ot.21	ot.22
Počet správnych odpovedí	48	10	42	8	39	1
Počet nesprávnych odpovedí	1	37	6	33	4	34
Počet, koľkokrát otázka nebola zodpovedaná	1	3	2	9	7	15
spolu	50					
	ľahká	ťažká	normálna	ťažká	normálna	ťažká

*Zhrnutie:* Testom odľahých pozorovaní boli identifikované ako odlišné (v štruktúre správnych odpovedí, nesprávnych odpovedí a vynechaní odpovede) otázky 11, 12, 14 a 22. Z nich je možné klasifikovať otázku 11 ako príliš jednoduchú, a otázky 12, 14 a 22 ako príliš ťažké. U týchto troch otázok je potrebné skontrolovať formuláciu otázky, prípadne v budúcnosti v procese výučby zdôrazniť oblasti, ktorých sa uvedené otázky dotýkajú.

Hodnotenie a zvyšovanie kvality zadávaných otázok je tiež súčasťou zvyšovania objektivnosti hodnotiaceho procesu.

## Záver

Na základe našich skúseností s používaním elektronického testu v procese výučby je možné konštatovať, že využitie takéhoto prostriedku významne prispieva k zvýšeniu efektívnosti procesu priebežného a aj záverečného hodnotenia vedomostí študentov. Z výsledkov testovania vyplývajú tri výstupné kategórie – bodové hodnotenie študenta, faktory ovplyvňujúce úspešnosť výsledkov a indikácia odlišných otázok.

Jednoduchá konfigurácia parametrov testu umožňuje prispôsobiť jeho nastavenie konkrétnym podmienkam a požiadavkám vyučujúceho. Medzi hlavné prednosti testu patrí skutočnosť, že okamžite po ukončení vypracovania testu má vyučujúci k dispozícii výsledky bodového hodnotenia, uložené v databázovej tabuľke.

Pozornosť sme venovali hľadaniu vplyvu vybraných faktorov na úspešnosť študentov (hodnotenú boli faktory opakovanie testu, dĺžka časového limitu, poznanie otázok a komunikácie medzi študentmi) a skúmaniu úspešnosti zodpovedania jednotlivých otázok (hodnotenú boli faktory poznania otázok a typu odpovede). Faktory opakovania, poznania otázok a komunikácie majú štatisticky významný vplyv na úspešnosť výsledkov študentov. Opakovaním testu dochádza k nadobudnutiu istoty študenta v rámci učebnej látky pri priebežnom hodnotení. Naopak, opakovanie testu viackrát po sebe pri záverečnom hodnotení nie je vhodné, ale je lepšie mať skupinu testov (aspoň 3 alebo 4), ktoré sa potom môžu cyklicky opakovať. Zadávanie tvorby otázok študentom (a nasledovné poznanie otázok) vhodne iniciuje priebežnú prípravu, zopakovanie si učiva, ale nezaručuje preniknutie do hĺbky a podstaty, vedie viac ku memorovaniu konkrétnych správnych odpovedí, ako ku dôkladnému štúdiu problematiky. Preto je vhodné optimalizovať štruktúru otázok a zväžiť podiel takýchto otázok pri hodnotení vedomostí.

Štatisticky významný vplyv na úspešnosť otázok bol preukázaný pre faktor poznania otázok a pre faktor typ odpovede. Na základe týchto informácií učiteľ môže pre hodnotenie vedomostí vytvoriť test, ktorý bude mať ním požadovanú náročnosť.

Faktor dĺžky časového limitu nemá štatisticky významný vplyv (záleží na formulácii a dĺžke otázok a odpovedí). Naše skúsenosti sú asi 20 minút na 60 otázok, pričom minimálny spotrebovaný čas bol pod 10 minút.

Celkové hodnotenie výsledkov jednotlivých otázok (faktor kvality) poskytuje učiteľovi istú spätnú väzbu vo forme indikácie príliš jednoduchých, resp. príliš ťažkých (alebo nepochopených) otázok. Táto indikácia umožňuje na jednej strane vykonať úpravu otázok a odpovedí testu, na druhej strane korekciu či doplnenie výučby. Objektivizáciu odlišných otázok umožňuje test odľahých pozorovaní.

