

Príručka

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

Štátny pedagogický ústav, Úsek merania výsledkov vzdelávania, štatistické oddelenie
PRÍRUČKA
vysvetlenie pojmov používaných v záverečných správach zo štatistického spracovania testov EČ MS

Zostavili: Mgr. Zuzana Juščáková, PhD.,
RNDr. Viera Ringlerová,

Bratislava, jún 2007

ŠPÚ/EČ MS

ÚVOD	4
1 ZÁKLADNÉ INFORMÁCIE O EXTERNEJ ČASTI MS	5
1.1 Cudzí jazyk	5
1.2 Postup výpočtu úspešnosti žiaka v teste EČ MS z cudzieho jazyka	6
1.3 Matematika	7
2 FÁZY SPRACOVANIA DÁT	8
2.1 Kontrola dát.....	9
2.2 Prvotné vyhodnotenie.....	10
2.3 Kvalitatívne vyhodnotenie testu	13
2.4 Druhotné vyhodnotenie.....	17
ZÁVER	20

Úvod

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

Správa zo štatistického spracovania obsahuje údaje a analýzy o danom predmete a úrovni. Informuje o štatistickom súbore – žiakoch, ktorí maturovali v danom predmete a úrovni a podrobne skúma kvalitu meracieho nástroja – testu.

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 samostatnosti porozumenia a výkladu uvádzaných štatistických výstupov.

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

1 Základné informácie o externej časti MS

1.1 Cudzí jazyk

Pre externú časť maturitnej skúšky (EČ MS) v cudzom jazyku boli pripravené testy dvoch úrovní, test vyššej úrovne A a základnej úrovne B. Cudzí jazyk a jeho úroveň si žiaci volili a do stanoveného termínu sa prihlasovali na EČ MS na pôde školy.

Maturitná skúška úrovne A 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 v úrovni A úrovni B2 Spoločného európskeho referenčného rámca Rady Európy.

Maturitná skúška úrovne B 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 v úrovne B 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 oboch úrovní 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. Úrovně testov boli vypracované s prihliadnutím na hodinovú dotáciu v danom predmete. Boli vytvorené dva varianty z každej úrovne testu, ktoré sa líšili poradím úloh, resp. pri úlohách s výberom odpovede poradím alternatív odpovede.

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

Obsahová štruktúra testov EČ z cudzích jazykov

1. tabuľka Rozdelenie položiek v teste z cudzích jazykov EČ MS

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

- 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 a doplňovanie chýbajúcich informácií.
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 a dopĺňať chýbajúce slová alebo časti 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.

1.2 Postup výpočtu úspešnosti žiaka v teste EČ MS z 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 v troch častiach testu tak, aby každá časť prispievala k výsledku žiaka rovnakou váhou. V súčasnosti sú to tri časti, ako sme uviedli vyššie, a každá z týchto častí meria inú vlastnosť (vedomosť, zručnosť) žiakovho výkonu.

Test, ktorý meria istú črtu žiaka založenú na teoretickom modeli, by mal byť podľa teórie merania konzistentným celkom, prípadne zložený z konzistentných celkov. Tým prispievame k zabezpečeniu validity meraného konštruktú. To, či vyhodnotíme jednu alebo viac charakteristík žiaka (jeden alebo viac konštruktov), určuje spôsob ďalšieho spracovania a interpretácie výsledkov.

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 vlastnosti žiakovho výkonu, potom tieto časti musíme chápať ako samostatné celky, pre ktoré je nutné zvlášť počítať výsledky a sledovať reliabilitu². 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 jednoduchým priemerom úspešností žiaka v troch častiach testu, aby sa jednotlivé časti podieľali na celkovej úspešnosti žiaka rovnakou váhou³.

Počet testových položiek 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.
- Z dosiahnutých bodových hodnôt v každej časti testu sa vypočíta úspešnosť v percentách. Úspešnosť v jednotlivých častiach testu sa počíta podľa nasledujúcich vzorcov:

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

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

$$\text{Čítanie s porozumením (č): } \quad \dot{U}_c = \frac{PSO_c}{20} \cdot 100.$$

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

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

² Rovnaké postupy používajú testovacie inštitúcie v zahraničí. Uvádžame 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.

³ Podobne je počítaný výsledok v skúške TOEFL (TEST OF ENGLISH AS A FOREIGN LANGUAGE): „Pri forme testu TOEFL papier - pero 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.)

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

$$\bar{U} = \frac{U_p + U_g + U_č}{3} \quad (1)$$

- Celkovú úspešnosť žiaka zaokrúhľujeme na jedno desatinné miesto.

