KÉRI HENRIK

EGY ISKOLAI TELJESÍTMÉNYTESZT MATEMATIKAI-STATISZTIKAI VIZSGÁLATA

A jó teszt érdemének tartják, hogy *a tanuló tudásszintjét objektív módszerekkel méri, az eredmények számszerű összefüggésekben fejezhetők ki.* E tanulmány egy teszt eredményeinek a matematikai statisztika módszereivel való elemzésére tesz kísérletet.[1]

A felmérés

1969 májusában a Petrik Lajos Vegyipari Technikum öt párhuzamos első osztályában a laboratóriumi anyagból (a minőségi elemzésből) tesztes felmérést végeztünk. Célunk *a párhuzamos osztályok felkészültségének összehasonlítása volt.* Alaptesztként az A. C. S. Cooperative Examination Qualitative Analysis Form 1960 c. tesztjét használtuk. A teszt a felelet-választás (multiple choice) módszerrel dolgozik, egyetemi hallgatók vizsgáztatásához készült. Az eredetileg hatvantételes tesztből harminchat tételt tartottunk meg, helyenként némi átalakítással, s a tételeket Six—Tietz *Kationentrennung* című programozott füzetének négy alkalmas tételével negyvenre egészítettük ki. Az eredeti teszt utolsó tíz tételét eleve nem használhattuk: az oldhatósági szorzattal, ionkoncentrációval, disszociációs állandóval kapcsolatos számításokat tartalmaznak, ezzel elsőseink nem foglalkoztak. Az ilyen jellegű kérdéseket az eredeti teszt első ötven kérdéséből is kihagytuk. A puskázási lehetőségeket a megfelelő ülésrenddel a minimumra csökkentettük. A résztvevő tanárok egy kivétellel a tesztet nem ismerték, a felügyeletet senki sem látta el a saját osztályában. A negyven tétel megoldására nyolcvan percet adtunk. A helyesnek vélt feleletet a sokszorosított tételekkel együtt kiadott papírlap megfelelő rovatába be kellett jelölni.

Példaképpen közlöm a negyedik kérdést.

4. Ólomkloridot és higany(I)-kloridot kell egymástól elválasztani. A legjobb kémszer az elválasztáshoz

a) híg sósav *b)* híg ammónium-hidroxid
c) ón(II)-klorid *d)* forró víz
e) kálium-kromát oldat

A kérdésben számos buktató van, az egyértelműen helyes feleletet a *d)* pont tartalmazza. A helyesen válaszoló tanulók a kiadott papírlap 4. sorának *d)* oszlopát x-szel megjelölték.

[1] *Felhasznált irodalom: Ashford,* Contributions of the ACS Examinations Committee to Chemical Education, Journal of Chemical Education, 42 (1965), 496—501.; *Ebel:* Measuring Educational Achievement. New Jersey, 1965.; *Findley* (Editor): The Impact and Improvement of School Testing Programs. Chicago, 1963.; *Harris* (Editor): Encyclopedia of Educational Research. New York, 1960.; Materials used in Workshop in Kandy, Ceylon — Summer, 1968. (Procedure for the Construction of ACS Tests; A Typical Item Analysis); HENDRICKSON, JUDGE: The ACS—NSTA Chemistry Tests. The Science Teacher 35 (1968); *Hajtman Béla:* Bevezetés a matematikai statisztikába. Pszichológusok számára. Budapest, 1968.

A beadott lapokat ketten (az osztályban tanító tanár és a felügyeletet ellátó tanár) sablonnal ellenőrizték és megállapították a pontszámokat. Minden helyes feleletért egy pontot adtunk. A helytelen feleletekért a szakirodalomban ajánlott

$$R - \frac{W}{n-1}$$

képlet alapján pontokat vontunk le (a javasolt kerekítéssel). A képletben

R = a helyes feleletek száma
W = a helytelen feleletek száma
n = egy tétel variánsainak a száma (nálunk $n = 5$).

Így négy hibás felelet után egy jó pontot vontunk le. Ez a levonás a vaktában való találgatás eredményességének valószínűségét erősen csökkenti, hiszen a hibás válaszokért annyi pontot vonunk le, amennyit a vaktában való találgatással a legnagyobb valószínűséggel éppen nyerne. A pontlevonásnak nincs büntető jellege, hiszen a levonás nagy valószínűséggel csak a találgatással nyert pontok számát érinti.

