

## A TESZTELMÉLETEK NÉHÁNY PROBLÉMÁJÁRÓL A PEDAGÓGIAI KUTATÁS SZEMSZÖGÉBŐL\*

OROSZ SÁNDOR

Veszprémi Egyetem Pedagógiai Pszichológiai Tanszék

*A tanulmány Horváth György könyvének ismertetéséből kiindulva a mai tesztemélet és tágabban a pszichológiai méréselmélet kérdéseit taglalja.*

Kulcsszavak: *modern tesztemélet, pszichológiai statisztika, méréselmélet*

1. Mindenekelőtt elismeréssel adózunk Horváth Györgynek évtizedes könyvterméséért; 1991: *Az értelem mérése*, 1993: *Bevezetés a teszteméletbe*, 1996: *Vélemények mérlegen*, 1997: *A modern tesztemmodellek alkalmazása* – tiszteletre méltó sor!

Természetesen a produktív periódust életének-munkásságnak egy jóval hosszabb „felhalmozási-érlelési szakasza” előzte meg. Tudjuk, hogy a szerző a nyelvtanári diploma, a pszichológiai tudományos fokozatok mellé „beszerezte” a matematikus oklevelet is: mindebből egy életre szóló, nagyon tervszerű felkészülés, kutatói „témahűség”, „problémamakacsság” habitusa bontakozik ki. Nem véletlenül lett Horváth György a teszt kutatás egyik legautentikusabb képviselője hazánkban.

S ha már ilyen summázatforma közléssel kezdtük a méltatást, érdemes előzetesen kiemelni munkáinak egynémely általános vonását is, mielőtt mondandónk lényegére térnénk.

a) Egyik jellemzője a *precizitás*. Ez megnyilvánul általában már a témamegjelölésben – téma-körülhatárolásban. A tesztemmodellekről szóló munkájának előszavában például ez olvasható: „A tesztemélet terminust a hagyományos pszichológiai értelmezésben használjuk, tehát a tesztemelés általános elméleténél szűkebb és elvontabb értelemben. Nem térünk ki mindazokra az előírásbeli, vizsgálatvezetési és tartalmi kérdésekre, amelyek ... a tesztemelés teljesebb pszichológiai elméletének

\* HORVÁTH GYÖRGY: *A modern tesztemmodellek alkalmazása* (Akadémiai Kiadó, Budapest, 1997) című könyve kapcsán

fontos területei (PETHŐ, 1974), ... adottnak tekintjük e követelmények teljesülését, bizonyítottnak vesszük a tesztelés objektivitásának feltételeit, és figyelmünket a mérőeszköz tulajdonságaira összpontosítjuk. *A pszichológiai tesztelmélet a tesztnek mint mérőeszköznek matematikailag leírható és elemezhető tulajdonságai foglalkozik.*” (11. old.; kiemelés: O. S.)

Jelentkezik ez a precizitás a pontos bibliográfiai adatolásban, a filológiai korrektségben, a következetes fogalomhasználatban is.

b) A mérésmethodikai kultúra következetes terjesztésének nevezhető a másik jellemzője. Ezen az értendő, hogy csaknem minden tárgyalt témát-problémát nem csupán „itt-és-most” állapotában mutat be, hanem időben kiterjesztve, kialakulását is felvillantva (kik, mikor, miképpen foglalkoztak vele) és a probléma fejlődéstörténetét is érintve-bemutatva; ezáltal olyan információkkal gazdagítja az olvasókat, amelyek birtokába nagyobb részük nem jutna, s amelyeket a mélyebben érdeklődők is csak nehezen, széles körű, fáradságos szakirodalom-feltárás útján lelhetnek fel.

c) A fogalmi pontosságra törekvő kifejtés (a témák természetéből adódóan) fezszes, „szikár” közlésmodot eredményez általában. Horváth tárgyyszerű stílusát átszövi azonban egy sajátos *humor*. Eszköze olykor csak a tények nem szokványos sorrendbe állítása, máskor – gyakran – nem közvetlenül, saját megfogalmazásban nyilvánul meg, hanem („nagyoktól vett”) idézetekben, mintegy ezek mögé rejtőzködve érzékelteti a maga egyéni-humoros látásmódját; ezért akár szemérmes humornak nevezhetjük. Csapán egy példát említsünk erre!

Ismeretes a képességek normális eloszlásáról kialakult felfogás körüli csaknem évszázados vita. A szerző a maga álláspontját többféleképpen is jelzi. Például modális, objektív kijelentésben, összegező-általánosító formában: „Mindenesetre tény, hogy a közhittel ellentétben a pszichológiai tesztelmélet klasszikusai kevés kivétellel nem fogadták el (vagy nem fogadták bizonyítottnak) az intelligencia normális eloszlásának feltevését.” (52. o.) – De a probléma humoros oldala iránti affinitását is érzékelteti azzal, hogy idézi a tesztelmélet egyik klasszikusának a problémával kapcsolatos eme megállapítását: „A pszichológusok azt hitték, hogy a matematikusok olyan tételt bizonyítottak, amelynek értelmében a normális eloszlásnak érvényesnek kell lennie a pszichológiai adatokra, a matematikusok pedig abban a hitben voltak, hogy ez az eloszlás olyan empirikus jelenség, amellyel a pszichológiai gyakorlatban nap mint nap találkozunk.” (52. o., Guttmantól idézi Horváth Gy.)

2. Az alábbiakban nem a szó eredeti értelme szerinti<sup>1</sup> recenziót kap az olvasó; nem következik a könyvnek még csak viszonylagosan teljes ismertetése sem, hanem egy általános bemutatás után néhány probléma pedagógiai aspektusú, meditatív jellegű taglalása következik. Egyetértésben a szerzőnek eme megállapításával: „... a tesztelmélet eredményeit ... felhasználják a pszichológián kívül más tudományok is, így mindenképp a pedagógia és a szociológia.” (11. o.) – Tegyük hozzá: nem változatlanul, hanem *adaptív módon*.

<sup>1</sup> [re]censeo = megszámlál, számba vesz, felbecsül, felsorol; recensio = számbavétel, szemle

Előre kell bocsátanunk, hogy a könyv – s vele a további fejtegetések – csak bizonyos statisztikai alapfogalmak, alapvető matematikai tudás és némi számítógépes ismeretek birtokában érthetőek.

A könyv három részre tagolódik.

Az első fejezetben („A klasszikus tesztelmélet problémái”) az általános jellemzés után néhány alapfogalom és a „hagyományos” tesztmutatók: reliabilitás, validitás, ekvivalencia-együttható, nehézségindex, elkülönítésmutató („discriminability”, „item discrimination power”, „differential validity”) kifejtése következik, majd a hagyományos tesztelméletek bírálatának ismertetésére, végül a normális eloszlás problematikájának elemzésére kerül sor.

„Az újabb tesztmodellek” című, második fejezet a „modern tesztelmélet” előzményeit és kialakulását vázolja, „természetesen” az első két Rasch-modell és Rasch logisztikus modelljének ismertetése kapcsán, majd áttekintést ad a több item-paraméteres modellekről, végül körvonalazza az e modellekkel kapcsolatos problémákat és távlati lehetőségeket.

A harmadik, „A gyakorlati alkalmazás problémái” című fejezetben először a tesztmodellek empirikus vizsgálatának két válfaját ismerteti. Az egyik a *számítógépes szimuláció*, amelyet az „in vitro” kísérletek modern válfajának tekinthetünk. A másik a gyakorlati, az eleven közegben („in vivo”) történő felhasználás, kipróbálás.

