

X 75515

BEVEZETÉS A PEDAGÓGIAI KUTATÁS MÓDSZEREIBE

SZERKESZTETTE:
FALUS IVÁN



MŰSZAKI KÖNYVKIADÓ, BUDAPEST

X 75515

Az egyes fejezetek szerzői:

Falus Iván	1., 2., 6., 16.
Tóthné Környei Márta	3.
Bábosik István	4., 8.
Szabolcs Éva	5., 11./b
Nádasi Mária	7., 11./a
Réthy Endréné	9.
Csapó Benő	10.
Nahalka István	12., 13., 14., 15.

Az V. rész szerkesztésében közreműködött:

Kotschy Beáta

Bírálta:

Ballér Endre

Vágó Irén

© Falus Iván, Tóthné Környei Márta, Bábosik István, Szabolcs Éva, Nádasi Mária,
Réthy Endréné, Csapó Benő, Nahalka István, 2004

© Műszaki Könyvkiadó, 2004

SZTE Egyetemi Könyvtár



J000665943



ISBN 963 16 2664 4

Azonosító szám: MK-3020501

X 755 15

CSAPÓ BENŐ:

TUDÁSSZINTMÉRŐ TESZTEK¹

A FEJEZETBEN KIFEJTÉSRE KERÜLŐ KÉRDÉSEK:

- A tudásszintmérés sajátosságai
- A klasszikus tesztelmélet alapjai
- Valószínűségi tesztelméletek
- A tesztek jóságmutatói
- Az itemek jellemzői
- Normaorientált és kritériumorientált tesztelés
- A tananyag elemzése
- Feladattípusok
- Az értékelési rendszer kialakítása
- Ekvivalens tesztváltozatok készítése
- A tesztek kipróbálása és standardizálása
- A feladatok, itemek jóságának vizsgálata
- A reliabilitásmutatók kiszámítása

1. A TESZTELÉS ELMÉLETE

1.1. TUDÁSSZINTMÉRÉS, TUDÁSSZINTMÉRŐ TESZTEK

A pedagógiai vizsgálatok egyik leggyakrabban alkalmazott formája, sajátos területe a tudásszintmérés. A tudásszintmérés elmélete a tesztelés pszichológiai elméleteiből fejlődött ki, de a mérések sajátos tárgya újszerű, eredeti gyakorlati megoldások kidolgozását tette szükségessé, ami hamarosan kihatott az elmélet fejlődésére is.

A közös eredetnek és a párhuzamos fejlődésnek köszönhetően nehéz lenne olyan definíciót alkotni, ami fogalmi szinten precízen elkülöníti a pszichológiai és a tudásszintmérő tesztekét. Találunk azonban számos olyan gyakorlati szempontot, amely alapján a tudásszintmérő tesztekét megkülönböztethetjük a pszichológiai tesztekétől. A pszichológiai tesztek a személyiség legkülönbözőbb vo-

¹ A fejezet a *Módszertani füzetek pedagógiai vezetőknek* c. sorozatban megjelent *Pedagógiai értékelés* c. füzet (Veszprém, 1987/88) rövidített változata alapján készült.

násainak vizsgálatára szolgálhatnak. Használatuk, eredményeik értelmezése sajátos felkészültséget igényel, az általuk nyert adatok bizonyos esetben az orvosi titokhoz hasonló védelmet élveznek. A gyakorlatban ennek megfelelően pszichológiai tesztnak kell tekintenünk mindazokat a teszteket, amelyeket kidolgozói annak minősítenek. A pedagógiai, közelebbről tudásszintmérő tesztek ezzel szemben az iskolai munka rutinszerűen használt munkaeszközei lehetnek, melyeket a megfelelő képzettségű tanárok széles körben, gyakorlatilag korlátozás nélkül használhatnak. Vannak olyan tesztek, amelyek a pszichológiai és pedagógiai tesztek között átmenetet képeznek. Ilyenek például az eredetileg pszichológiai vizsgálatokra készített, azonban az iskolai felmérésekre is alkalmazhatóak tett tesztek, amelyek használata egyszerű, eredményei közérthetőek, és a tanulók jobb megismerésének hatékony eszközei. Ilyen lehet például egy figyelem-teszt és bizonyos képesség-tesztek. (Kozéki, 1984, ismertet több ilyen tesztet.)