A vaktában való kitöltéssel még akkor is aligha lehetne eredményt elérni, ha a levonás elmaradna. Száz ötvariánsos tételt (minden tétel egy helyes és négy helytelen feleletből, distractor-ból áll) véletlenszerűen kitöltve a helyes válaszok várható értéke húsz, a helyes válaszok száma e körül az érték körül mozog, e körül „szórnak". Ha például $\sigma = 8$ elméleti szórást tételezünk fel (s ezt több teszt átnézése alapján reális értéknek vehetjük), akkor annak valószínűsége, hogy valaki a véletlenszerűen kitöltött teszttel harminc vagy annál több pontot ér el

$$u = \frac{30-20}{8} = 1{,}25$$

A Φ (1,25) eloszlásfüggvény értéke (táblázat alapján)

$$\Phi(1{,}25) = 0.894$$

E fölé esik

$$1-0{,}896 = 0{,}106$$

A keresett valószínűség $p = 0{,}106$.
Ez százalékban kifejezve annyit jelent, hogy pusztán találgatással a felelők 89,4%-a nem éri el a harminc pontot, s csak 10,6% ér el harminc vagy annál több pontot. Negyven pontot és annál többet a feltételezett szórás mellett csak

$$u = \frac{40-20}{8} = 2{,}5$$

$$\Phi(2{,}5) = 0{,}994$$

$$1 - 0{,}994 = 0{,}006$$

$p = 0{,}006$ valószínűséggel lehetne elérni, ez a felelők 0,6%-át jelenti. A gyakorlatban a teszt nehézségétől függően ugyan jelentősen eltérhet a még elfogadható minimális pontszám, általában mégis a lehetséges pontszám 50%-ával számolhatunk, esetünkben tehát ötven ponttal. Így annak valószínűsége, hogy a feleletlapot vaktában kitöltve valaki elfogadható eredményt ér el, gyakorlatilag nulla, még akkor is, ha a hibás feleletekért nem vonnánk le pontokat.

Eredmények

A teszt alapján öt osztályt objektíven össze lehetett hasonlítani. *(1. táblázat)*. Négy osztálynál a félévi átlag (amiben az elmélet és a gyakorlati eredmények is benne vannak) és a pontszám szoros párhuzamban áll. Ez arra mutat, hogy a tanárok ezekben az osztályokban azonos vagy közel azonos mértékkel mértek. Élesen elüt ettől az 1/b osztály: a magas félévi eredmény mellett igen alacsony pontszámot ért el. A tanárt az iskolában is liberálisnak tartják, érdemjegyei a többi osztályéval egybevetve nincsenek összhangban a tanulók tényleges teljesítményével.

1. táblázat

A félévi átlag és az átlagos pontszám összehasonlítása

Osztály	A laboratóriumi átlag félévkor	A teszt pontszámainak számtani átlaga
1/a	3,4	28,1
1/b	3,7	21,5
1/c	3,2	25,4
1/d	3,0	24,6
1/e	3,3	28,0

A kapott pontokat érdemjegyekké alakítottuk át. Itt önkényesen jártunk el: a pontszámok közti határokat úgy húztuk meg, hogy az eredmények a jól értékelhető négy osztálynál hozzávetőleg a félévi átlageredménynek feleljenek meg. Az osztályok helyes és helytelen feleleteit táblázatba foglaltuk. Érdekes volt a helytelen feleletek számának eloszlását tanulmányozni. Itt már jelentős eltérések mutatkoztak az egyes osztályok között. Fény derült arra, hogy egyes kérdések milyen mértékben maradtak tisztázatlanok, melyek a leggyakoribb tévedések, félreértések, így a tanév befejezése előtt számos kérdést még egyszer megtárgyalhattunk.

Az eredmények statisztikai értékelése

A további feldolgozásban a számítások egyszerűsítésére egy osztályt, az 1/e osztályt emeltem ki *(2. táblázat)*. Az egy osztályra való korlátozást az is indokolja, hogy a teszt eredményét az osztály más (dolgozati) eredményével is összehasonlítom. Ezt a vizsgálatot azonban az eltérő feltételek miatt nem lehet az egész csoportra, hanem csak egy osztályra elvégezni.