A szimuláció jelentőségét a következőkben foglalja össze: „A tesztmodellek vizsgálatában a szimulációs módszerek rendkívüli szerepét az is indokolja, hogy a gyakorlatban szinte sohasem teremthető meg az a fajta ideális kísérleti szituáció, amelyet a számítógép az elmélet követelményeihez igazodva tetszés szerint tud előállítani (pl. párhuzamos tesztek, ismétlődő mérések).” (185. o.) – De – igen dicséretre méltó módon – rögtön kitér a szimuláció korlátaira, hátrányára: „A szimuláció mégsem pótolhatja a modellek valóságos gyakorlati alkalmazásával szerzett tapasztalatokat: a szimuláció a maga *idealizált feltételeivel* (kiemelés: O. S.) nem tudja tükrözni a teszt szerkesztés és -alkalmazás *tartalmi* meghatározottságait. Az emberi pszichikum legfontosabb jellegzetessége... a tartalmi meghatározottság (amiben a természeti és társadalmi pszichikus egysége megvalósul), ezért az *emberi pszichikum vizsgálatának eszközei* is csak korlátozott mértékben, *speciális esetekben vonatkoztathatnak el a tartalmi összefüggésektől*. (186. o.; kiemelés: O. S.) – E megállapítás igazolásaként is felfogható az 1984-ben végzett beállítódásvizsgálatának az a nagyon instruktív, elemző-demonstráló bemutatása, amely e kutatás eszköztárának kialakítási folyamatát, a vizsgálat lefolyását, az eredmények elemzését ismerteti olyan részletesen, hogy egészében reprodukálható, mások által megismételhető; vagyis a tudományosság alapkövetelményeit messzemenően kielégítő és *tanulható* módon.

3. Korábban is említettük: a szerző többször jelzi, hogy a pszichológiai tesztelmélet eredményeit (egyebek között) a pedagógia is felhasználja: „Erre lehetőséget kínál a tesztelmélet viszonylagos elvontsága és általánossága.” (11. o.) – Ez az elvontság és általánosság azonban nem azt jelenti, hogy a pszichológiai tesztelmélet akár „legstabilabb” elemei is változtatás nélkül hasznosíthatók lennének más területeken. S azt se feledjük, hogy a könyvben bemutatott apparátus a pszichológián

belül is elsődlegesen a kognitív szférában „mozgó” intelligenciavizsgálatokban adekvát.

Minden diszciplína a valóság egy-egy szűkebb szféráját, speciális területét vizsgálja sajátos aspektusokból és eszközökkel. És egyazon tudomány illetékességébe tartozó valóságszférán belül is vannak olyan elkülönülő terrénumok, amelyek vizsgálata speciális eszközöket kíván. Közhely, hogy a pedagógiai kutatás szempontjából sajátos „nagy” területek a neveltség, a szociabilitás (szocializáltság) és a tudás.

Mínt hogy a pedagógiai szférában a tudás mérése mutatja a legtöbb rokonságot a pszichológiai tesztelmélet tárgyával, mert ezen a területen hasznosítottuk leginkább az intelligenciavizsgálatok tapasztalatait, illetve a tesztelmélet tételeit, ezért a továbbiakban csupán a tudásmérés területén mutatkozó néhány probléma bemutatására-elemzésére térünk ki.

A recenzeálásra való felkérést alkalomnak tekintjük sajátos problémáink némelyikének kifejtésére. Mínt hogy mind a „hagyományos”, mind a „modern” tesztelmélet (és modellek) jó néhány elemét hasznosítjuk a tudásmérésben, ezért nem e tesztelméletek mentén, hanem problémák köré csoportosítva következnek az elemzés.

4. A problémakörök: objektivitás, tartalomfüggőség, „tananyagskála”, validitás, reliabilitás, probabilitás, „discriminability”, eloszlás jellege, mérés és értékelés viszonya. Ezek sokféleképpen összefüggő kérdések, ezért nem vállalkozunk lineáris kifejtésükre, csupán csomópontok megragadására.

4.1. A pszichológiai teszt és a pedagógiai tudásteszt közötti egyik alapvető különbséget abban látjuk, hogy a pszichológiai teszt az egyén valamely tulajdonságának, mentális képességének mint állandó jellemzőnek a mérését célozza: „ahogy van” („an sich”, „sui generis”) állapotában; a tudásteszt valamely ismeret, kognitív képesség meglétét, fejlettségét, „tanítás által befolyásolt” (fejlesztett), aktuális állapotában igyekszik megragadni, hogy támpontokat találjon a továbbfejlesztéshez.

Ezen az alapon a tantárgyi tudásteszt érvényességének, *tartalmi validitásának* objektív alapja, mércéje a tantervi követelményrendszer. A mérőeszköz minősítésének kritériuma, hogy adekvát módon „lefedje” a tantervben-tankönyvben objektívált tananyagot. Ennek elsajátítását szükséges mérni, „visszajelezni” az érintetteknek: a tanulóknak, pedagógusoknak s egy sor egyéb „visszacsatolási kör” érintettjeinek.

A *validitás biztosításának első* (s talán legfontosabb) *lépése* ennek megfelelően a tananyag totális: mennyiségi, minőségi és strukturális elemzése, feltárása. Ez konkrétan a tanítandó minden ismeret: fogalom, tény, törvény, elmélet, elv, tétel, axióma, szabály, kognitív és operatív tevékenység stb. „listázását”, majd a fogalom- és ténystruktúrák, tevékenységi algoritmusok elemző feltárását jelenti.

Meglepő eredményekre vezethet az ilyen elemzés. Például: kiderült, hogy az általános iskolai anyanyelvi nyelvtan „mondatrész” tematikus egységében (szerkezetileg: öt szempont szerint, három mélységi szintbe tagolva) 30-nál több fogalmat tartalmaz a tananyag, s hogy ezek közül csupán egyetlen fogalomhoz, az állítmány

fogalmához tartozó tények száma 33. Vagy: az alsó tagozaton a „háromszögek” fogalmát két felosztási alappal hat részhalmazra (fogalomra) tagolva tanítják, e fogalmak kifejtése 35 tényt, 10 felismerési-besorolási, hat számítási algoritmust (összesen 16 féle kognitív műveletet) és 8-10 szerkesztési algoritmust (operatív feladattípusát) tartalmaz.

A tananyagfeltárást követi az *adekvát tesztfeladatok* megalkotása. A tudás sokféle alkotóelemének a kritériumokat kielégítő szinteken történő mérésére alapvetően két feladattípus: a feleletválasztó (alternatív és multiple choice) és a feleletalkotó típus sokféle válfaja s ezek különféle kombinációi szolgálnak. – Mérési gyakorlatunkban (tartalomtól függően) mintegy félszáz feladattípusát alkalmazunk rendszeresen.

Nem térhetünk ki itt e nagyon fáradtságos, sokféle hozzáértést kívánó elemző és konstruáló munka ismertetésére, csupán azt jegyezzük meg, hogy a tananyagfeltárást és feladatkonstruálást elméleti és gyakorlati szakemberek: kutatók és gyakorló pedagógusok (nem kis létszámú) munkacsoportjai végzik.

