

Príručka

**(vysvetlenie pojmov, používaných v záverečných správach zo
štatistického spracovania testov EČ MS)**

Oddelenie hodnotenia výsledkov meraní

ÚVOD	3
1 ZÁKLADNÉ INFORMÁCIE O EXTERNEJ ČASTI MS	4
1.1 Vyučovací jazyk	4
1.2 Cudzí jazyk	5
1.3 Postup výpočtu úspešnosti žiaka v teste EČ MS z vyučovacieho a cudzieho jazyka	6
1.4 Matematika	8
2 FÁZY SPRACOVANIA DÁT.	9
2.1 Kontrola dát	9
2.2 Prvotné vyhodnotenie	10
2.3 Položková analýza testu	Chyba! Záložka nie je definovaná.
2.4 Druhotné vyhodnotenie	16
ZÁVER	19
Literatúra	21

Úvod

Štandardizácia metodológie spracovania externej časti maturitnej skúšky (EČ MS) nás dovedla k rozdeleniu dokumentov štatistického vyhodnotenia na dve časti – *Správa...* a *Príručka*.

Správa zo štatistického spracovania obsahuje údaje a analýzy o danom teste. Informuje o štatistickom súbore – žiakoch, ktorí maturovali v danom predmete a úrovni, porovnáva výkony podskupín súboru z rôznych hľadísk a podrobne skúma kvalitu testu ako meracieho nástroja .

Príručka, ktorú predkladáme v tomto dokumente, informuje o spôsobe získavania a spracovania dát, vysvetľuje použité štatistické analýzy a dopĺňa niektoré matematické formule s tým spojené. Poradie častí a kapitol *Príručky* sleduje metodológiu štatistického spracovania.

Tento dokument je určený čitateľovi *Správy* a má dopomôcť k výkladu v nej uvádzaných štatistických výstupov.

Očakávame, že interpretácia štatistika spolu s interpretáciou tvorcu testu, skúseného pedagóga a ďalších odborníkov smerom k ďalším inštitúciám dopomôže ku komplexnému využitiu štatistického spracovania EČ MS.

1 Základné informácie o EČ MS

Žiaci sa do stanoveného termínu prihlasovali na EČ MS na pôde školy povinne vyučovacieho jazyka, povinne voľbou cudzieho jazyka a úrovne, nepovinne voľbou matematiky.

1.1 Vyučovací jazyk

Pre EČ MS z vyučovacích jazykov boli pripravené testy, ktoré obsahom a úrovňou náročnosti zodpovedali cieľovým požiadavkám na vedomosti a zručnosti absolventa strednej školy.

Cieľové predmety: slovenský jazyk a literatúra,
slovenský jazyk a slovenská literatúra,
maďarský jazyk a literatúra,
ukrajinský jazyk a literatúra.

Úlohou testu bolo preveriť čítanie s porozumením a jazykové kompetencie žiaka, zvládnutie literatúry, teórie literatúry, fonetiky, morfológie, lexikológie, syntaxe, ortografie a štylistiky. Zvládnutie rôznych myšlienkových operácií (analýza, syntéza,...), intelektuálnych a logických zručností.

Test EČ MS z vyučovacieho jazyka trval 90 minút a pozostával z 8 textov, na každý text sa vzťahovalo 5 položiek s výberom odpovede a 3 položky s krátkou odpoveďou, spolu 64 položiek. Všetky úlohy boli v kontexte. Zhrnutie uvádza nasledujúca tabuľka:

Cieľ	Preveriť pomocou produktívnych kontextualizovaných úloh, ako žiak zvládol rozsah učiva predpísaný Cieľovými požiadavkami na vedomosti a zručnosti absolventa strednej školy. Preveriť schopnosť žiaka čítať rôzne typy textov s porozumením a pracovať s textom.
Čas	90 minút
Počet textov	8 textov rôznej dĺžky
Charakteristika textov	Autentické, prípadne upravené texty lyrické, epické, dramatické, publicistické, náučné, vedecko-populárne na primeranej úrovni náročnosti. Texty sú obsahovo zamerané na témy vymedzené Cieľovými požiadavkami na vedomosti a zručnosti absolventa strednej školy.
Zameranie textov, zdroje	Texty literárne, náučné, publicistické, písané pre širokú verejnosť, ktoré spadajú do rámca žiakových skúseností - denníky (nie bulvár), časopisy, beletria, ...
Tematické okruhy	Texty obsahujú literárne ukážky, citácie, všeobecné témy, s ktorými sa žiaci stretávajú v každodennom živote, sú v rozsahu vedomostí a skúseností žiakov a sú v súlade s Cieľovými požiadavkami na vedomosti a zručnosti absolventa strednej školy.
Testované vedomosti a zručnosti	Celkové porozumenie textu. Overiť zvládnutie literatúry, teórie literatúry, fonetiky, ortografie, morfológie, lexikológie, syntaxe a štylistiky v rozsahu Cieľových požiadaviek na vedomosti a zručnosti absolventa strednej školy.

Počet testových položiek	64 z toho 40 úloh s výberom odpovede a 24 úloh s krátkou odpoveďou.
Druhy, typy testových úloh	Úlohy s výberom odpovede (výber zo štyroch možností). Doplňovacie – s krátkou odpoveďou.
Kritériá hodnotenia	Každá položka je hodnotená za správnu odpoveď 1 bod, za nesprávnu alebo žiadnu 0 bodov), maximálny počet bodov je 64.

1.2 Cudzí jazyk

Pre EČ MS v cudzom jazyku boli pripravené testy dvoch úrovní, test vyššej úrovne B2, základnej úrovne B1.

Maturitná skúška úrovne B2 je odporúčaná maturantom všetkých typov stredných škôl so študijnými odbormi, ktorí súčasne spĺňajú tieto dve podmienky:

1. pripravujú sa na maturitnú skúšku z cudzieho jazyka,
2. počas svojho štúdia dosiahli úroveň pokročilého (napr. v nemeckom jazyku *Mittelstufe*, v anglickom jazyku *upper-intermediate*).

Obsahom a úrovňou náročnosti zodpovedal test EČ MS úrovne B2 Spoločného európskeho referenčného rámca Rady Európy.