A pontszámok érdemjeggyé való átalakításánál önkényesen jártunk el. A teszttel így elértük dolgozataink szokott átlagát, csak éppen a határok megállapításánál tologattuk ide-oda a pontokat ahelyett, hogy ezt objektív kritérium alapján

2. táblázat

A számtani átlag és a szórás számítása

Pontszám	*f*	*d*	*fd*	d^2	fd^2
40	2	12	24	144	288
39	1	11	11	121	121
38	1	10	10	100	100
37	—	—	—	—	—
36	2	8	16	64	128
35	2	7	14	49	98
34	1	6	6	36	36
33	3	5	15	25	75
32	2	4	8	16	32
31	2	3	6	9	18
30	—	—	—	—	—
29	2	1	2	1	2
A = **28**	3	0	0	0	0
27	2	—1	—2	1	2
26	2	—2	—4	4	8
25	1	—3	—3	9	9
24	3	—4	—12	16	48
23	2	—5	—10	25	50
22	2	—6	—12	36	72
21	—	—	—	—	—
20	2	—8	—16	64	128
19	2	—9	—18	81	162
18	2	—10	—20	100	200
17	—	—	—	—	—
16	—	—	—	—	—
15	1	—13	—13	169	169
összeg (Σ)	40		2		1946

Jelölések

A = kiválasztott alap, osztályköz

n = elemszám (tesztelt tanulók száma)

f = gyakoriság (a megadott pontszámot elért tanulók száma)

x_i = egyes adat, osztályközép

$d = x_i - A$ (eltérés a választott alaptól)

$\bar{x}$ = számtani átlag

s = szórás

$A = 28$

$$\bar{x} = \frac{\Sigma fd}{n} + A = \frac{2}{40} + 28$$

$$\bar{x} \cong 28$$

$$s = \sqrt{\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2} = \sqrt{\frac{1946}{40} - \left(\frac{2}{40}\right)^2}$$

$$= \sqrt{48{,}6}$$

$$s \cong 7{,}0$$

tettük volna. Ha az érdemjegyek eloszlását normál eloszlásúnak vesszük, a második táblázat segítségével könnyen jutunk célba.

Néhány megjegyzést a táblázathoz. Az első oszlopban helyet foglalnak a pontszámok. Az oszlop az elért legmagasabb pontszámmal kezdődik és az elért legalacsonyabb pontszámmal fejeződik be. A második oszlop tartalmazza a pontszámok gyakoriságát (f), vagyis a jelzett pontszámot hány tanuló érte el. A harmadik oszlop a további számítások könnyítésére szolgál. A pontszámoszlop közepe táján önkényesen kiválasztottunk egy sort, egy alapot (A), s ezt nullával jelöltük. Az első oszlopban jelzett pontszámnak ettől az alaptól való eltérését (d) az oszlop egyes soraiba beírtuk. A továbbiakban már evvel az egyszerűbb (transzformált) értékkel számoltunk. A negyedik oszlop (fd) a gyakoriság és a transzformált pontszámok szorzatát, az ötödik oszlop (d^2) a transzformált pontszámok négyzetét tartalmazza. A hatodik oszlop (fd^2) a gyakoriságok és a transzformált pontszámok négyzetének szorzatát foglalja magában. A második, negyedik és hatodik oszlopot összegeztük. $\Sigma f = n = 40$, $fd = 2$, $fd^2 = 1946$).

Ezután került sor a számtani átlag számítására az

$$\bar{x} = \frac{\Sigma fd}{n} + A$$

képlet alapján. A képlet első tagja

$$\frac{\Sigma\, fd}{n} = \frac{2}{40} = 0{,}05$$

kis értékénél fogva elhanyagolható, így

$$0{,}05 + 28 \cong 28$$

A számtani átlag ez alkalommal egyezik az önkényesen választott alappal.

Az eredmények szórását a

$$s = \sqrt{\frac{\Sigma\, fd^2}{n} - \left(\frac{\Sigma\, fd}{n}\right)^2}$$

képlet alapján számítjuk ki, értéke

$$s \cong 7{,}0$$

A számtani átlag és — normál eloszlást véve alapul — a szórás ismeretében ($s = \sigma = 7{,}0$) feladatunkat már könnyen elvégezhetjük, ha a pontszámok eltérését a számtani átlagtól ($\bar{x} = 28{,}0$)

— 1,5 σ alatt elégtelennek
— 1,5 σ és — 0,5 σ között elégségesnek
— 0,5 σ és + 0,5 σ között közepesnek
+ 0,5 σ és + 1,5 σ között jónak
+ 1,5 σ fölött jelesnek vesszük. *(3. táblázat)*

A 0,5 σ (kerekítve) négy pontnak, a 1,5. σ tíz pontnak felel meg. Így a 28 $\pm$ 4 pont (24—32 pont) alapján közepes osztályzatot adunk. Ettől alacsonyabb pontszám (egészen 28 — 10 = 18 pontig) elégségest kap, a 18-nál kevesebb pontszámért elégtelen jár. A 32 pont felett egészen 38 pontig az eredmény jó, ennél magasabb pontszám jeles osztályzatot érdemel.