## Literatúra

- [1] BENKOVÁ, M., FLOREKOVÁ, Ľ.: Informatizácia vysokoškolského štúdia. In: *Informatika a algoritmy '2000* : Vedecká konferencia s medzinárodnou účasťou : Prešov 7. a 8. septembra 2000. Prešov: TU-FVT, 2000. s. 9-11. ISBN 80-88941-13-X.
- [2] FLOREKOVÁ, Ľ., BENKOVÁ, M.: Komunikácia človek-stroj a medzilidská komunikácia vo výučbe na technických univerzitách. In: 80-88941-00-8. Information system in state and local

- government administration, the management of... / Zborník príspevkov. Košice. Košice : FVT TU, 1998. ISBN 80-88941-00-8
- [3] FLOREKOVÁ, Ľ., BENKOVÁ, M.: Štatistické metódy. Košice : FBERG TU, 1999. 120 s. ISBN 80-7099-411-8.
  - [4] HEBÁK, P. HUSTOPECKÝ, J.: Vícerozmerné štatistické metódy s aplikaciami, SNTL Alfa, Praha 1987.
  - [5] HOROVČÁK, P.: Internet Technologies Application on Knowledge Testing. In ICETA 2003 (2nd International Conference on Emerging Telecommunications Technologies and Applications and the 4th Conference on Virtual University) Conference Proceedings, elfa Košice, 2003, ISBN 80-89066-67-4, pp.239 – 244
  - [6] HOROVČÁK, P.: Didaktický test – nič jednoduchšie. Informatika v škole č. 26, 2003, ÚIPŠ Bratislava, ISSN 1335-616X, str. 3 - 9
  - [7] KOMENDA S., KLEMENTA J. Analýza náhodného v pedagogickém experimentu a praxi. SPN Praha, 1981, 316 str.
  - [8] ŽELEZNÍK, O., LAVRIN, A., - HURNÁ, S. Elektronická podpora dištančného vzdelávania. 1999. ISBN 80-88964-11-3 Spolupráca medzi univerzitou a priemyselnou a spoločenskou praxou / Zborník zo seminára s.54-58 Košice TU 1999

**Poznámka:** Príspevok bol riešený v rámci projektov KEGA 3/3084/05, KEGA 1/3126/05 (B), KEGA 1/3124/05 (L) a ICOTEL (LdV) SK/02/B/F/PP/-142261

Doc. Ing. Pavel Horovčák, CSc., Ing. Beáta Stehlíková, PhD.  
Technická univerzita v Košiciach  
Fakulta BERG  
Katedra informatizácie a riadenia procesov  
pavel.horovcak@tuke.sk, beata.stehlikova@tuke.sk

# Chosen aspects of electronic knowledge testing results' valuation

## Summary

Electronic test has its own unique place in the whole education process, as a highly formalised instrument of evaluation of the students' preparation and knowledge. The contribution is devoted to testing conditions, quantification of criteria for evaluation of results, analysis of success factors like: repetition of test, duration of testing, appropriate formulation of test questions, awareness of students about the questions. The article specifies selected aspects of quantification of criteria for evaluation (quality) of results of electronic knowledge testing.

The attention was devoted to examination of impact of selected factors on success of students (evaluated factors included: factor of quality, factor of time, factor of question recognition, and factor of communication among students), and on success of answering individual questions (evaluated factors included: factor of question recognition and factor of answer type). Factors of repetition, question recognition and communication have statistically significant effect on success of students. Through repetition of test, the student acquires certain level of security about the learning topic during the ongoing evaluation process. On contrary, repetition of test more times in a row is not appropriate during the final examination of students. It is more suitable to have a group of tests (3 or 4), which can be cyclically repeated. Assignment of question creation to students (and consequent recognition of questions) positively motivates ongoing preparation of students and repetition of the learning material, however, it does not guarantee deep understanding of the essence of the topic, since it leads more to memorization of specific correct answers than to in-depth study of the topic. Because of that reason, it is advisable to optimise structure of questions, and carefully choose the proper percentage of such questions on the test.

Statistically significant impact on success of questions was proved for the factor of question recognition and the factor of answer type. Based on this information, the teacher can design tests with appropriate level of difficulty for evaluation of students' knowledge. Factor of time has no significant statistical impact (it depends more on formulation and length of questions and answers). Our experience proved approximately 20 minutes for 60 questions, while the minimum completion time was below 10 minutes.

**Key words:** *electronic test, database, fruitfulness factors, statistical analysis*