Vagyis: *a tartalmi validitás biztosítása elsősorban az előkészítő, nagyon alapos-fáradtságos team-tevékenységen múlik; alkalmazzuk természetesen a mérést követően a „hagyományos” és a „modern” tesztelméletek nyújtotta hitelesítő matematikai apparátust, és korrekciókat is végzünk a teszteken.*

4.2. A valóság bármely területén ki lehet alakítani metrikus skálát. Ennek alapfeltétele, hogy *a vizsgálandó valóságszférában találjunk olyan objektív viszonyítási alapot, amely állandó, változatlan, illetőleg nagy a konstanciája.* Ebből kiemelhető egy *egység-érték*, amelynek többszörösével/hányadosával minden mérték kifejezhető, azaz *skála építhető.* Az élettelen természettel foglalkozó tudományok számtalan ilyen alakítottak ki.

Az egyén intelligenciája, tudása szubjektív, mérhetetlen, amíg nem ölt objektív formát valamilyen: mások által is felfogható jelben-jelrendszerben, produktumban, alkotásban.

Az igazán nehéz probléma ezen objektivációkban közös viszonyítási alapot találni: melyik legyen a fentebb felsorolt ismeretfajták közül a viszonyítási alap? – A fogalom mint minden ismeret „építőköve”? – A gond az, hogy a fogalom objektívált formája vagy egy név (szó, szó szerkezet), vagy definíciót alkotó kijelentések sora, vagy másféle leírás-körülírás, azaz nem egységnyi terjedelmű jelölő.

Az ismeretfajtákból (mint egyedi tényezőkből) kiindulva nem találunk olyan viszonyítási alapot, amely *metrikai egység* lehetne. Az objektivációból: a közlés formájából közelítve viszont kiemelhető ilyen tartalmas elem: az a legkisebb, legkevésbé összetett *elemi közlés* (kijelentés), amelynek megléte-hiánya vagy/és helyes-helytelen volta egyértelműen megállapítható, *alternatív döntéssel minősíthető*, pozitív vagy negatív kategóriába sorolható, számszerű (számítógépes) összesítéshez 0 vagy 1 jellel ellátva azonosítható. Az ilyen tudáselemet nevezzük *alternatív egységnek*, illetve a „tesztnevezéktan” szerint *ítemnek*.

A „mit mérünk”, vagyis a tartalmi validitás kérdése ezzel – elvileg – megoldódik. Nagyobb volumenű mérésekben persze („megnyugtató”) a kapott adatokkal elvégezzük azokat a validitászámításokat is, amelyeket a figyelmét „a mérőeszköz tulajdonságaira” (11. o.) összpontosító tesztelmélet nyújt.

4.3. A pedagógiai tudásvizsgálatokban másképp jelentkezik a pszichológiai tesztelméletekben annyit vitatott „populációfüggőség”.

A pszichológia olyan tesztek megalkotására törekszik, amelyek az *egyén* vizsgált tulajdonságát „semleges” módon mutatják meg, függetlenül attól, hogy milyen populációban helyezkedik el. Másképp fogalmazva: a *populációfüggetlenség* mértéke a teszt egy *fő jóságmutatója*, megbízhatóságának, *reliabilitásának* alapja. (Ugyanakkor egyes pszichológiai tesztelméletek – Horváth Gy. által is kiemelt – paradoxonja, hogy a vizsgált tulajdonság populációbeli eloszlását alapul véve skálát alakítanak ki, például az intelligencia minősítésére: lényegében 7 (+4) fokú skálán helyezik el az egyént az IQ-ja alapján.

A pedagógiai kutatót (s nem az egyes tanulók tudását értékelni, minősíteni kívánó egyes gyakorló pedagógust!) pedig éppen az érdeklő főképpen, hogy *mi jellemzi a populációt*, esetünkben: milyen egy adott tanulói populáció aktuális tudása: egészében és egyenként. Az iskolarendszer működésére jellemző fontos információk nyerhetők a tudásvizsgálatok eredményeinek eloszlásából.

A „hagyományos” pszichológiai tesztelméletektől eltérően a pedagógiai tudáskutatást igazából nem érdekli, hogy a tudás eloszlása a populációban normális-e vagy sem, mert nem akarunk a (feltételezett) normális eloszlás törvénye alapján „szabályos” értékelő skálát kialakítani. Nem akarjuk (a kutatás folyamán) az itemeket súlyozni, csak számba venni: (minden) egy (jól megoldott) item = 1 (egység, vonás, rovátka, strigula, score); minden rossz vagy hiányzó item = 0. Ezeket össze-számláljuk, és arányuk adja az eredményt; egyéenként is, rétegenként is, a populációt tekintve is.

*Nem az a fő kérdésünk tehát, hogy mire használhatók az elméleti (valószínűségi) eloszlás törvényszerűségei (szabályszerűségei) a tesztkészítés folyamatában és minősítésében, hanem hogy mit mond a pedagógiai valóságról a mért értékek valós eloszlása valamely nagy validitású (és utólag számított reliabilitású) tudástesztrel végzett vizsgálatban.*

A további fejtegetéseket két *reprezentatív tudásvizsgálatunk* adataira–tapasztalataira alapozzuk. Mindkét mérés az általános iskola kibocsátó tudásszintjének megállapítását célozta, ezért a 8. osztályban, a tanév végén bonyolítottuk, egyazon (az 1978-tól bevezetett) tanterv érvényességi idején belül, lényegében azonos tesztekkel; így lehetővé vált a mintegy fél évtized alatt végbement változások feltárása is. Az első mérést az 1988/89., az 1989/90. és az 1990/91. tanév végén szerveztük, összesen 11 tantárgyból. A második mérés 1996. június első hetének három napján folyt, összesen 7 tantárgyból. A mérési tapasztalatok alapján az a (számításokkal alátámasztott) konvenció alakult ki, hogy ilyenféle vizsgálatokban megbízható eredmény „legalább négyszáz fős minta” (lásd: Horváth, 43. o.) alapján várható. Első vizsgálatunkban 400 körüli, a másodikban 1220–1240 közötti volt a minták elemszáma. (Az ingadozás az esetenként hiányzó tanulók miatt volt.)

Az eloszlás jellemzésében az elfogadott mutatókat: az átlagot (számtani közép,  $\bar{x}$ ), a szóródás terjedelmét (range), a szórásmutatót ( $\pm s$ ), a relatív szórást ( $[s/\bar{x}] \times 100$ ) és az eloszlásábrázolást fogjuk alkalmazni.

a) E mérések (s más méréseink) eredményei alapján úgy tűnik, hogy – megszorításokkal ugyan, de – létezik a valóságban normális, az elméletihez közel álló eloszlás. Nem vele egyező, hisz „... a realitások világa a korlátozások világa” (MORONEY, 1970, 103).

Az 1996. évi mérésből rendelkezésre áll 1227 tanuló nyolc adata: a hét tantárgytestt ítemenkénti jó–rossz megoldásainak százalékos arányszáma (százalékpont), valamint a 7 tárgy százalékpont-átlagaiból számított átlag. Az 1227 tanuló hét tantárgy-átlagának az átlaga 54,2 százalékpont, a szórásérték  $s = 16,2$ , a relatív szórás  $29,9 \approx 30\%$ . Az elméleti (50 százalékpont átlaghoz tartozó) szimmetrikus eloszlás mellett a szórás terjedelmébe tartozik a populáció elemeinek 68%-a; mintánkban az elemek  $66,88 \approx 67$  százaléka. Vagyis: az általános iskola kibocsátó tudásszintjének jellemzői a *közepes átlag*, a *normálishoz nagyon közeli eloszlással*, *némi jobboldali aszimmetriával*. Az erős relatív szórás viszont „gyanússá” teszi a normalitást. Szükség van részletesebb elemzésre. Ennek egyik lehetőségét az eloszlás grafikus képe (a gyakorisági poligon) adja.