3. táblázat
Érdemjegy-számítás a szórás alapján

Pontszám	Érdemjegy	Gyakoriság	Százalék
15—17	1	1	2,5
18—23	2	10	25,0
24—32	3	17	42,5
33—38	4	9	22,5
39—40	5	3	7,5

Akár a gyakorisági, akár a százalékos oszlopot nézzük, ennél az önkényesen választott osztálynál is egyenletes eloszlást látunk. Evvel elkerültük, hogy indokolatlanul magas mércét állítunk, vagy pedig az alacsony követelmények miatt a jelesek száma magasra szökik. A módszer viszont avval a hátránnyal jár, hogy az osztály átlaga felkészültségétől függetlenül mindig közepes lesz. Egy osztályon belül azonban megbízhatóan, egyenletesen oszlanak el az érdemjegyek.

Ennél tagoltabb, differenciáltabb eredményhez jutunk, ha a tanulókat *százalékosan rangsoroljuk.* Ezt a kumulált gyakorisági táblázat segítségével végezhetjük el. Ez a következő módon készül:

4. táblázat
Százalékos rangsorolás

Pontszám	Gyakoriság	Kumulált gyakoriság	Százalékos rangsor
41		80	100,0
40	2 2	78	97,5
39	1 1	75	94,0
38	1 1	73	91,2
37	— —	72	90,0
36	2 2	70	87,5
35	2 2	66	82,5
34	1 1	63	78,7
33	3 3	59	73,7
32	2 2	54	67,5
31	2 2	50	62,5
30	— —	48	60,0
29	2 2	46	57,5
28	3 3	41	51,5
27	2 2	36	45,0
26	2 2	32	40,0
25	1 1	29	36,2
24	3 3	25	31,5
23	2 2	20	25,0
22	2 2	16	20,0
21	— —	14	17,5
20	2 2	12	15,0
19	2 2	8	10,0
18	2 2	4	5,0
17	— —	2	2,5
16	— —	2	2,5
15	1 1	1	1,25
14	—	—	0,0

a) Az elért pontokat csökkenő sorrendbe állítjuk teljes számsort alkotva. A sort eggyel magasabb számmal kezdjük és eggyel alacsonyabb számmal fejezzük be, mint tényleges pontszámsorunk.

b) Tesztenként az elért pontszámnak megfelelő szám mögé egy-egy vonást húzunk. Az egyik vonás magasabban, a másik alacsonyabban helyezkedik el. Ezeket a vonásokat összegezzük (gyakoriság).

c) Minden pontszám közepéig a lefelé eső gyakoriságokat összeadjuk (kumulált gyakoriság).

d) A kumulált gyakorisági értékeket százalékban fejezzük ki.

A százalékos rangsor a tanuló teljesítményét csoportjának teljesítményéhez egyértelműen méri. Azt mutatja, hogy csoportjának hány százaléka ért el az övénél alacsonyabb teljesítményt. *(4. táblázat)*

A teszt érvényességének és megbízhatóságának vizsgálata

A továbbiakban azt vizsgálom, mennyire volt alkalmas a teszt a tanulók teljesítményének objektív mérésére. Ezért az elért eredményt az eredeti teszthez csatolt statisztikai feldolgozással vetem egybe, alkalmazva azokat a számítási módszereket, amelyeket a feldolgozásnál használtak. Ez csak bizonyos korlátok között végezhető el, de a kapott eredményekből vállalkozásunk több problematikus pontjára fény derül.

Egy teszt érvényességén értjük *azt a hatásosságot, amellyel egy mérni szánt teljesítményt valóban mérni tud.* A teszt használójának tisztában kell lenni a céllal, amelyet el akar érni. Ennek alapján választja ki vagy konstruálja meg a tesztet. A teszt alkalmazójának hozzávetőlegesen ismernie kell a tanulók felkészültségét, előképzettségét, műveltségi fokát, korát. A teszt érvényessége bizonyos körülményekhez van kötve, más körülmények között (pl. más felkészültségű, más korosztályú tanulókkal) érvényessége csökken, illetve érvényét veszti. A teszt érvényessége megközelíthető fejlődéslélektani, logikai oldalról, de a megközelítésnek statisztikai módszerei is vannak.

A teszt érvényességének vizsgálatában úgy járhatunk el, hogy más tesztekkel elért eredményekkel vagy az iskolai osztályzatokkal vetjük egybe. Vitathatatlanul egyértelmű eredményhez azonban aligha juthatunk. Az erre vonatkozó szakirodalomban bizonygatják, hogy a jó tesztek ugyanazzal a tanulócsoporttal meglepően jól egyező eredményeket hoznak. Ezt saját anyagommal — egyetlen teszt birtokában — bizonyítani nem tudom. Marad tehát, hogy a teszt eredményét más módon nyert érdemjegyekkel vetem egybe.