Maturitná skúška úrovne B1 je odporúčaná maturantom všetkých typov stredných škôl so študijnými odbormi, ktorí súčasne spĺňajú tieto dve podmienky:

1. pripravujú sa na maturitnú skúšku z cudzieho jazyka,
2. počas svojho štúdia dosiahli úroveň stredne pokročilého (napr. v nemeckom jazyku *Grundstufe 3*, v anglickom jazyku *intermediate*).

Obsahom a úrovňou náročnosti zodpovedal test EČ MS úrovni B1 Spoločného európskeho referenčného rámca Rady Európy.

Každá úloha testu – položka – musí korešpondovať s cieľovými požiadavkami na vedomosti a zručnosti maturantov z cudzieho jazyka pre príslušnú úroveň.

Testy úrovne B2 obsahujú **80 úloh**: 46 úloh s výberom odpovede, 34 úloh s krátkou odpoveďou. Za správnu odpoveď získal žiak 1 bod, za nesprávnu, alebo ak neodpovedal, 0 bodov. Testy boli vypracované s prihliadnutím na hodinovú dotáciu v danom predmete. Boli vytvorené dva varianty testu, ktoré sa líšili poradím úloh, resp. pri úlohách s výberom odpovede poradím alternatív odpovede.

Na vypracovanie testu externej časti úrovne B2 mali žiaci **120 minút**.

Testy úrovne B1 obsahujú **60 úloh**: 36 úloh s výberom odpovede, 24 úloh s krátkou odpoveďou. Za správnu odpoveď získal žiak 1 bod, za nesprávnu, alebo ak neodpovedal, 0 bodov. Testy boli vypracované s prihliadnutím na hodinovú dotáciu v danom predmete. Boli vytvorené dva varianty testov jednotlivých úrovní, ktoré sa líšili poradím úloh, resp. pri úlohách s výberom odpovede poradím alternatív odpovede.

Na vypracovanie testov externej časti úrovne B1 mali žiaci **100 minút**.

Obsahová štruktúra testov EČ z cudzích jazykov v úrovni B2

Časti testu	Položky	
	s výberom odpovede	s krátkou odpoveďou
Počúvanie	1 - 13	14 - 20
Gramatika	21 - 40	41 - 60
Čítanie	61 - 73	74 - 80

Tab 1. Rozdelenie položiek v teste EČ MS z cudzích jazykov v úrovni B2

Obsahová štruktúra testov EČ z cudzích jazykov v úrovni B1

Časti testu	Položky	
	s výberom odpovede	s krátkou odpoveďou
Počúvanie	1 - 13	14 - 20
Gramatika	21 - 30	31 - 40
Čítanie	41 - 53	54 - 60

Tab 2. Rozdelenie položiek v teste EČ MS z cudzích jazykov v úrovni B1

1. časť: **Počúvanie s porozumením** – v tejto časti majú žiaci preukázať, ako porozumeli základným informáciám z počutého textu. Riešili úlohy, ako je určenie správnosti alebo nesprávnosti výpovede, výber z viacerých možností, doplňovanie chýbajúcich informácií alebo priradovanie.

2. časť: **Gramatika a lexika** – testujú sa tu znalosti gramatických štruktúr a slovnej zásoby a ich správne použitie v kontexte. Žiaci majú do textových ukážok dopĺňať adekvátne gramatické tvary na základe výberu z viacerých možností, dopĺňať chýbajúce slová alebo časti slov, prípadne vyberať správnu lexikálnu jednotku z banky slov.

3. časť: **Čítanie s porozumením** – cieľom tejto časti je zistiť ako žiaci porozumeli obsahu textu, či pochopili jeho hlavné myšlienky, základné súvislosti a vyriešili úlohy doplňovacieho typu, výberu z viacerých možností, resp. určenie správnosti alebo nesprávnosti výpovede.

1.3 Postup výpočtu úspešnosti žiaka v teste EČ MS z vyučovacieho a cudzieho jazyka

Hodnotenie výkonu žiakov v testoch EČ MS vyjadrujeme celkovou úspešnosťou žiaka v percentách a tiež percentilom¹.

Celková úspešnosť žiaka v teste z cudzieho jazyka sa skladá z čiastkových výsledkov žiaka v troch častiach testu tak, aby každá časť prispievala k výsledku žiaka rovnakou váhou. V súčasnosti sú to tri obsahové časti, ako sme uviedli vyššie, a každá z týchto častí meria inú vlastnosť (vedomosť, zručnosť) žiakovho výkonu.

Ak predpokladáme, že zručnosť žiakov počúvať s porozumením, správne používať gramatické javy a slovnú zásobu a čítať s porozumením v cudzom jazyku, sú tri rozdielne

¹ Percentil vyjadruje umiestnenie žiaka vzhľadom na jeho poradie v usporiadanom súbore všetkých žiakov podľa úspešnosti.

vlastnosti žiakovho výkonu, potom tieto časti musíme chápať ako samostatné celky², pre ktoré je nutné zvlášť počítat' výsledky a sledovať reliabilitu. Výsledné hodnotenie testu potom získame zložením výsledkov jednotlivých častí.

Pri určovaní postupu výpočtu celkovej úspešnosti žiaka sme prihliadali na svetové trendy, ktoré kladú dôraz na schopnosť aktívne používať jazyk. Celkovú úspešnosť v teste preto počítame aritmetickým priemerom úspešností žiaka v troch častiach testu, aby sa jednotlivé časti podieľali na celkovej úspešnosti žiaka rovnakou váhou³.

Počet testových položiek úrovne B2 v jednotlivých častiach je 20 : 40 : 20. Výpočet celkovej úspešnosti žiaka v teste EČ MS z cudzích jazykov je nasledovný:

- Správna odpoveď žiaka v každej z položiek v každej časti testu je hodnotená 1 bodom. Nesprávna alebo žiadna odpoveď je hodnotená 0 bodmi.
- K dosiahnutým bodovým hodnotám v každej časti testu sa vypočíta úspešnosť v percentách.