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:

2. tabuľka Príklad výpočtu celkovej úspešnosti žiaka v z teste cudzích jazykov EČ MS

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%	

1.3 Matematika

Pre EČ MS z matematiky boli pripravené testy dvoch úrovní - test vyššej úrovne A a základnej úrovne B. Žiaci sa do stanoveného termínu prihlasovali na maturitu z matematiky na pôde školy voľbou úrovne A alebo B.

Maturitná skúška úrovne A je odporúčaná maturantom všetkých typov stredných škôl so študijnými odbormi, ktorí sa pripravujú na maturitnú skúšku z matematiky na vyššej úrovni.

Maturitná skúška úrovne B je odporúčaná maturantom všetkých typov stredných škôl so študijnými odbormi, ktorí sa pripravujú na maturitnú skúšku z matematiky na základnej úrovni.

V predmete matematika boli vypracované testy, 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. Úrovne testov boli pripravené s prihliadnutím na hodinovú dotáciu v predmete matematika. V každej úrovni boli vytvorené dva varianty testu, ktoré sa líšili poradím úloh, resp. pri úlohách s výberom odpovede poradím alternatív odpovede.

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

3. tabuľka Rozdelenie položiek v teste EČ MS z matematiky

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

2 Fázy spracovania dát

Odpovede všetkých testov externej časti maturitnej skúšky zapisovali žiaci do samoprepisovacích odpoveďových hárkov. Kópiu overil žiak podpisom. Oddelené kópie bez zásahu hodnotiteľa alebo inej osoby sú archivované na školách.

Z cudzieho jazyka hodnotenie otvorených úloh s krátkou odpoveďou vykonáva hodnotiteľ pod externým dozorom predsedu PMK podľa centrálne vypracovaných pokynov. Predsedovia PMK zodpovedajú za jeho správnosť.

Hodnotenie testov z matematiky je spracovávané automatizovane v software SPSS.

Originály všetkých typov odpoveďových 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 skenovania odpoved'ových hárkov,
- kontrola kódu školy,
- kontrola označenia variantu testu (kódu testu),
- 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 kľúča 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 Kódom žiaka je jeho rodné číslo. Databáza neobsahovala meno a priezvisko žiaka.

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

⁶ Hodnovernosť testu, miera vhodnosti testu pre daný účel (test meria to, čo deklaruje, že meria).

⁷ Spoľahlivosť merania. Spočíva v tom, že pri opakovanom riešení testu získame tie isté výsledky a nastane čo možno najmenší rozptyl nových hodnôt okolo pôvodných. Koefficient reliability r teoreticky nadobúda hodnoty $z \in (-1; 1)$, my posudzujeme $|r| \in (0; 1)$.

2.2 Prvotné vyhodnotenie

Kvantitatívne analýzy opisujú štatistický súbor, teda žiakov, ktorí sa do maturity v danom predmete zapojili.

Vypočítali sme 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}) je definovaná podľa štandardného vzorca (2) pre aritmetický priemer.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (2)$$

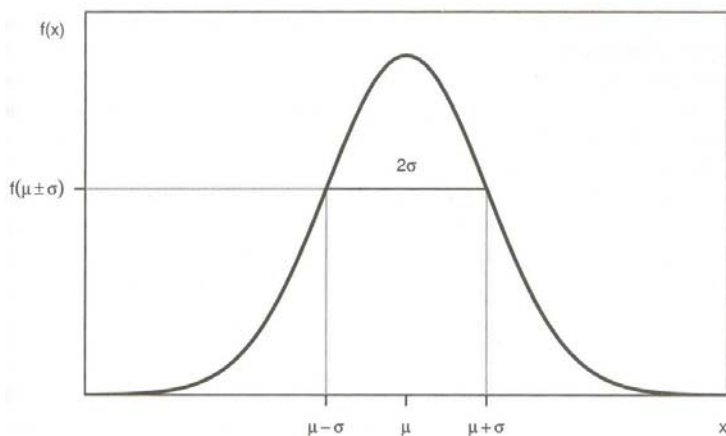
x_i ... úspešnosť i-teho žiaka

n ... počet žiakov

Štandardná odchýlka (SD) vyjadruje mieru rozptýlenia úspešnosti žiakov od strednej hodnoty (\bar{x} alebo μ). 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 počtom testovaných žiakov štandardná odchýlka spravidla klesá.

$$SD = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (3)$$

Štandardnú odchýlku 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 $\mu \pm 3,29\sigma$,

Intervalový odhad úspešnosti populácie počítame zo štandardnej odchýlky. Pri použití pravidla približne 2 štandardných odchýlok sa určuje interval (4), v rozmedzí ktorého sa umiestnilo 95% testovaných žiakov.