Az 1968/69. tanév második felében a tanulók négy dolgozatot írtak. A dolgozatok körülbelül azt a tananyagot fogják át, amelyet a teszt feldolgoz: az iskolában tanított minőségi elemzést, megtűzdelve néhány számolási feladattal. A dolgozatok általában hét kérdést tartalmaztak. A kérdések megfogalmazásánál arra törekedtem, hogy a dolgozat egyértelműen értékelhető legyen. Minden hibás feleletért egy jegyet vontam le. Négy helyes feleletért a tanuló elégséges jegyet kapott, háromért még egykettedet, kettőért már csak elégtelent. Sok esetben még további féljegyeket kellett adnom, mivel gyakran egy kérdésre adott felelet részben mégis elfogadható volt. A féljegyek az osztályozó naplóba nem kerültek, a tanulók a gyakorlati munkával, egybevetett teljes jegyet kaptak. A négy dolgozat jegyét átlagoltam és a teszt pontszámaival vetettem össze.

Az összevetés a PEARSON-féle korrelációs koefficienssel történt. *(5. táblázat)* A kapott index (r) azt mutatja, hogy a két változó egymásnak megfelelő mérték-

5. *táblázat*

Korrelációs koefficiens számítása egy teszt és négy dolgozat átlaga alapján

y \ x	0,67—1,17	1,17—1,67	1,56—2,17	2,17—2,67	2,67—3,17	3,17—3,67	3,67—4,17	4,17—4,67	4,67—5,17	*f*	*d*	*fd*	*fd*²	*xy*
41,5—38,5				1			1	1		3	+5	15	75	20
38,5—35,5				1				1	1	3	+4	12	48	24
35,5—32,5				1	1	1		2	1	6	+3	18	54	30
32,5—29,5				1	2		1			4	+2	8	16	2
29,5—26,5		2		1	1	3	1			8	+1	8	8	—2
26,5—23,5				2	2			1		5	0	0	0	0
23,5—20,5		1		2	1		1			5	—1	—5	5	3
20,5—17,5	2	2	1							5	—2	—10	20	32
17,5—14,5		1								1	—3	—3	9	9
f	2	6	1	9	7	4	4	5	2	40	.	43	235	118
d	—4	—3	—2	—1	0	+1	+2	+3	+4					
fd	—8	—18	—2	—9	0	4	8	15	8	—2				
*fd*²	32	54	4	9	0	4	16	45	32	196				

$$C_x = \frac{\Sigma fd_x}{n} = \frac{43}{40} = 1{,}07$$

$$C_y = \frac{\Sigma fd_y}{n} = \frac{-2}{40} = -0{,}05$$

$$C_x\, C_y = -0{,}05$$

$$\frac{\Sigma_{xy}}{n} = \frac{118}{40} = 2{,}95$$

$$s_x^2 = \frac{\Sigma fd_x^2}{n} = \frac{235}{40} = 5{,}87$$

$$s_y^2 = \frac{\Sigma fd_y^2}{n} = \frac{196}{40} = 4{,}90$$

$$\sigma_x = \sqrt{s_x^2} = \sqrt{5{,}87} = 2{,}42$$

$$\sigma_y = \sqrt{s_y^2} = \sqrt{4{,}90} = 2{,}42$$

$$r = \frac{\frac{\Sigma_{xy}}{n} - C_x\, C_y}{\sigma_x\, \sigma_y} = \frac{2{,}95 + 0{,}05}{2{,}21 \cdot 2{,}42} = \frac{3{,}00}{5{,}34}$$

$$r = 0{,}562$$

számai mennyiben térnek el a hozzájuk tartozó számtani átlagtól. $r = 1$ teljes egyezésre, $r = -1$ teljesen negatív, fordított viszonyra, $r = 0$ pedig minden kapcsolat hiányára mutat.