Az eloszlási görbe az adatok (ábrázoláshoz történő) csoportosításától függően valamelyest módosul: több vagy kevesebb, általánosabb vagy részletesebb információkat ad ugyanazon adattömegekről. (Vagyis az eloszlás „populációfüggősége” mellett „ábrázolásfüggőség”-ről is lehetne beszélni.)

Ennek érzékeltetését szolgálja az 1. ábra, ahol egymásra másoltuk az 1227 tanuló hét tantárgyi átlagából képzett átlagteljesítmények eloszlását: az adatok 5-ös (vastag vonal) és 1-es (vékony vonal) intervallumokba sorolásával. Mindkét poligon többé-kevésbé normál ogiva alakú. Az 5-ös intervallummal készült görbe is jelez bizonyos bimodalitást (két csúcs az átlag közelében); az 1-es intervallum poligonján ez egyrészt határozottabban kivehető, másrészt határozottan jelentkezik két „árok”: a 30 és a 80 százalékpont táján, illetve ezek mellett további csúcsok rajzolódhatnak ki (polimodalitás).

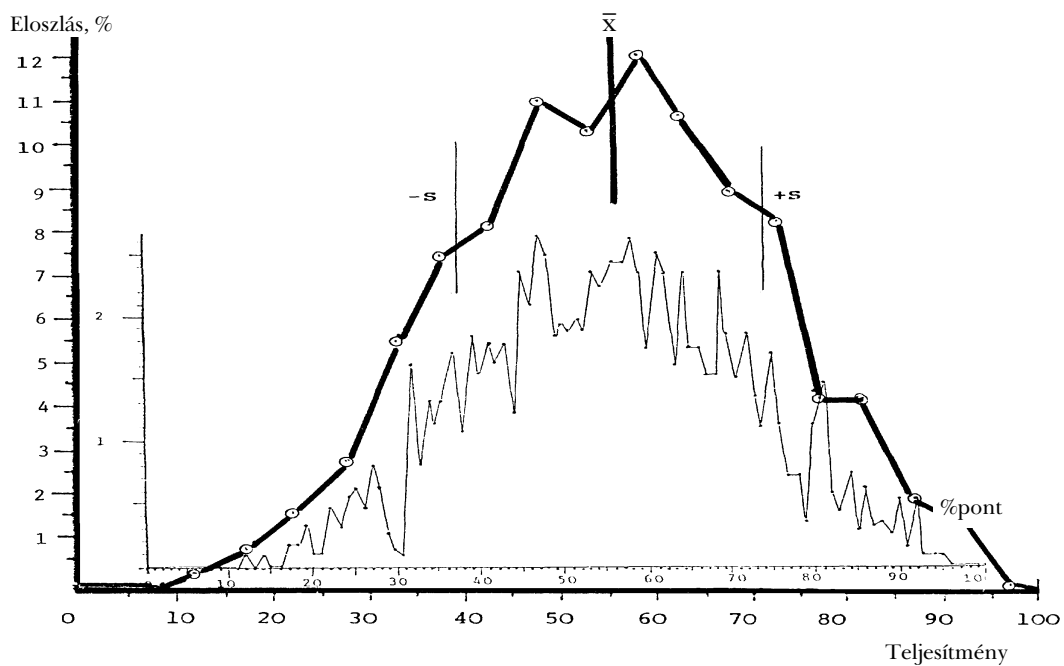
Mindez azt jelzi, hogy a minta (populáció) nem homogén: az egyedenkénti átlagok elfedik, hogy belsőleg tagolódik valamely ismérv szerint. – Ilyenfajta elkülönítést a pedagógiában gyakorta okoznak a települési viszonyok.

Válasszuk hát szét a városi és a községi iskolák tanulóit! – Az eloszlást a 2. ábra mutatja.

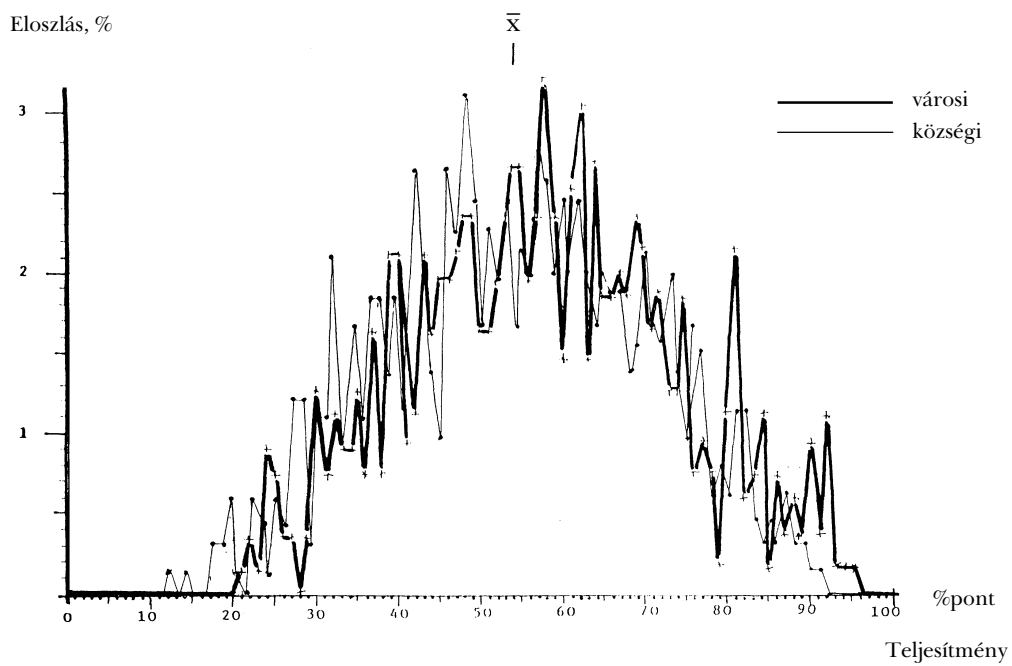
A két görbe jellege nagyon hasonló egymáshoz, és az 1. ábra jelezte polimodalitás is felfedezhető mindkét görbén. Csupán annyi különbség látszik, hogy a városi tanulók eredményeinek eloszlásgörbéje kissé jobb oldali, a községi gyerekek poligonja enyhe bal oldali aszimmetriát mutat. Valóban: a városi gyerekek átlaga ( $\bar{x} = 57,3$  százalékpont) valamivel magasabb a községekben lakókéknál ( $\bar{x} = 53,4$ ), de a 3,9 százalékpont különbség alig szignifikáns, illetőleg a szignifikáns eltérés mindössze 1,2 százalékpont, ami önmagában nem magyarázza az említett jelenséget.

Ismeretes, hogy az iskolák színvonala között jelentős különbségek vannak. (Mint ahogy országosan az átlagos osztálylétszám a mérés idején 19,1 volt, az iskolák száma körülbelül egyhuszada a mérésben részt vett tanulókénak, az iskolai átlageredmények eloszlását nem volt célszerű 5-ösnél kisebb intervallumokba sorolni.)