Úspešnosť ($Ú$) v jednotlivých častiach testu sa počíta podľa nasledujúcich vzorcov:

$$\text{Počúvanie s porozumením } (p): \quad Ú_p = \frac{PSO_p}{20} \cdot 100,$$

$$\text{Gramatika a lexika } (g): \quad Ú_g = \frac{PSO_g}{40} \cdot 100,$$

$$\text{Čítanie s porozumením } (č): \quad Ú_č = \frac{PSO_č}{20} \cdot 100.$$

PSO_{\square} - počet správnych odpovedí v príslušnej časti

7

- Váha jednotlivých častí na celkovej úspešnosti žiaka je 1 : 1 : 1. Pri výpočte jednoduchého aritmetického priemeru sa používa vzorec:

$$Ú = \frac{Ú_p + Ú_g + Ú_č}{3}.$$

- Celkovú úspešnosť žiaka zaokrúhľujeme na jedno desiatinné miesto.

Počet testových položiek úrovne B1 v jednotlivých častiach je 20 : 20 : 20. Celkovú úspešnosť žiaka v teste z cudzieho jazyka úrovne B1 určujeme podľa vzt'ahu

² Rovnaké postupy používajú testovacie inštitúcie v zahraničí. Uvádzame príklad spôsobu spracovania výsledkov testov v ČR: „Štatistické charakteristiky sa zisťujú z celkových výsledkov uchádzačov, pokiaľ sa však písomná časť skúšky skladala z viacerých častí (napr. testy z rôznych predmetov, **testovanie rôznych zručností**, testy v niekoľkých kolách), alebo mala niekoľko variantov, je nutné pracovať s výsledkami uchádzačov **za každú časť zvlášť**.“ http://www.Scio.cz/tvorba_testu/hodnoceni_kvality/zakl_stat.htm. [10]

³ Podobne je počítaný výsledok v skúške TOEFL (TEST OF ENGLISH AS A FOREIGN LANGUAGE): „Pri papierovej forme TOEFL testu je možné dosiahnuť výsledok maximálne 677 bodov a minimálne 310. Výsledok pozostáva zo sčítania dosiahnutých bodov z častí „počúvanie“, „jazyková skladba“ a „čítanie“ a ich následného vynásobenia desiatimi a delením tromi (maximálny počet bodov získaných v jednej časti je 68 a minimálny 32).“ (Jazykové skúšky a štandardizované testy 1. časť: In: Bulletin SAIA Slovenská akademická informačná agentúra, Informačný mesačník o štúdiu v zahraničí č. 9, ročník XIV, september 2004.) [5]

$$\dot{U} = \frac{PSO}{60} \cdot 100,$$

kde PSO je počet správnych odpovedí.

Príklad:

Ak žiak X správne odpovedal na 9 položiek v časti počúvanie s porozumením, na 38 položiek v časti gramatika a lexika a na 19 položiek v časti čítanie s porozumením, jeho úspešnosť počítame nasledovne:

Časti testu	Počet správnych odpovedí žiaka X	Úspešnosť žiaka X	Celková úspešnosť žiaka X
Počúvanie s porozumením	9	45,0%	78,3%
Gramatika a lexika	38	95,0%	
Čítanie s porozumením	19	95,0%	

Tab 3. Príklad výpočtu celkovej úspešnosti žiaka v teste cudzích jazykov EČ MS

Počet testových položiek v jednotlivých častiach úrovne B1 je 20:20:20, preto sa výpočet úspešnosti v prípade úrovni B1 líši od výpočtu v úrovni B2 v \dot{U}_g , kde je počet položiek 20, takže

$$\dot{U}_g = \frac{PSO_g}{20} \cdot 100.$$

1.4 Matematika

V predmete matematika boli vypracované testy jednej úrovne, ktoré obsahovali **30 úloh**, z toho 10 úloh s výberom odpovede a 20 úloh s krátkou odpoveďou. Za správnu odpoveď získal žiak 1 bod, za nesprávnu (alebo ak neodpovedal) 0 bodov. Test bol pripravený s prihliadnutím na hodinovú dotáciu v predmete matematika a boli vytvorené dva varianty, ktoré sa líšili poradím úloh, resp. pri úlohách s výberom odpovede poradím alternatív odpovede.

Na vypracovanie testu externej časti mali žiaci **120 minút**.

Položky	s krátkou odpoveďou	1 - 20
	s výberom odpovede	21 - 30

Tab 4. Rozdelenie položiek v teste EČ MS z matematiky

2 Fázy spracovania dát

Odpovede všetkých testov externej časti maturitnej skúšky zapisovali žiaci do samoprepisovacích odpoved'ových hárkov (OH). Kópiu OH overil žiak podpisom, originál podpísal po vyhodnotení hodnotiteľ. Oddelené kópie bez zásahu hodnotiteľa alebo inej osoby sú archivované na školách.

V cudzích jazykoch hodnotenie riešení testov otvorených úloh s krátkou odpoveďou vykonával hodnotiteľ pod externým dozorom podľa centrálne vypracovaných pokynov. Predsedovia PMK kontrolovali hodnotenie a zodpovedajú za jeho správnosť.

Hodnotenie testov z matematiky bolo spracovávané automatizovane v software SPSS podľa kľúča správnych odpovedí.

Originály všetkých typov odpoved'ových hárkov boli zaslané na centrálne spracovanie, následne skenované, automaticky vyhodnocované a postúpené na štatistické spracovanie v software SPSS.

V prvej fáze sme urobili kontrolu dát. Nasledovalo prvotné spracovanie údajov. Fáza položkovej analýzy overovala kvalitu testu. Poslednou fázou bolo kvantitatívne a kvalitatívne štatistické vyhodnotenie súboru.

2.1 Kontrola dát

Po elektronickom spracovaní odpoved'ových hárkov sme v rámci kontroly kvality dát vykonali procedúry súvisiace s jednotlivými premennými:

- kontrola úplnosti naskenovania dát,
- kontrola kódu školy,
- kontrola označenia variantov testu (kódov testov),
- kontrola kódu žiaka⁴ a jeho duplicitnosti v databáze,
- kontrola chýbajúceho označenia pohlavia žiaka,
- kontrola prepojenia kódu a pohlavia žiaka,
- kontrola chýbajúceho uvedenia známky žiaka⁵,
- kontrola bodovania,
- kontrola správnosti kľúčov odpovedí.