A táblázat készítésénél a következőképpen járunk el: Minden tanulóra nézve két adatunk van: pontszáma és dolgozatainak az átlaga. E két „jegy" alapján kell osztályba sorolnunk, kétdimenziós gyakorisági táblázatot kell készítenünk. Az x-tengelyre kerülnek a dolgozatátlagok növekvő sorrendben, az y-tengelyre a pontszámok csökkenő sorrendben. Az adatokat arányosan csoportosítjuk. A határokat úgy állapítjuk meg, hogy a pontszámok is, a dolgozatátlagok is a megadott határok *közé* (nem pedig a határra) essenek. Minden tanulónak a pontszámának *megfelelő sor*, illetve *dolgozatátlagának megfelelő oszlop* kereszteződésében egy vonalat húzunk. A vonalakat összegezzük, és a végső táblázatban már csak ezek a számértékek szerepelnek. A táblázat alapján pl. megállapítható, hogy három olyan tanuló volt, akinek dolgozatátlaga 3,17 és 3,67 közé esett, tesztük pontszáma pedig 29,5 és 26,5 között volt. Az f-oszlop az egyes pontszámcsoportok gyakoriságát tartalmazza. Adatai megegyeznek a 2. táblázat f-oszlopának adataival, csakhogy itt az adatokat kilenc csoportba vontuk össze. Az adatok csoportosításával járó pontatlanság elhanyagolhatóan kicsi, viszont a számítást megkönnyíti. Az f-sor az egyes dolgozatátlagok gyakoriságát tünteti fel. A további számítások megkönnyítésére ismét kiválasztunk egy tetszőleges alapot (A), amit nullával jelölünk. A csoportokat is tovább egyszerűsítjük: a csoporthatárok közé eső értéket egy egységnek vesszük, így a d-sorban illetve a d-oszlopban a nullától való eltérés egységeit pozitív és negatív irányban beírjuk. Az fd illetve fd^2 sor és oszlop kiszámítása az eddigiek alapján megy. Az f, az fd és az fd^2 sorok és oszlopok vége ismét az összegezett adatokat tartalmazza.

Az xy-oszlop adatainak kiszámításánál úgy járunk el, hogy a korrelációs táblázat minden egyes gyakorisági adatának vízszintes és függőleges d-értékét összeszorozzuk és a szorzatokat soronként összegezzük. Így jön ki az első sorra

$$-1.5 + 2.5 + 3.5 = 20$$

vagy a negyedik sorra

$$-1.2 + 0.2 + 2.2 = 2$$

A végén az oszlopot összegezzük.

A számítás további menete a táblázat mellől leolvasható.

A kapott eredmény ($r = 0{,}562$) csak közepes erősségű összefüggésre mutat. Nagyfokú egyezés esetén az x- és y-tengely alá írt jegyek a táblázat bal alsó sarkából a jobb felső sarokba húzott vonal mentén tömörülnének, itt pedig szemmel láthatóan erősen szórnak. Ezen a módon tehát nem bizonyítható, hogy tesztünk a más módon nyert információkat helyettesítheti, illetve azokat fölöslegessé teszi.

Helyes volt-e összehasonlítási eljárásunk, mennyiben megbízhatók a dolgozatok eredményei? Mint már jeleztem, a dolgozatok értékelésénél teljesen egyöntetű, egyértelmű eljárásra törekedtem. A következő táblázatban *(6. táblázat)* önkényesen két dolgozat (a harmadik és negyedik) eredményét emeltem ki és vetettem egybe. Az egybevetésnél kapott eredmény ($r = 0{,}630$) az előzőnél valamivel jobb, értékét azonban erősen lerontja, hogy teljesen azonos módszerrel mért tudásszintnél sem kaptunk ennél jobb értéket. A korrelációs koefficienssel ugyanis függetlenítettük magunkat attól, hogy az egyik dolgozat esetleg nehezebb volt: mindegyik dolgozatot a belőle is számított számtani átlaggal vetettük össze.

6. táblázat

Korrelációs koefficiens számítása két dolgozat alapján

	0,75—1,25	1,25—1,75	1,75—2,25	2,25—2,75	2,75—3,25	3,25—3,75	3,75—4,25	4,25—4,75	4,75—5,25	*f*	*d*	*fd*	*fd²*	*xy*
5,25—4,75						1	3	1	2	7	+4	28	112	72
4,75—4,25			2				1			3	+3	9	27	—6
4,25—3,75			2	1		1		1		5	+2	10	20	—2
3,75—3,25		3			1					4	+1	4	4	—9
3,25—2,75		1	2	2	1					6	0			
2,75—2,25											—1			
2,25—1,75	1	1	3				1			6	—2	—12	24	22
1,75—1,25	1				1	1				3	—3	—9	27	9
1,25—0,75	5	1								6	—4	—24	96	92
f	7	6	9	3	3	3	5	2	2	40		6	310	178
d	—4	—3	—2	—1	0	+1	+2	+3	+4					
fd	—28	—18	—18	—3	0	3	10	6	8	—40				
fd²	12	54	36	3	0	3	20	18	32	278				