Iskolák és település szerinti bontásban a hét tárgy átlagainak eloszlását a 3. ábra mutatja.

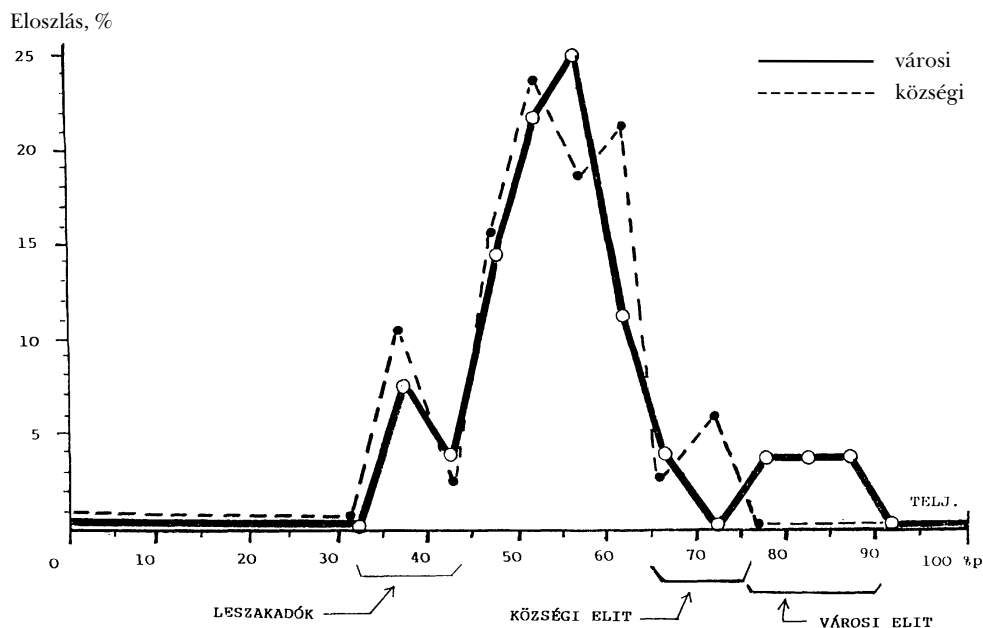


1. ábra. Hét tantárgy átlagának eloszlása, tanulónként (1996)



2. ábra. Hét tantárgy átlaga, tanulónként (1996)





3. ábra. Az iskolai eredmények polarizálódása (1996)  
(7 tantárgy iskolai átlagának eloszlása – 8. osztály)

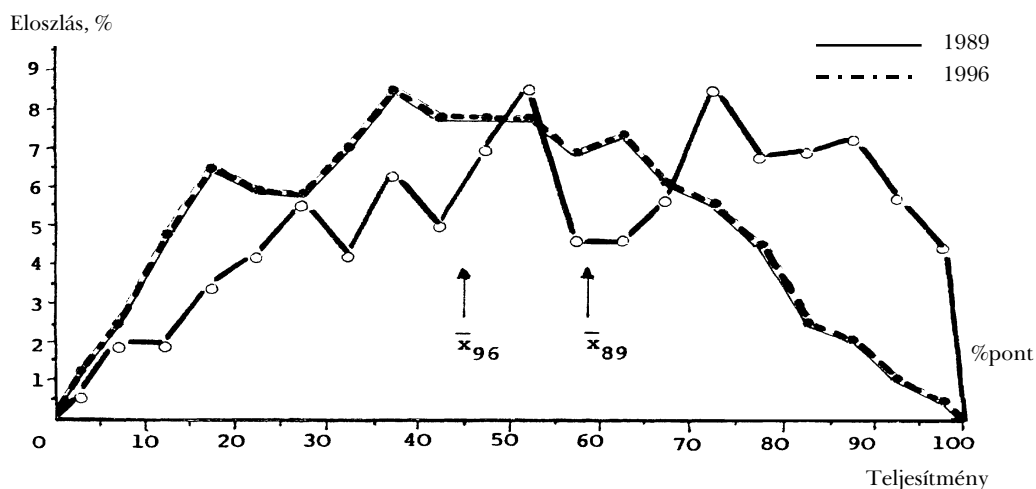
Az ábra nem szorul különösebb magyarázatra: nagyfokú polarizáltság jellemző mind a városi, mind a községi iskolák eredményeire. Ez az egyik oldalon erős *elitesedéssel*, a másik oldalon pedig a gyenge iskolák *leszakadásával* jár együtt. A leszakadók között van a városi iskolák 7-8%-a, a községiek közül körülbelül 11-12%; falun az iskolák 6-7%-a elit, a városiak 12-13%-a. A városi iskolák elitjének az átlagteljesítménye mintegy 12 százalékponttal jobb, mint a községi elitiskolák átlageredménye.

Kiegészítő adatok: az 1989–91. évi vizsgálataink idején a leggyengébb és a legjobb iskolai átlageredmény közötti különbség 38 százalékpont volt, 1996-ban viszont 52 százalékpont. Az elitesedéssel együtt járt egy igen jelentős polarizálódás: míg az egyik oldalon az iskolák több mint 10%-ában erősen növekedett az átlagos eredmény, addig a másik oldalon, az „alsó régióban” az iskolák egytizede szinte véglegesen elszakadt az iskolák többségétől. Nem nehéz e jelenséget összefüggésbe hozni a rendszerváltást követő iskolai szabadságnövekedéssel, illetőleg az átalakulás sokféle nehézségével, a kialakult szabadversennyel.

b) Az eloszlásvizsgálat olykor meglepő, de igen lényeges tényeket eredményez. Ezek egy részét a gyakorisági poligon szokásos elemzésével, más részét ettől eltérő ábrázolásmóddal, s nem kis hányadát „csupán” az adatok sajátos szempontú csoportosításával lehetséges felderíteni. – Ezekre az esetekre is hozunk példákat.

ba) Az 1989 júniusában végzett reprezentatív mérésben a matematika átlaga 58 százalékpont volt, ami a szakemberek szerint meglepően magas érték. Ehhez azonban szélsőséges, csaknem 50%-os relatív szórás tartozott. A gyakorisági poligon (4. ábra, folyamatos vastag vonal) „vészjeleket” ad: az átlagértéknél nem csúcs, hanem mély árok látható, ami arra utal, hogy az általános iskolát elhagyók a matematikatudás tekintetében két elkülönülő populációt alkotnak. A -s értéke, vagyis az „átlag alatti populáció átlaga” 30 százalékpont, s ez gyakorlatilag azt jelenti: nem rendelkeznek a továbbhaladáshoz szükséges minimális tudással sem. Ez a tömeg (több tízezer tanuló!) nagy teherterhelésnek, lehúzó erőnek, a jövőre nézve igen rossz előjelnek minősült akkor (1989) is.