Cieľom uvedených kontrolných procedúr bolo vyčistiť dáta, zvýšiť ich validitu⁶ a prispieť k zvýšenej hodnovernosti a reliabilite⁷ spracovaných výsledkov. Výsledky prvej fázy spracovania dát sme sumarizovali vo forme kontrolných protokolov pre jednotlivé testy, ktoré umožňujú kedykoľvek verifikovať proces spracovania dát.

⁴ Kód žiaka obsahuje rodné číslo žiaka. Databáza neobsahovala meno a priezvisko žiaka.

⁵ Klasifikačný stupeň žiaka v 1. polroku 4. ročníka z predmetu, v rámci ktorého písal test externej časti maturitnej skúšky.

⁶ Hodnovernosť, že test meria to, čo deklaruje, že meria; miera vhodnosti testu pre daný účel

⁷ Spoľahlivosť merania spočíva v tom, že keď test zopakujeme, získame tie isté výsledky, čiže čo možno najmenší rozptyl nových hodnôt okolo pôvodných.

Koeficient reliability r teoreticky nadobúda hodnoty z $(-1; 1)$. My posudzujeme $|r| z (0; 1)$ a za prijateľnú považujeme hodnotu $r \geq 0,80$.

2.2 Prvotné vyhodnotenie

Kvantitatívne analýzy opisujú štatistický súbor, teda žiakov, ktorí sa do maturity v danom predmete zapojili. Ďalej sme vypočítali prvotné **charakteristiky testu** obsahujúce percentuálne parametre úspešnosti a určili prvotný histogram rozloženia početností percentuálnych úspešností.

Úspešnosť definujeme ako percentuálny podiel bodov za položky, na ktoré žiak správne odpovedal z celkového počtu bodov, ktoré mohol v teste získať.

Maximum je najvyššia dosiahnutá úspešnosť, maximálny počet percent, ktoré dosiahol niektorý z testovaných žiakov.

Minimum je najnižšia dosiahnutá úspešnosť, minimálny počet percent, ktoré dosiahol niektorý z testovaných žiakov.

Priemerná úspešnosť (\bar{x} , Mean) je počítaná podľa vzorca pre aritmetický priemer,

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} .$$

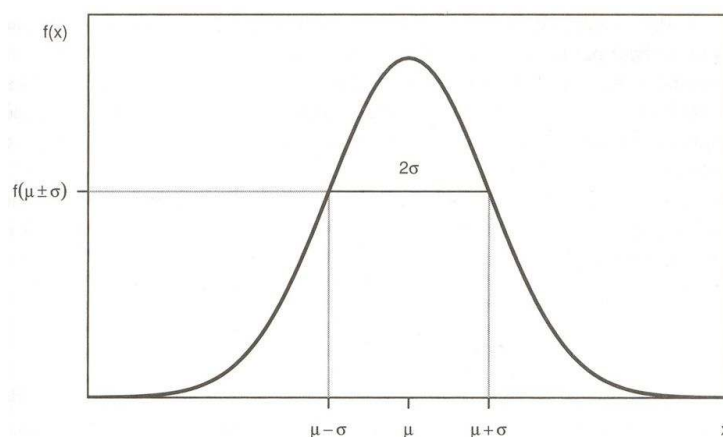
x_i ... úspešnosť i-teho žiaka

n ... počet žiakov

Štandardná odchýlka (SD - Standard Deviation) vyjadruje mieru rozptýlenia úspešnosti žiakov od aritmetického priemeru, zjednodušene je to priemer nameraných odchýliek od aritmetického priemeru. Čím je SD väčšie, tým je úspešnosť rozptýlenejšia a teda existujú veľké rozdiely vo výkonoch žiakov. S rastúcim počtom testovaných žiakov štandardná odchýlka spravidla klesá.

$$SD = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} .$$

Štandardnú odchýlku σ za predpokladu normálneho rozloženia interpretujeme v histograme rozloženia početností nasledovne:



- interval $\mu \pm \sigma$ obsahuje 68,8 % populácie,
- interval $\mu \pm 2\sigma$ obsahuje 95,5 % populácie,
- interval $\mu \pm 3\sigma$ obsahuje 99,7 % populácie;

alebo naopak

- 95,0 % populácie je v intervale $\mu \pm 1,96\sigma$,
 - 99,0 % populácie je v intervale $\mu \pm 2,58\sigma$,
 - 99,9 % populácie je v intervale $\mu \pm 3,29\sigma$,
- kde μ je stredná hodnota, napr. \bar{x} .

Intervalový odhad úspešnosti populácie počítame zo štandardnej odchýlky. Pri použití pravidla približne 2 štandardných odchýlok určuje interval, v rozmedzí ktorého sa umiestnilo 95% testovaných žiakov,

$$\langle -1,96 \cdot SD ; 1,96 \cdot SD \rangle.$$

Štandardná chyba priemernej úspešnosti (SE – Standard Error Mean) dokumentuje s akou presnosťou je vypočítaná hodnota priemernej úspešnosti. Čím je menšia štandardná chyba priemeru, tým presnejšie charakterizuje priemer testovanú populáciu,

$$SE = \frac{SD}{\sqrt{n}}.$$

Obvykle platí, že čím väčšia vzorka (čím väčšie n), tým menšia štandardná chyba.

Interval spoľahlivosti pre populačný priemer počítame zo štandardnej chyby priemeru. V našej práci reprezentuje interval

$$\langle -1,96 \cdot SE ; 1,96 \cdot SE \rangle,$$

v ktorom sa s 95%-nou pravdepodobnosťou nachádza populačný priemer (pravidlo dvojnásobku SE).