$$C_x = \frac{\Sigma\, fd_x}{n} = \frac{6}{40} = 0{,}15$$

$$C_y = \frac{\Sigma fd_y}{n} = \frac{-40}{40} = -1{,}00$$

$$C_x C_y = -0{,}15$$

$$\frac{\Sigma_{xy}}{n} = \frac{178}{40} = 4{,}45$$

$$s_x^2 = \frac{\Sigma fd_x^2}{n} = \frac{310}{40} = 7{,}75$$

$$s_y^2 = \frac{\Sigma fd_y^2}{n} = \frac{278}{40} = 6{,}95$$

$$\sigma_x = \sqrt{s_x^2} = \sqrt{7{,}75} = 2{,}78$$

$$\sigma_y = \sqrt{s_y} = \sqrt{6{,}95} = 2{,}63$$

$$\sigma_x\, \sigma_y = 7{,}30$$

$$r = \frac{\frac{\Sigma xy}{n} - C_x C_y}{\sigma_x\, \sigma_y} = \frac{4{,}45 + 0{,}15}{7{,}30} = \frac{4{,}60}{7{,}30}$$

$$r = 0{,}630$$

Egy teszt megbízhatóságán értjük azt a hatásfokot, amellyel egy teljesítményt mér. Az érvényesség nemcsak a teszt tartalmára, hanem a felhasználás körére is vonatkozik. A teszt csak a megfelelő korú, megfelelő felkészültségű tanulókra alkalmazható teljes érvényességgel. Itt azonban most az a kérdés, hogy az *érvényes* teszt a teljesítményt milyen megbízhatósággal méri. A teszt megbízhatóságát némiképp az érvényességi koefficienssel is jellemezhetnénk. Mivel azonban a teszt megbízhatóságát más tényezők is befolyásolják, ezeket is figyelembe kell vennünk.

Ezek a következők:

a) A hosszabb teszt megbízhatóbb, mint a rövid teszt.

b) Heterogén tételek alapján megbízhatóbb eredményt kapunk, mint homogén tételek alapján.

c) A teszt megbízhatósága nagymértékben függ az egyes tételek megkülönböztető, diszkrimináló képességétől.

d) A közepes nehézségű kérdések megbízhatóbb eredményt adnak, mint a túl nehezek vagy a túl könnyűek.

e) Eltérő képességű tanulók megbízhatóbban tesztelhetők, mint a közel azonos képességűek

E szempontok felsorolását még folytatni lehetne. Nehéz azonban őket egy matematikai formulába foglalni. A megbízhatóságra nézve így csak becsléssel élhetünk. Gyakran alkalmazzák erre a célra a KUDER—RICHARDSON-féle 20. számú képletet. E képlet alapján számítottam ki a tesztelt 1/e osztályra a teszt megbízhatósági koefficiensét *(7. táblázat)*. A táblázat p értékét megkapjuk, ha egy tételre adott helyes feleletek számát a tételre adott összes feleletek számával elosztjuk, $q = 1 - p$.

Mint látjuk, a számítás meglehetősen hosszadalmas. Ezt a módszert akkor alkalmazzák, ha a becslésben nagy biztonságra törekszenek. Gyakran folyamodnak azonban a KUDER—RICHARDSON-féle 21. képlethez:

$$r = \frac{n}{n-1}\left(1 - \frac{n\bar{p}\bar{q}}{\sigma^2}\right)$$

ahol

$$\bar{p} = \frac{\bar{x}}{n},$$

$\bar{q} = 1 - \bar{q}$

$r =$ a megbízhatósági koefficiens

$n =$ a tételek száma

$\bar{x} =$ az elért pontszámok számtani átlaga

$\sigma =$ a szórás

Esetünkben a következő eredményhez jutunk:

$$r = \frac{40}{39}\left(1 - \frac{40 \cdot 0{,}7 \cdot 0{,}3}{48{,}6}\right) \qquad \bar{p} = \frac{28}{40} = 0{,}7$$

$$r = 0{,}846 \qquad \bar{q} = 1 - 0{,}7 = 0{,}3$$

$$\sigma^2 = 48{,}6$$

Az eredmény ($r = 0{,}846$) jóval alacsonyabb, mint a 20. számú képlettel számított ($r = 0{,}909$). Az esetek többségében ez a helyzet. Viszont nagy előny, hogy evvel a képlettel számolni rendkívül egyszerű, az alacsony eredmény meg biztosíték arra, hogy túlbecslés ne forduljon elő. További előnye még, hogy a tesz-