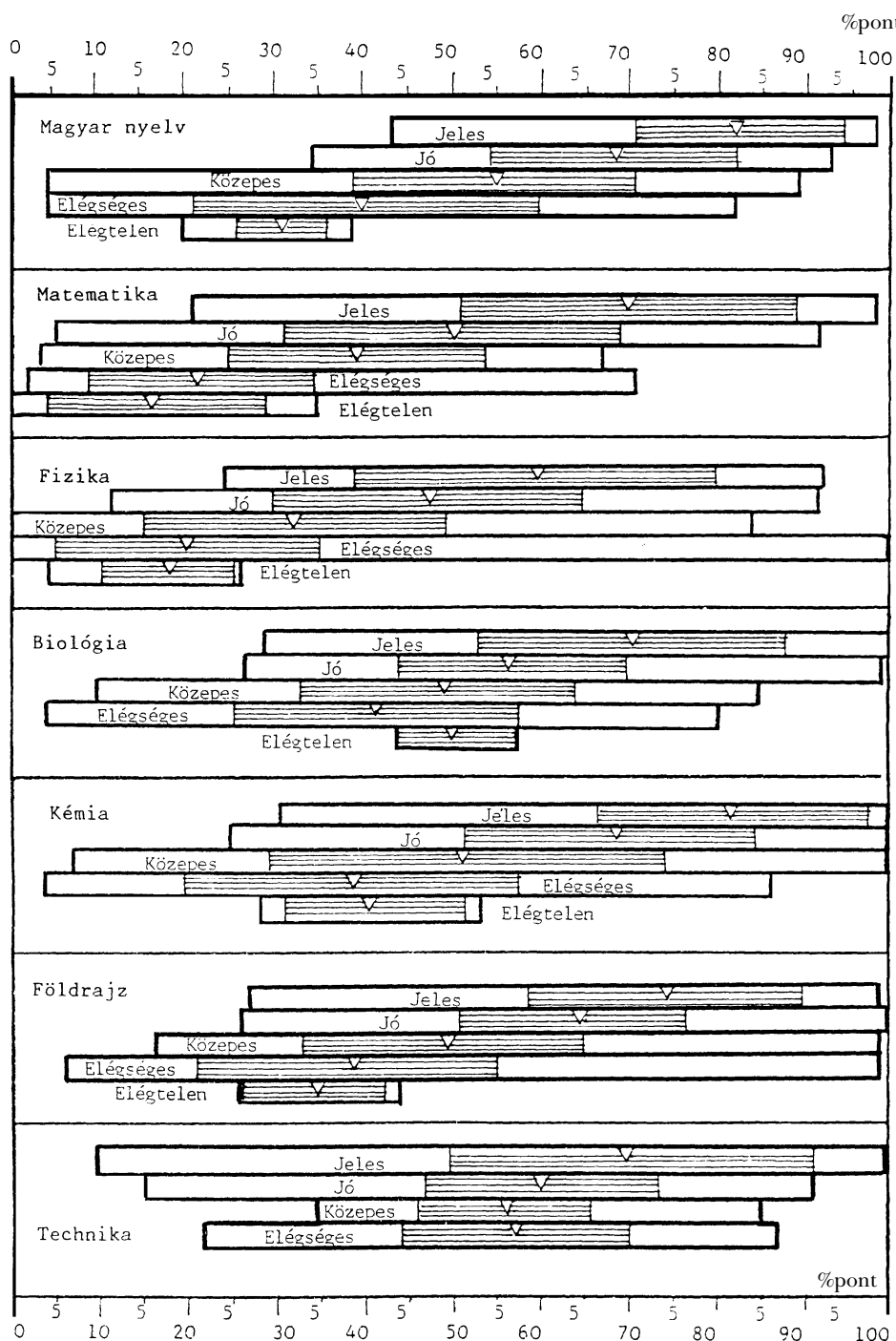
S valóban: bő fél évtized alatt, 1996-ig igen erős negatív irányú változás következett be: az átlag 12 százalékponttal csökkent ( $\bar{x} = 46$  százalékpont), ezzel együtt járt egy *negatív* irányú kiegyenlítődé (lásd 4. ábra, balra aszimmetrikus eloszlási görbe) következett be.



4. ábra. Az egyéni eredmények eloszlása matematikából

bb) Sajátos ábrázolásmódot választottunk egy akut pedagógiai probléma, az osztályozás objektivitásának vizsgálatára. – Abból indultunk ki, hogy az iskola részéről (a gyerekek továbbtanulási jelentkezése miatt) a 8. osztályos félévi tantárgyi osztályzatok hasonló szummatív szerepet töltenek be, mint a mi év végi kibocsátó tudásszintméréseink.

Tantárgyanként részmintákat képeztünk a jeles, jó stb. félévi osztályzatú gyerekek mért eredményeiből, átlagot és szórást számítottunk, és feltüntettük a szóródás terjedelmét is. – Az 5. ábrán tantárgyanként egy-egy téglalap képvisel minden egyes érdemjegyet, a téglalap hossza a mért eredmények szóródásterjedelmét képviseli, benne a vonalkázott rész a szórásterjedelmet, a csúcsával lefelé álló háromszög az átlagot mutatja. A felső és alsó határolóvonalon százalékpont-skála segíti az adatok leolvasását.



5. ábra. A mért eredmények eloszlása osztályzatonként

Nem az eredmények értékelése itt most a cél, hanem az eloszláselemzés egy sajátos, ha tetszik: „modell jellegű” változatának bemutatása. Ezért csupán egy-két szempont szerinti főbb tanulságot emelünk ki.

Látható, hogy a mért eredmények eloszlása minden tantárgyban és szinte minden érdemjegy-kategóriában aszimmetrikus, ami „belső kiegyensúlyozatlanság”-ról vall; hogy a szóródás terjedelme igen nagy, azaz nagyon eltérő eredményeket fednek az érdemjegyek minden tantárgyban, például nyelvtanból jelest kaphat olyan tanuló is, akinek a mért teljesítménye 43 százalékpont, vagy a matematikából jelesek között akad, akinek objektív (mért) eredménye alig 20%-os; hogy az érdemjegyek nem differenciálnak a tényleges teljesítmény (tudás) szerint, hiszen a 40-80 százalékpontos teljesítményzónában bármilyen osztályzat lehetséges bármely tantárgyból és így tovább.

c) Az eloszlás behatóbb vizsgálata ábrázolás nélkül is gyakran eredményezi nagyon fontos tények és összefüggések feltárását. Csupán két példát lássunk erre is az 1996. évi kibocsátó tudásszintmérésből és egy régebbi képességfejlődés-vizsgálatból!

ca) Az ítemenkénti jó–rossz megoldások aránya fontos információ az osztályt tanító tanár részére, például a kompenzálás tervezéséhez: egy 20 %pont körüli, alacsony átlag a „rátanítás” (újratanítás) szükségességére figyelmezteti. Ugyanezen ítemátlag az országos reprezentatív mintában a tanterv felülvizsgálatára hívhatja fel a figyelmet.

A tudásszint polarizációjával kapcsolatos fejtegetésből kiderült, hogy nagy jelentősége lehet az iskolánkénti (részmintákban történő) elemzésnek. Az 1996. évi matematika mérőlap egy feladata ez volt: „Egy téglatest élei:  $a = 10$  cm,  $b = 8$  cm hosszúak. Mekkora a felszíne és a térfogata? – A téglatestet az  $a$ – $b$  téglalapra mérőleges átlósíkkal elmetsszük. Mennyivel változott a keletkezett testek felszíne és térfogata az eredeti téglatestéhez képest?” – A feladat (kódolt javítókulccsal azonosított) 10 íteimből áll. Az utolsó ítem helyes megoldása: „A térfogat nem változott.”