Cronbachovo alfa je koeficient reliability testu. Pri binárnom hodnotení položiek je špeciálnym prípadom Cronbachovho α (1) Kuder-Richardsonov vzorec KR_{20} (3):

$$Cronbachovo \alpha = \frac{k}{k-1} \cdot \frac{SD^2 - \sum_{i=1}^k SD_i^2}{SD^2}, \quad (1)$$

V prípade binárneho hodnotenia platí

$$SD_i^2 = p_i(1 - p_i), \quad (2)$$

a teda

$$KR_{20} = \frac{k}{k-1} \cdot \frac{SD^2 - \sum_{i=1}^k p_i \cdot (1 - p_i)}{SD^2} \quad (3)$$

je Cronbachovho α , pričom

$k...$	počet úloh v teste
$p_i...$	úspešnosť i-tej úlohy (aká časť žiakov ju vyriešila správne, $p_i \in \langle 0; 1 \rangle$)
$SD^2 ...$	celková variabilita
$SD_i^2 ...$	variabilita i-tej položky.

Hlavné faktory ovplyvňujúce reliabilitu testu:

- Počet položiek – čím dlhší test, tým vyššia reliabilita;
- Obťažnosť položiek – reliabilitu znižuje veľký počet príliš ľahkých alebo ťažkých položiek;
- Diskriminačná sila položiek, parametrom ktorej je *P.Bis.*, ktorý je vysvetlený v položkovej analýze a vystihuje rozlišovaciu silu položky podľa výkonnosti žiaka;
- Časový limit riešenia testu – ak je rýchlosť riešenia položiek významný faktor úspešnosti, reliabilita bude umelo zvýšená;
- Rozptyl výkonnosti testovaných žiakov – čím je väčší (čím je súbor rôznorodjší), tým je vypočítaná hodnota reliability vyššia.

Štandardná chyba merania (SEM – Standard Error of Measurement) je ukazovateľom presnosti merania, do ktorého vstupujú faktory štandardná odchýlka a reliabilita testu. Tento ukazovateľ má význam pri určovaní intervalového odhadu skutočnej úspešnosti individuálneho žiaka. Čím je štandardná chyba merania menšia (čím je nižšia štandardná odchýlka a vyššia reliabilita), tým presnejšie je určený bodový odhad individuálneho výsledku žiaka,

$$SEM = \sqrt{(SD^2 \cdot (1 - \alpha))}.$$

Intervalový odhad úspešnosti individuálneho žiaka je interval, v ktorom sa s 95%-nou pravdepodobnosťou nachádza skutočná úspešnosť individuálneho žiaka. Tento interval dostaneme, ak k nameranej úspešnosti žiaka pripočítame a odpočítame 1,96-násobok štandardnej chyby merania:

$$\langle x_i - 1,96 \cdot SEM; x_i + 1,96 \cdot SEM \rangle.$$

2.3 Položková analýza testu

Prvým krokom je overenie ekvivalencie variantov testu. Porovnali sme súbory žiakov z hľadiska administrácie variantov podľa krajov, zriaďovateľa, typu školy a pohlavia. Potom sa zisťuje štatistická signifikancia zistených rozdielov a počíta sa aj vecná signifikancia, aby sa zistilo, či medzi variantmi je rozdiel natoľko významný z pedagogického hľadiska, aby sme sa ním museli zaoberať. Pokiaľ sa vecná signifikancia rozdielu nepotvrdí, teda oba varianty daného testu sú z hľadiska obťažnosti porovnateľné, môžeme žiakov, ktorí písali akýkoľvek variant testu, medzi sebou porovnávať.

Po dôkaze ekvivalentnosti variantov sme jeden variant – zástupný – podrobili položkovej analýze, ktorá sleduje vlastnosti jednotlivých položiek: obťažnosť, citlivosť,

neriešenosť, vynechanosť, nedosiahnutosť, medzipoložkovú koreláciu a ďalšie, ktoré dokladujú charakter a kvalitu testu.

Charakteristiky položkovej analýzy

Úspešnosť položky je percentuálny podiel žiakov, ktorí správne riešili danú položku. Ak je hodnotenie zložitejšie, úspešnosť počítame ako percentuálny podiel počtu bodov, ktoré žiaci získali z počtu bodov, ktoré mohli získať. Čím je úspešnosť v riešení danej položky nižšia, tým je položka **obťažnejšia**.

Obťažnosť položky je opakom jej úspešnosti. Vzťah medzi obťažnosťou a úspešnosťou položiek je nasledovný:

$$\text{obťažnosť} = 100\% - \text{úspešnosť}.$$

Z hľadiska zabezpečenia vysokej reliability je optimálne, aby sa obťažnosť väčšiny položiek pohybovala okolo 50 %.

Za *veľmi obťažné* pokladáme položky s obťažnosťou 80 % a viac (nad 90 % hovoríme o *extrémnej* obťažnosti), *lahké* sú položky s obťažnosťou pod 20 % (pod 10 % hovoríme o *extrémne ľahkých* položkách).

Medzipoložková korelácia (P.Bis)

Test je reliabilný, ak sú jeho položky homogénne – vnútorne konzistentné. Vnútorná homogenita - konzistencia spočíva v tom, že jednotlivé položky medzi sebou korelujú, pretože merajú spravidla tú istú vlastnosť. **Medzipoložková korelácia je koreláciou medzi položkou a zvyškom testu** a vyjadrujeme ju prostredníctvom koeficientu korelácie P.Bis. (Point Biserial) medzi úspešnosťou vybranej položky a sumou úspešností všetkých ostatných položiek.

$$r_{P.Bis.} = \frac{M_1 - M_0}{SD} \sqrt{\frac{N_1 N_0}{N(N-1)}}$$

pričom

N ... je celkový počet žiakov,

N_1 ... je počet tých žiakov, ktorí v skúmanej otázke odpovedali správne,

N_0 ... je počet tých žiakov, ktorí v skúmanej otázke odpovedali chybné,

M_1 ... priemerný počet bodov za celý test bez skúmanej položky, ktorí získali tí žiaci, ktorí skúmanú otázku zodpovedali správne,

M_0 ... priemerný počet bodov za celý test bez skúmanej otázky, ktorí získali tí žiaci, ktorí skúmanú otázku zodpovedali chybné,

SD ... štandardná odchýlka celkového počtu bodov za test bez skúmanej otázky.