7. táblázat

Megbízhatósági koefficiens kiszámítása

A tétel száma	A helyes feleletek száma	p	q	pq
1	36	0,90	0,10	0,090
2	8	0,20	0,90	0,160
3	39	0,975	0,025	0,024
4	34	0,85	0,15	0,128
5	27	0,675	0,325	0,219
6	30	0,75	0,25	0,187
7	38	0,95	0,05	0,047
8	39	0,975	0,025	0,024
9	23	0,575	0,425	0,224
10	40	1,000	0,000	0,000
11	39	0,975	0,025	0,024
12	24	0,60	0,40	0,240
13	29	0,725	0,275	0,200
14	20	0,50	0,50	0,250
15	34	0,85	0,15	0,128
16	35	0,875	0,125	0,110
17	15	0,375	0,625	0,234
18	35	0,875	0,125	0,110
19	40	1,000	0,000	0,000
20	30	0,75	0,25	0,187
21	31	0,775	0,225	0,175
22	23	0,575	0,425	0,244
23	30	0,75	0,25	0,187
24	28	0,80	0,30	0,210
25	32	0,80	0,20	0,160
26	15	0,375	0,625	0,234
27	22	0,55	0,45	0,247
28	25	0,625	0,375	0,234
29	37	0,925	0,075	0,079
30	26	0,65	0,35	0,237
31	21	0,525	0,475	0,249
32	39	0,975	0,025	0,024
33	20	0,50	0,50	0,250
34	33	0,825	0,175	0,145
35	34	0,85	0,15	0,128
36	39	0,975	0,025	0,024
37	38	0,95	0,05	0,047
38	39	0,975	0,025	0,024
39	38	0,95	0,05	0,047
40	39	0,975	0,025	0,024
				5,565

$$r = \frac{n}{n-1}\left(1 - \frac{\Sigma pq}{\sigma^2}\right) = \frac{40}{39}\left(1 - \frac{5{,}565}{48{,}6}\right)$$

$$r = 0{,}909$$

tekhez mellékelt statisztikai adatok éppen az ezzel a képlettel becsült megbízhatósági koefficienst közlik, így megvan az összehasonlítási lehetőségünk. Az eredeti teszt megbízhatósági koefficiense az első húsz tétel alapján (a KUDER—RICHARDSON-féle 21. számú képlettel számolva) 0,848, a teljes, hatvantételes teszt alapján azonban 0,971.

Mi az oka annak, hogy tesztünk negyven tétel alapján ugyanazt a megbízhatósági koefficienst adja, mint az alapul vett teszt feleannyi tétel alapján? És sokkalta rosszabbat, mint az eredeti teszt hatvan kérdés alapján? A választ itt a 7. *táblázatban* megkereshetjük. Tételeink jelentős része igen kicsi megkülönböztető képességgel rendelkezett. A 10. és a 19. kérdést nyugodtan lehetett volna elhagyni, mert minden tanuló helyesen felelt meg rájuk ($p = 1,0$), e tételek segítségével a tanulók között különbséget tenni, diszkriminálni nem tudunk. A másik végletről ugyan kevesebb példa van, de a második tételre például nyolc tanuló felelt meg helyesen ($p = 0,20$), így ennek megkülönböztető képessége is igen kicsi. Így sorról sorra haladva beláthatjuk, hogy negyvenkérdéses tesztünk nem volt jobb, mint egy jól megszerkesztett húszkérdéses teszt. Ez alátámasztja azt az eredményünket, hogy ezt a tesztet erre a tanulócsoportra teljes érvényességgel alkalmazni nem lehetett.

Összefoglalás

A teszt a tanulói teljesítmény mérésének fontos eszköze. A jó teszt eredménye objektív, reprodukálható, tanárt és diákot egyaránt tájékoztat az elért eredményről, mindkét félnek segít a hiányosságok felszámolásában. A teszt érvényessége, megbízhatósága statisztikai módszerekkel jól becsülhető. A szükséges matematikai apparátus egyszerű, megnyugtatóan egyértelmű eredményekhez vezet. A jó pedagógiai felkészültség, a tanári tapasztalat önmagukban nem elégségesek ahhoz, hogy egyből megbízható teszteket készítsünk. A tesztek fordítása, átültetése problematikus vállalkozás, az átültetéssel a jó teszt érvényességéből veszít. A tesztek javításának legfontosabb útja statisztikai elemzésük. Az így feltárt hibákat kijavíthatjuk, a kevésbé megkülönböztető kérdéseket kicserélhetjük egy újabb tesztben. Egy egész országra szóló programnál elengedhetetlenül szükséges az előteszteléshez folyamodni, s csak az eredmények gondos statisztikai elemzése és a hibák kijavítása után szabad a program végrehajtásához hozzáfogni. Az előteszt elemzése azonban nem teljesen azonos az itt tárgyalt módszerrel.