$$\langle -1,96 \cdot SD ; 1,96 \cdot SD \rangle \quad (4)$$

Štandardná chyba priemernej úspešnosti (SE) dokumentuje, s akou presnosťou je vypočítaná hodnota priemernej úspešnosti (5). Čím je menšia štandardná chyba priemeru, tým presnejšie charakterizuje priemer testovaných populácií.

$$SE = \frac{SD}{\sqrt{n}} \quad (5)$$

Cronbachovo alfa je koeficient reliability testu. Pri binárnom hodnotení položiek používame Kunder-Richardsonov vzorec KR-20 (6). Ak je táto podmienka oslabená, používame vzorec Cronbachovo α (7):

$$KR - 20 = \frac{k}{k-1} \cdot \frac{SD^2 - \sum_{i=1}^k p_i \cdot (1-p_i)}{SD^2} \quad (6)$$

$$Cronbachovo \alpha = \frac{k}{k-1} \cdot \frac{SD^2 - \sum_{i=1}^k SD_i^2}{SD^2} \quad (7)$$

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.
- Obtiažnosť 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 triediacu schopnosť 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) je ukazovateľom presnosti merania, do ktorého vstupujú faktory ako 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.

$$SEM = \sqrt{(SD^2 \cdot (1 - \alpha))} \quad (8)$$

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

$$\langle -1,96 \cdot SE; 1,96 \cdot SE \rangle, \quad (9)$$

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

Intervalový odhad úspešnosti individuálneho žiaka je interval (10), 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.

$$\langle x_i - 1,96 \cdot SEM; x_i + 1,96 \cdot SEM \rangle \quad (10)$$

2.3 Kvalitatívne vyhodnotenie testu

Prvým krokom je overenie ekvivalencie variantov testu. Porovnali sme súbory žiakov z hľadiska administrácie variantov podľa krajov, zriaďovateľa, typu školy a pohlavia. Potom sme zisťovali štatistickú signifikanciu rozdielu medzi variantmi a vecnú signifikanciu rozdielu, 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 testu rovnakej úrovne 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.

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.

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).

Korelácia (r) 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 od nuly väčšie, tým je vzťah skúmaných veličín tesnejší.

Medzipoložková korelácia (*P.Bis*)

Test je reliabilný, ak sú jeho položky homogénne – vnútorne konzistentné. Vnútorne homogénna – 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.

Záporná hodnota vypovedá o tom, že položka nerozlišuje dobrých a slabších žiakov. Dobrí žiaci (žiaci, ktorých úspešnosť v teste bola vyššia) na položku vo väčšine prípadov odpovedali nesprávne a naopak slabší žiaci (ktorí dosiahli v teste nižšiu úspešnosť), napríklad hádaním, uviedli správnu odpoveď. Ak je táto hodnota blízka nule, položka taktiež slabo rozlišuje dobrých 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 *P.Bis.* uvádzali pre lepšiu čitateľnosť v stonásobku. V cudzích jazykoch vypočítaný koeficient *P.Bis.* určuje medzipoložkovú koreláciu v danej časti testu (Počúvanie, Gramatika a Čítanie).

Kľúče a distraktory

Táto analýza skúma proporčnosť voľby správnej a nesprávnej odpovede v položkách s výberom odpovede medzi žiakmi rôznej výkonnosti. Pomáha charakterizovať položku z viacerých aspektov, ako sú diskriminačná sila položky, obťažnosť, formulácia problému/otázky v položke, jednoznačnosť a podobne.

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

V položkách s výberom odpovede (v testoch z cudzích jazykov sú to položky 1-13, 21- 40, 61- 73 a v teste z matematiky položky 21 – 30) sme pre každú položku vypracovali tabuľku. 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 vyznačená žltou farbou.

Kritériá hodnotenia položiek s výberom odpovede:

1. podiel žiakov, ktorí si vybrali správnu odpoveď 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 záporná.

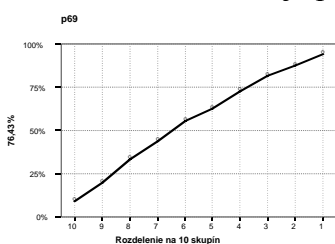
Akékoľvek nedodržanie týchto kritérií zvýrazňujeme červenou alebo hnedou farbou. Červená farba identifikuje v správnej odpovedi hodnotu *P.Bis.* menšiu ako 0,2. Hnedá farba identifikuje distraktory, ktoré majú kladnú hodnotu medzipoložkovej korelácie (*P.Bis.*).

Citlivosť alebo diskriminačná sila položky je schopnosť položky rozlíšiť dobrých a slabších žiakov. Žiakov usporiadame do poradia podľa úspešnosti v teste. Zoradených žiakov 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.