Záporná hodnota vypovedá o tom, že položka nerozlišuje „dobrých“ a „slabších“ žiakov. Dobrí žiaci, ktorých úspešnosť v teste bola nadpriemerná, na položku vo väčšine prípadov odpovedali nesprávne a naopak slabší žiaci, ktorí dosiahli v teste nižšiu úspešnosť, uviedli správnu odpoveď, napríklad hádaním. Ak je táto hodnota blízka 0 položka taktiež slabo rozlišuje výkonných a menej úspešných žiakov. Aby sme položku považovali za vhodnú, hodnota medzipoložkovej korelácie musí dosahovať úroveň minimálne 0,20. Položka s hodnotou 0,25 a vyššou je v pedagogických meraniach považovaná za dobrú.

V našich grafoch usporiadania položiek podľa medzipoložkovej korelácie sme hodnotu $r_{P.Bis.}$ uvádzali v 100 násobku pre lepšiu čitateľnosť. V cudzích jazykoch vypočítaný koeficient $r_{P.Bis.}$ určuje medzipoložkovú koreláciu aj v danej časti testu (Počúvanie, Gramatika a Čítanie).

Kľúče a distraktory

V prípade položiek s výberom odpovede má žiak v ponuke okrem (jedinej) správnej aj jednu alebo viac nesprávnych odpovedí, ktoré nazývame **distraktory**.

Cieľom skúmania je zistiť, aká je proporcionalita voľby distraktorov a správnej odpovede, usudzovať, akí žiaci volili správnu odpoveď, akí jednotlivé distraktory. Pomáha charakterizovať položku z viacerých aspektov súvisiacich s diskriminačnou silou položky, akými sú obťažnosť, formulácia problému/otázky v položke, jednoznačnosť správnej odpovede a podobne.

Výstupom tejto analýzy je tabuľka pre každú položku. V hlavičke tabuľky sú uvedené možnosti odpovedí A, B, C, D (podľa počtu distraktorov), označenie X sa vzťahuje na žiakov, ktorí neodpovedali.

V prvom riadku tabuľky sú uvedené hodnoty medzipoložkovej korelácie ($P.Bis.$). V druhom riadku tabuľky p znamená podiel žiakov, ktorí si vybrali danú možnosť. V treťom riadku tabuľky N znamená počet žiakov, ktorí si vybrali danú možnosť.

Správna odpoveď je súčasťou kľúča, ktorý je v tabuľkách vyznačený žltou farbou.

Kritériá hodnotenia položiek s výberom odpovede:

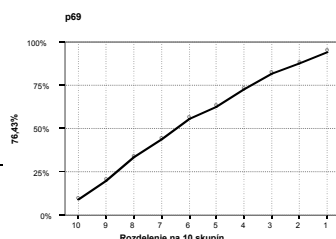
1. podiel žiakov, ktorí si vybrali správnu odpoveď (kľúč) by mal byť najväčší,
 2. hodnota $P.Bis.$ pri správnej odpovedi by mala byť väčšia ako 0,20 (optimálne väčšia ako 0,25).
 3. hodnota $P.Bis.$ pri nesprávnej odpovedi (distraktore) by mala byť hodnota záporná.
- Akékoľvek nedodržanie týchto kritérií zvyčajne zdôrazňujeme červenou alebo hnedou farbou. Červená farba identifikuje v správnej odpovedi hodnotu $P.Bis.$ menšiu ako 0,20. Hnedá farba identifikuje distraktory, ktoré majú kladnú hodnotu medzipoložkovej korelácie ($P.Bis.$).

Citlivosť alebo diskriminačná sila položky je schopnosť položky rozlíšiť dobrých a slabších žiakov. Zoradených žiakov podľa úspešnosti v teste rozdelíme do piatich (prípadne desiatich) skupín. Citlivosť položky predstavuje rozdiel medzi priemernou úspešnosťou najslabšej a najlepšej pätiny (desatiny) testovaných žiakov. Citlivosť nižšiu ako 10 % považujeme za nedostatočnú, záporná hodnota identifikuje kritickú položku.

Distribúcia úspešnosti a citlivosť sa interpretuje grafmi, ktoré podrobnejšie znázorňujú vzťah medzi výkonnosťou v celom teste (delenie na 10 skupín na x-ovej osi) a ich priemernými úspešnosťami v danej položke na y-ovej osi. V prvej desatine (1) sa nachádzajú žiaci s najvyššou percentuálnou úspešnosťou riešenia testu, poslednú skupinu (10) tvoria najmenej úspešní žiaci. Grafy distribúcie úspešnosti pomáhajú charakterizovať, ako v danej položke súvisí úspešnosť jej riešenia s celkovou výkonnosťou žiaka. Vľavo pri každom grafe je uvedená citlivosť položky.

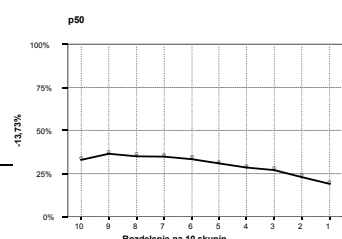
Podľa výslednej čiary rýchlo odhadneme

- charakter „dobrej“ položky,

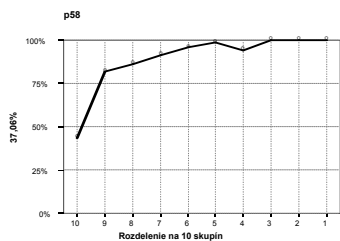


alebo úplne nevhodnej položky;

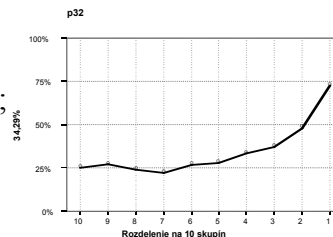
NÚCEM, Bratislava 2009



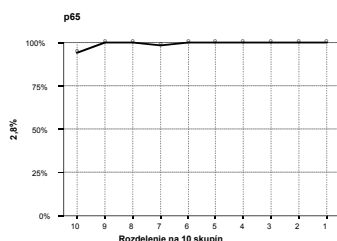
- položky, ktorá výrazne oddeľuje najslabších žiakov od ostatných,



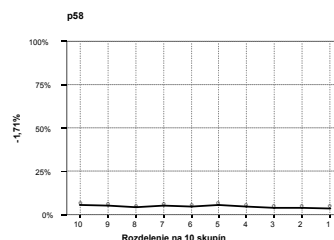
alebo naopak - najlepších od ostatných;



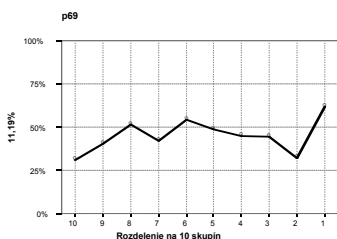
- položky, ktorá slabo diferencuje, pretože je pre všetkých veľmi ľahká,



alebo veľmi ťažká.