Eltekintve attól, hogy a teljes mintában az egész feladat megoldása igen alacsony szintű, külön meghökkentő, hogy ezt a térfogatra vonatkozó, a 8. osztályban már „kötelező” evidenciaként tudnivaló összefüggést csak a tanulók 10 (tíz) százaléka adta meg helyesen. – Iskolánkénti bontásban ez az átlag úgy fest, hogy vannak iskolák, ahol a tanulók 70, 60 ... 10 százaléka adott jó megoldást, de az iskolák 79%-ában egyetlen tanuló sem jutott e felismerésre, sem számítással, sem „egyszerű” logikai úton. Ennek oka viszont már minden bizonnyal a tanítás hogyanjában keresendő.

A matematikában igen komoly problémák lehetnek (amire nemcsak az igen alacsony, 46 pontos átlag utal), mert a tudásteszt ítemeinek bő 50%-ában az iskolák fele 0 (nulla) teljesítményt produkált. Matematikatanároktól érdeklődve ennek okáról, gondolkodás nélkül mindegyik kérdezett ezt válaszolta: „Ezekben az iskolákban nem tanítják azt az anyagrészt.”

Ezek után a többi tantárgy eredményeit is megvizsgáltuk ilyen szempontból: a technikában kevésbé jellemző, de a másik hat tantárgynál átlagosan az iskolák egyötödében a tesztítemek 25%-át egyetlen tanuló sem oldotta meg: visszajelzés legfőképp a közoktatás irányítóinak, oktatáspolitikusoknak.

cb) Az alapképességek tekintetében két fő pedagógiai problémát vizsgálunk: az aktuális fejlettséget és a fejlődés jellegzetességeit; ez utóbbinak a fejlesztés-fejlesztettség tényezőinek felderítése szempontjából van jelentősége.

A helyesírás fejlettségét vizsgáltuk az 1960-as évtized végén. A mérés elméleti alapjait, metodikáját itt nem ismertetjük (lásd OROSZ, 1974). – A transzverzális mérés négy „metszetben” mutatta a fejlettséget: az általános iskola 5. és 8., a középiskola 2. és 4. osztályában. Kiszámítottuk – többek között – azt is, hogy az egyes évfolyamok tanulóinak hány százaléka érte el egy magasabb, illetve csak egy alacsonyabb évfolyam átlagát. Ezt tünteti fel az alábbi táblázat (OROSZ, 1974, 38):

1. táblázat. Magasabb, illetve alacsonyabb évfolyamok átlagát elérő tanulók százalékos aránya

		Általános iskola		Középiskola	
		5. osztály	8. osztály	2. osztály	4. osztály
Általános iskola	5. osztály	—	24	1	0
	8. osztály	32	—	17	10
Középiskola	2. osztály	3	8	—	43
	4. osztály	1	7	44	—

Az adatokból kétirányú következtetés adódik. Az egyik a „kihasználatlan lehetőségek”, hogy például az ötödikesek egynegyede elérte a nyolcadikosok szintjét, és a nyolcadikosok egynegyede a középiskolás másodikosok, illetve negyedikesek színvonalát, továbbá a középiskolás másodikosok közel fele a negyedikesekét; feltehető, hogy megfelelő eljárásokkal rövidebb idő alatt is elérhető lenne az optimális fejlettség. – A másik következtetés az „elmulasztott lehetőségek” irányába mutat, hogy ti. a nyolcadikosok egyharmada az ötödikesek átlagát, a középiskolás másodikosok közel 10%-a csak az ötödikes és nyolcadikos átlagot éri el, illetve a középiskolai negyedikesek majd felének a helyesírása a másodikosok átlagos níveljén van.

Mindkét jelenség a *közoktatás veszteséges működését* mutatja. És az ez irányú mutatók kidolgozása bizonyos rokonságot mutat az IQ-skála metodikájával.

4.4. Végül a *tesztek jóságmutatóiról* ejtünk néhány szót A teszt *bonításának, jóságának* két alapkritériuma, hogy *a)* azt mérje, amit mérni kívánunk vele; konkrétan: validitása, érvényessége legyen minél nagyobb mérvű, *b)* legyen megbízható, reliábilis, más szóval: amely eredményt egy teszttel egy adott mintában kapunk, az minél pontosabban képviselje a teljes populációban várható, valószínű (lehetséges) eredményt.

*a)* A „*hagyományos tesztelmélet*” abból a feltevésből indul ki, hogy a tesztegésszel vizsgálható tulajdonságok valamilyen szabályszerűség szerint oszlanak el az emberek tömegében, a populációban. Ezt a törvényszerűséget a matematika különféle elméleti eloszlásmodellek (normális, binomiális, Poisson-eloszlás stb.) segítségével írja le, ezért a *tesztvaliditás* annál nagyobb, minél erősebben illeszkedik a vele mért eredmények, adatok eloszlása valamely elméleti eloszlás értékeihez. (Mint már említettük, a hagyományos tesztelmélet a normális eloszlást részesíti előnyben.)

Az egyezéseket – többféle posztulátumból kiindulva – különféle számításmódokkal („modellekkel”) vizsgálják, ezek közös neve az *illeszkedésvizsgálat*, az egyes technikák, számításmódok pedig a *próbák*: t-próba,  $\chi^2$ -próba stb. (E problematika és a technikák kiváló ismertetése található Horváth Gy. könyvének első fejezetében.)

Mint korábban jeleztük, a tudásteszték validitását a tananyagelemzés–tesztkészítés fázisában igyekszünk biztosítani, beleértve e folyamatba a próbamérések eredményei alapján végzett validitásszámításokat és teszt- (item-) korrekciókat is.

*b)* A reliabilitás igen bonyolult problematikáját tulajdonképpen megkerüljük a tudásmérésekben azáltal, hogy nem hagyatkozunk a modern (pszichológiai) tesztelmélet egyes modelljeire, már csak azért sem, mert „...nincs olyan 'igazi' (a valószínűt tökéletesen tükröző) modell, hanem csak 'jól alkalmazható'... modell.” (Horváth, 104. o.)

*ba)* Ahogyan a validitás biztosítását előzetesen, a tananyagelemzés és a tesztkészítés fázisában kezdjük, úgy a teszteredmények *reliabilitását* is előre biztosítani igyekszünk.

Ez a *teszt-egészet* illetően a véletlen kiválasztással nyert, elegendően nagy elemszámú *reprezentatív minta* előállításával indul. Az eljárás az általános statisztikából is jól ismert (lásd pl. KÖVES, PÁRNECZKY, 1975) „maximum likelihood” módszeren alapul. Általa előre biztosítható az eredmények előre rögzített valószínűségi szintű reliabilitása, tényleges értéke pedig a kapott adatok alapján utólag kiszámítható. A pedagógiában – a vizsgált elemek (emberek) nagy variabilitása miatt – általában  $p = 0,05$  valószínűségi szintet veszünk alapul, ami köznapi nyelven 95%-os megbízhatóságot jelent. Ez a szint emelhető a minta elemszámának növelésével. A számításhoz általában a Student-féle t-eloszlási táblázatot használjuk; eszerint az 1000-nél nagyobb elemszámú minta végtelen nagy, vagyis az elemszám további növelése sem biztosít „abszolút” reliabilitást, más szóval: a teljes populációban végzett vizsgálatával azonos eredményt. (Csak megjegyezzük, hogy tudásmérés esetén mindig reprezentatív mintát alkalmazunk, hiszen a teljes populáció vizsgálata gyakorlatilag megvalósíthatatlan, minthogy például az általános iskolában csak egyetlen évfolyam létszáma – a jelenlegi igen alacsony lélekszámú korosztályokban is – 100 ezernél jóval nagyobb tömeget alkot.)