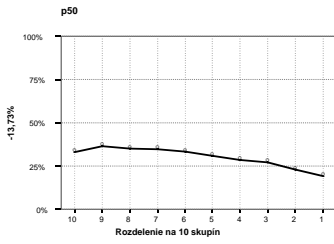
Distribúcia úspešnosti a citlivosti sa interpretuje grafmi, ktoré podrobnejšie znázorňujú vzťah medzi výkonnosťou v celom teste (delenie na 10 skupín) na x-ovej osi a ich priemernými úspešnosťami v danej položke na y-ovej osi.

Podľa výslednej čiary rýchlo odhadneme

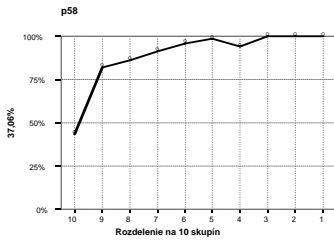
- charakter „dobrej“ položky,



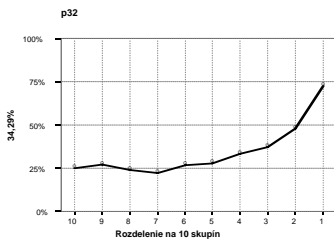
alebo úplne nevhodnej položky;



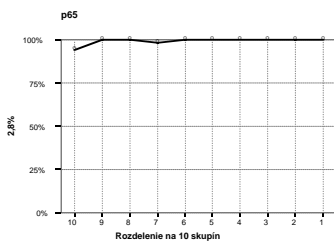
- položky, ktorá výrazne oddeľuje najslabších žiakov od ostatných,



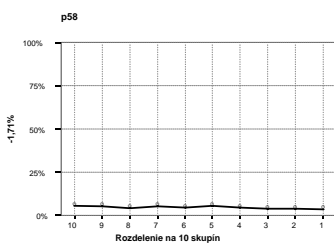
alebo naopak - najlepších od ostatných



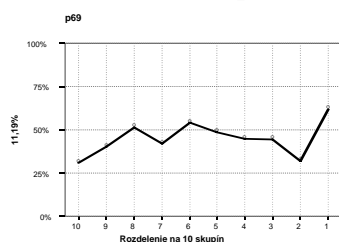
- položky, ktorá slabo diferencuje, pretože je pre všetkých veľmi ľahká



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 tvorcoch testov, pretože takýto graf



môže byť aj dôsledkom 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 pre nedostatok času neriešil. Za nedosiahnutú považujeme každú položku, po ktorej žiak žiadnu z položiek neriešil. Poslednú položku v teste posudzujeme podľa predposlednej položky v teste. Predpokladáme, že nedosiahnuteľnosť 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.

2.4 Druhotné vyhodnotenie

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

Výkonnosť žiakov v teste sme popisovali nasledovnými štatistickými charakteristikami:

Úspešnosť, ktorá bola vysvetlená vyššie.

Percentil stanovuje 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, koľko žiakov dosiahlo horší výsledok. Napríklad, ak sa žiak umiestnil v 72. percentile, 72 % 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 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: typu školy, krajov, zriaďovateľa, pohlavia a známky, ktorú žiaci získali na polročnom vysvedčení v danom predmete.

Rozdiely podľa typu školy

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

Gymnázia

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

Združené stredné školy

Stredné odborné učilištia

Rozdiely podľa kraja sú dané územným členením 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.

Výsledky, ktoré opisujú priemernú úroveň vedomostí žiakov, sú doplnené o štatistické t-testy a ANOVA a najmä o zisťovanie vecnej významnosti rozdielu.

Vecná signifikancia

Určenie vecnej, niekedy nesprávne nazývanej pedagogickej významnosti rozdielu, je dôležité pri veľkých súboroch žiakov, kedy testy štatistickej významnosti rozdielov sú signifikantné aj pri malých vecných rozdieloch výsledkov.

Pre určenie vecnej signifikancie ako korelačnej miery r používame vzorec (11).

$$r = \sqrt{\frac{t^2}{t^2 + df}} \quad (11)$$

t ...hodnota testovacieho kritéria

df ...počet stupňov voľnosti

4. tabuľka Škála koeficientu vecnej signifikancie

r	Miera signifikancie
0 až 0,1	žiadna
0,1 až 0,2	veľmi mierna
0,2 až 0,3	mierna
0,3 až 0,4	stredná
0,5 až 1	silná, až veľmi silná

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

Záver

Výsledky boli vyhodnotené v štatistickom systéme SPSS 13.0. 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. 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.

Zostavovateľky

Bratislava, ŠPÚ, jún 2007