- Iná položka môže postihovať špeciálne schopnosti, ktoré nie sú pre didaktické testy cieľom testovania, ale toto posúdenie je na skúsených pedagógoch a tvorcach testov, pretože takýto graf



môže byť aj dôsledkom napríklad málopočetného súboru.

Neriešenosť položky predstavuje percentuálny súčet nedosiahnutých a vynechaných riešení. **Vynechanosť položky** nastáva vtedy, ak žiak položku neriešil, ale niektorú z nasledujúcich ešte riešil. **Nedosiahnuté** položky sú tie, ktoré žiak napríklad pre nedostatok času neriešil. Za nedosiahnutú považujeme každú položku, po ktorej žiak už žiadnu z položiek neriešil. Poslednú položku v teste posudzujeme podľa predposlednej položky v teste. Predpokladáme, že nedosiahnutosť poslednej položky v teste je rovnaká ako nedosiahnuteľnosť predposlednej položky. Vyššia ako 30 % neriešenosť položky by mala tvorcov testu vyzvať k úvahe nad príčinami, viac ako 50 % neriešenosť, obzvlášť pri veľkých súboroch, môžeme považovať za kritickú.

V prípade odhalenia formálnej chyby alebo iných nevhodných parametrov položky sme navrhli úpravu bodovania. Každému žiakovi bol priznaný v takejto položke 1 bod, bodovanie ostatných položiek zostalo nezmenené.

2.4 Druhotné vyhodnotenie

Po úprave bodovania sme určili druhotné charakteristiky testu a príslušný histogram. Stanovili sme úspešnosť a percentil každému žiakovi, čo je primárny cieľ testovania.

Výkonnosť žiakov v teste sme popisovali nasledovnými štatistickými charakteristikami:

Úspešnosť, ktorá bola vysvetlená vyššie; hodnota zaokrúhlená na dve desatinné miesta sa uvádza na maturitnom vysvedčení.

Percentil stanovuje percentuálne poradie žiaka v testovanej populácii.

Neoddeliteľnou súčasťou hodnotenia výsledku žiaka v testoch EČ MS sa stal percentil, ktorý sa uvádza aj na maturitnom vysvedčení. Tento údaj vypočítame tak, že účastníkov testovania usporiadame do poradia podľa dosiahnutej celkovej úspešnosti v teste a ich poradie vyjadríme hodnotou percentilu na stupnici 0 - 100. Hodnotu percentilu zaokrúhľujeme na jedno desatinné miesto. Percentil vyjadruje, aké percento žiakov dosiahlo horší výsledok.

Napríklad, ak sa žiak umiestnil v 72. percentile, 72 % žiakov z testovaných žiakov dosiahlo horší výsledok v teste a 28 % žiakov dosiahlo lepší alebo rovnaký výsledok. Hodnotu úspešnosti žiaka je dôležité sledovať v súvislosti s percentilom. Ak bude žiak riešiť test s úspešnosťou napr. 92 % a umiestni sa v 75. percentile, jeho výsledok interpretujeme nasledovne: učivo obsiahnuté v teste žiak zvládol na 92 % a $\frac{3}{4}$ žiakov získalo horší výsledok v danom testovaní. Iný príklad, kedy žiak riešil test s úspešnosťou 70 % a umiestnil sa v 85. percentile, interpretujeme nasledovne: testované učivo obsiahnuté v teste žiak zvládol na 70 % a 85 % žiakov riešilo test horšie ako porovnávaný žiak.

Ďalej analyzujeme rozdiely v dosiahnutých úspešnostiach rôznych skupín žiakov podľa zvolených kritérií: typ školy, kraj, zriaďovateľ, pohlavie a známka, ktorú žiaci získali na polročnom vysvedčení v 4. ročníku v danom predmete.

Rozdiely podľa typu školy

Pre účely vyhodnocovania výsledkov EČ MS rozdeľujeme školy na štyri skupiny:

Gymnázia

Ostatné:

stredné priemyselné školy
stredné vinárske, poľnohospodárske, lesnícke, ... školy
obchodné akadémie
umelecké školy
zdravotnícke školy
stredné odborné školy, dievčenské odborné školy,
združené stredné školy

Rozdiely podľa kraja sú dané územným spravovaním štátu. Používali sme nasledovné skratky:

BA Bratislavský kraj,
TT Trnavský kraj,
TN Trenčiansky kraj,
NR Nitriansky kraj,
ZA Žilinský kraj,
BB Banskobystrický kraj,
PO Prešovský kraj,
KE Košický kraj.

Rozdiely podľa zriaďovateľa

Zriaďovateľom školy môže byť: štát,
cirkev,
súkromník.

Rozdiely podľa známky

Použili sme údaje klasifikácie žiaka na polročnom vysvedčení. V niektorých prípadoch žiaci známku neuvedú. Neuvedené známky označujeme v tabuľkách číslom 9.

Uvedené rozdiely výkonov podskupín žiakov sú overované t - testom a ANOVA a zisťovaním vecnej významnosti rozdielu.

Štatistická signifikancia (Sig.)

Pri porovnávaní dvoch súborov (napríklad skupiny dievčat a skupiny chlapcov) vychádzame z predpokladu, že medzi súbormi nie je rozdiel vo výkone. Toto tvrdenie nazývame *Nulová hypotéza* a označujeme ho H_0 . Spolu s H_0 uvažujeme o *Alternatívnej hypotéze* H_1 , ktorá tvrdí opak H_0 , teda že porovnávané charakteristiky daných súborov sa nerovnajú. Na začiatku testovania si stanovíme úroveň spoľahlivosti p , to je miera rizika prijatia alternatívnej hypotézy. Bude to pravdepodobnosť, že naše tvrdenie môže byť nepravdivé s mierou pravdepodobnosti: $p=0,05$ (alebo $p=0,01$ či $p=0,001$), čo je možnosť omylu 5% (alebo 1%, či 0,1%). Ak nám výpočtom vyjde, že štatistická významnosť – signifikancia je $p \leq 0,05$, alternatívnej hypotéze môžeme veriť na 95% a viac. Konštatujeme, že medzi súbormi na zvolenej hladine významnosti je štatisticky významný rozdiel v charakteristike, ktorú sledujeme. V tomto prípade hovoríme, že rozdiel je štatisticky významný - rozdiel je signifikantný a H_0 zamietame.