Eme valószínűségek azt jelentik, hogy minden mért érték kiszámítható értékhatárok között, az ún. *konfidencia-intervallumban* helyezkedik el a meghatározott szinten, az említett  $p = 0,05$  esetén a vizsgálatok 95%-ában. Utólag ellenőrizhető ennek teljesülése. Saját vizsgálatunkban a 7 tantárgy 54,1 százalékpontos átlaghoz  $\pm \Delta = 0,91$ -es konfidencia-intervallum tartozik, ami azt jelenti, hogy a nyolcadikosok populációjában az átlageredmény 53,2–55,0 százalékpont között van, az ellenőrző számítások szerint nem is csak 95%-os, hanem 98%-osnál nagyobb biztonsággal ( $p = 0,982$ ).

*bb)* Igen jelentős probléma az *itemek reliabilitása*. Ezt általában vagy párhuzamos mérések, vagy ismételt mérések eredményei közötti korreláció alapján állapítják meg, de gyakori a tesztfelezéssel előállított részminták itemei közötti *korreláció* számítása is. Tudásszintméréseinkben mindhárom eljárást alkalmazzuk.

A felezéses eljárást minden tudástesztnél használjuk, itemenként külön mintát

alkotva a páratlan és páros sorszámú elemek (tanulók vagy iskolaátlag) adataiból. Az 1996. évi vizsgálat 7 tantárgyában az item-reliabilitásokat általában az  $r=0,9000$  fölötti korreláció jellemzi.

Az itemek megbízhatóságához kapcsolódó követelmény a „discriminability”, a „*differenciáló erő*”: az a tulajdonság, hogy biztonságosan megkülönböztesse a jó és a gyenge tanulókat. A hagyományos tesztelmélet szerint ezzel a tulajdonsággal azok az itemek rendelkeznek leginkább, amelyek megoldási gyakorisága 50% körül van. Nem rendelkezik megfelelő differenciáló erővel az olyan item, amelyet a vizsgált személyek kevesebb mint 20%-a vagy több mint 80%-a old meg helyesen (lásd: nehézségi index). Ezért a 20% alatti, illetőleg a 80% fölötti megoldású itemeket kihagyják a tesztből (vagy átalakítják).

Eleven probléma, de tudástesztjeinkben ez is sajátosan jelentkezik.

S itt kell kitérnünk a *párhuzamos* (parallel) tesztekre. Tudástesztjeinkben legalább kettő, de gyakrabban négy s még több, általunk *tesztváltozat*nak nevezett párhuzamos tesztet alkalmazunk. Ezekkel kapcsolatos követelmény az *ekvivalencia*: legyenek egyenértékűek a *terjedelem* (azonos számú feladat és item változatonként), a *tartalom* és a *szerkezet szempontjából*; a *nehézségi fok azonossága* a mérés adataiból derül ki. A mintavétel pedig úgy történik, hogy a minden egyes változatot megoldó tanulók külön-külön is reprezentatív mintát alkossanak.

A terjedelmi és strukturális egyenértékűséget általában azzal biztosítjuk, hogy az egyes változatokban azonos itemszámú, azonos szerkezetű, de eltérő adatokat, tényeket tartalmazó feladatokat illesztünk be. Például egy korábbi, négy változattól álló nyelvtani mérőlapon elemezni kellett olyan ekvivalens mondatokat, amelyekben azonos mondatrészek, szó szerkezetek voltak, csak az egyes mondattani funkciókban más-más szó szerepelt (OROSZ, 1973). Az elemezendő minőségjelzős szerkezetek és megoldási arányuk a következő volt:

2. táblázat. Tartalmi momentumok hatása a tevékenység-algoritmus működésére

Teszt-változat	Jelzős szerkezet	Jó megoldás, %
A	tiszta forrásból	68
B	száraz gallyat	64
C	derék úttörők	44
D	hétfői napon	24
$\bar{x}_{A-D}$		50

Feltűnő a C változatban a „derék úttörők” s még inkább a D változatban a „hétfői napon” szó szerkezetben a minőségjelző (derék, hétfői) alacsony szintű megoldása. Utóbbinak – mondhatnók – nincs kellő differenciáló ereje. – Hagyjuk ki a tesztből? – semmiképpen se, hiszen a tanítás szempontjából lényeges információval szolgál mindkettő. Azzal ugyanis, hogy egy algoritmus, esetünkben a minőségjelző besoroló algoritmus nem „csak úgy általában” működik, hanem tartalomtól függetlenül: a 'derék' manapság ritkán használt, ezért a gyerekek jelentős része számára

nem világos jelentésű szó, ezért mondatrész-besorolása is mérsékelten sikerült; a 'hétfői' képzett melléknévben pedig a szó főnév szófaj jellege maradt domináns, s ez a gyerekek nagy hányada számára „kizárta” a jelző funkciót.

S folytathatnánk a példák sorolását.

5. Horváth György pszichológiai tesztmodellekről írt könyve egy sor tudásmérési jellegzetesség átgondolását involválta. Egyértelmű, hogy a tesztek minősítésére alkalmas matematikai apparátus ('modell') jelentős részét adaptálhatjuk – adaptáltuk.

Egyik fő különbség a pszichológiai tesztek és a pedagógiai tudásmérő tesztek között az utóbbiak jóval erősebb (domináns) tartalomfüggőségében lelhető fel. No meg abban, hogy a tudásméréssel egyidejűleg más dimenziókban is végzünk tényfeltárást, például a gyerekek körülményeire, személyiségére, a pedagógiai közegre és a társadalmi háttérre vonatkozóan, és ezeknek az eredményekkel való összefüggéseit vizsgálva, a tudásszint alakulására vonatkozó magyarázatokat, okokat keresünk – találunk.

Úgy tűnik, ezen munka könnyítése érdekében célszerű (s az összegyűlt tapasztalatok alapján lehetséges is) lenne a pedagógiai mérésekkel kapcsolatos „*elemzési modellek*” kidolgozása.

## IRODALOM

- GUTTMAN, L. (1971) Measurement as structural theory. *Psychometrika*, 1971 Dez., 346–356.
- KÖVES P., PÁRNICZKY G. (1975) *Általános statisztika*. Közgazdasági és Jogi Könyvkiadó, Budapest
- MORONEY, M. J. (1970) *Számoktól a tényekig*, Tankönyvkiadó, Budapest
- OROSZ S. (1973) *Standardizált témazáró tesztek. Magyar nyelvtan. Általános iskola 5. osztály*. József Attila Tudományegyetem, Szeged
- OROSZ S. (1974) *A helyesírás fejlődése*. Tankönyvkiadó, Budapest
- PETHÓ B. (1974) *Technikai javaslatok pszichológiai tesztek alkotására és a vizsgálatok lefolytatására*. Akadémiai Kiadó, Budapest

## SOME PROBLEMS OF TEST THEORIES – FROM THE POINT OF VIEW OF EDUCATIONAL INVESTIGATION

OROSZ, SÁNDOR

*The paper starting from the review of a Hungarian book on test theory discusses basic issues in present day psychological testing and measurement theory.*

Key words: *modern testtheory, psychological statistics, measurement theory*