Ak nám výpočtom vyjde $p > 0,05$, nulovú hypotézu na zvolenej hladine významnosti nezamietame. Zistený rozdiel medzi charakteristikami môžeme považovať za náhodný, povieme, že rozdiel nie je signifikantný.

Pre správnu interpretáciu treba zdôrazniť, že nezamietnutie nulovej hypotézy ešte neznamená, že je správna. Je to len výrok, ktorý sa vzťahuje na konkrétne dáta a tieto v danom prípade nám nedovolili nulovú hypotézu zamietnuť. Lepšie povedané „Nepodarilo sa dokázať“ (za daných podmienok), že rozdiel je signifikantný“.

Opačne to neplatí. Je to tým, že ak by sme napríklad zväčšili početnosť výberu, tak nový test by mohol ukázať, že rozdiel je signifikantný. Opačne, ak potvrdíme signifikantnosť, tak zväčšenie počtu prvkov vo výbere nevedie k opačnému výsledku. [Fischer 2x2] [4]

Vecná signifikancia rozdielu priemerných úspešností je dôležitá pri veľkých súboroch žiakov, kedy testy štatistickej významnosti rozdielov vykazujú signifikantnosť aj pri malých vecných rozdieloch výsledkov a môže mať dôležité pedagogické interpretácie.

Pre určenie vecnej signifikancie ako korelačnej miery r používame vzorec

$$r = \sqrt{\frac{t^2}{t^2 + df}} \cdot$$

t... hodnota testovacieho kritéria

df...počet stupňov voľnosti

<i>r</i>	<i>Miera signifikancie</i>
0 až 0,10	žiadna
0,11 až 0,20	veľmi mierna
0,21 až 0,30	mierna
0,31 až 0,50	stredná
0,51 až 1	silná, veľmi silná až úplná

Tab 5. Škála koeficientu vecnej signifikancie

Odhad veľkosti populačnej diferencie určitej vlastnosti dokáže bez väčších rizík interpretovať iba skúsený výskumník. Ukazovateľ vecnej signifikancie mu má byť pri tom len pomôckou.

Korelácia opisuje mieru obojstrannej závislosti dvoch náhodných veličín. V našich výpočtoch sme používali Pearsonov koeficient korelácie $r \in \langle -1; 1 \rangle$. Čím sú absolútne hodnoty odchýlky r od nuly väčšie, tým je vzťah skúmaných veličín tesnejší.

Pre naše posudzovanie miery kauzality – korelácie – prijímame členenie podľa Cohena (1988):

<i>r</i>	<i>Miera korelácie</i>
pod 0,10	zanedbateľná
0,11 – 0,30	malá
0,31 – 0,50	stredná
0,51 – 0,70	veľká
0,71 – 0,90	veľmi veľká
0,91 – 1,00	(takmer) dokonalá

Tab 6. Posudzovanie korelačného koeficientu v analýzach maturitných testov

Záver

Výsledky vyhodnocujeme v štatistickom systéme SPSS 15.0. Na spracovanie výsledkov maturitnej skúšky a položkovej analýzy testov boli použité metódy štatistickej deskripcie, inferencie a vecná signifikancia rozdielov. V deskriptívnych častiach boli použité absolútne a relatívne početnosti, priemer, štandardná odchýlka, štandardná chyba priemeru, intervaly spoľahlivosti, pedagogické ukazovatele - štandardná chyba merania, pre názornosť histogramy rozloženia početnosti úspešnosti, grafy hodnôt skúmaných parametrov položiek. Štatistická inferencia spočívala v aplikácií t-testov a ANOVE. Vecná signifikancia rozdielov bola overovaná zodpovedajúcimi korelačnými mierami. Pre výpočet reliability testov bol použitý vzorec KR-20, pretože všetky úlohy boli hodnotené binárne (0-1).

Tento dokument považujeme za otvorený pre všetky odborné pripomienky, aby splňal svoj cieľ deklarovaný v úvode.

Oddelenie hodnotenia výsledkov meraní

Bratislava, NÚCEM, marec 2009

Literatúra

1. Burjan, V.: *Tvorba a využívanie školských testov vo vzdelávacom procese*. Exam : Bratislava 1999.
2. Cohen, 1988 in <http://rimarcik.com/sk/navigator/ordinal2.html> (14.3.2008).
3. Hendl, J.: *Přehled statistických metod zpracování dát*. Portál : Praha 2004.
4. Fischer2x2
http://www.jodo.sk/07-CHI_Fischer%5Cindex_FISHER_2x2.html
5. Grošeková, M. : *Jazykové skúšky a štandardizované testy 1. časť*. In: Bulletin SAIA Slovenská akademická informačná agentúra, Informačný mesačník o štúdiu v zahraničí č. 9, ročník XIV, september 2004.
[http://www.saia.sk/images/Bulletin%20SAIA/www9\[3\].pdf](http://www.saia.sk/images/Bulletin%20SAIA/www9[3].pdf) (20.6.2006)
6. Kolektív: *Standards pro pedagogické a psychologické testování*. Testcentrum : Praha 2001.
7. Onderčová, V.: *Vlastnosti didaktických testov*, MPC Prešov.
8. Sklenárová, I. - Zelmanová, O.: *Metodika spracovania dát z maturity 2005 v systéme SPSS*. ŠPÚ: Bratislava 2005.
9. Wimmer, G.: *Štatistické metódy v pedagogickom výskume*. Gaudeamus: Hradec Králové 1993
10. URL: http://www.scio.cz/tvorba_testu/teorie_testu/index.asp (15.06.2006)