A tudásszintmérő tesztek is pszichológiai tulajdonságokat mérnek, azok egy sajátos formáját, a tudást. Mégpedig általában az iskolában elsajátított tudást, vagy kicsit általánosabban: azt a tudást, amelyik az iskolai tanulásban szerepet játszik.

A tudásszintmérő teszteket is további csoportokba lehet sorolni. Az egyik gyakran alkalmazott felosztás megkülönbözteti a *standardizált* és a *tanárok által készített* teszteket. A standardizált tesztek általában hivatásos tesztkészítők, specialisták által kidolgozott mérőeszközök. Hosszabb fejlesztési folyamat, többszöri kipróbálás eredményei, a jóságukra vonatkozó mennyiségi mutatókat ismerjük. Ezekről megkülönböztetjük a tanárok által készített (teacher-made) teszteket. Ezekre az jellemző, hogy a tanárok maguk készítik általában saját vagy szűk körű használatra, a minőségükre vonatkozó számszerű mutatók nem ismertek.

Egy másik felosztás megkülönbözteti a tulajdonság- (adottság, aptitude) teszteket a teljesítmény- (achievement) tesztektől. A tulajdonságtesztek általában a „jövőbe” tekintenek, előrejelző, prediktív funkciójuk van. Azt vizsgálják, mi várható az adott tulajdonságokkal rendelkező tanulóktól a jövőben. (Ezek a tesztek állnak közelebb a pszichológiai tesztekhez.) A teljesítménytesztek a múltban elsajátított tudást mérik fel.

Az előzőekben említett felosztások inkább csak a lehetséges pólusokat jelzik, azok között mindenféle határeset is előfordulhat, és a tesztek néha sem formailag, sem elméletileg nem különböztethetők meg. A magyar pedagógiai szaknyelvben előfordul még a „feladatlap”, „mérőlap” kifejezés is. Ezeket ugyancsak nem lehet szigorúan definiálni, illetve megkülönböztetni. Annyiban tekinthetők teszteknek, amennyiben rendelkeznek a tesztektől megkövetelt tulajdonságokkal.

1.2. A KLASSZIKUS TESZTELMÉLET ALAPJAI

A tesztelméletek matematikai, elsősorban valószínűségi és statisztikai összefüggések rendszerei. A tesztelméletek gyökerei e század első évtizedeire nyúlnak vissza, teljesen kifejlett formájukban pedig az 50-es évek végén jelentek meg.

A tesztelméletek alapfeladatuknak tekintették, hogy a tesztekkel kapcsolatos matematikai illetve statisztikai törvényszerűségeket egységes rendszerbe foglalják. Néhány alapvető kiinduló feltevést rögzítenek, majd ebből a szokásos matematikai eljárásokkal tételeket vezetnek le, amely tételeket azután a tesztekkel kapcsolatos gyakorlati kérdések megválaszolásában is használhatunk.

Ma már klasszikus tesztelméletként szokás említeni – az újabb tesztelméletektől való megkülönböztetésként – azt a tesztelméletet, amelynek alapfogalmait itt felvázoljuk. Ez az elmélet is néhány alapfogalomra és néhány alapfeltevésre épül, magát a tesztelméletet a belőlük levezetett összefüggések rendszere alkotja. (A klasszikus tesztelmélet matematikai igényességű kifejtését adja Lord és Novick, 1966; az alkalmazási lehetőségekkel együtt ismerteti Lienert, 1967 és Horváth, 1991.)

Az alapfogalmak közül természetesen először a teszt fogalmát kell tisztáznunk. A teszt ebben az értelemben pszichológiai (pedagógiai) mérőeszköz jelent, amely a pszichikus tulajdonságokat megfelelő skálán méri. A teszt kisebb, önállóan értékelhető egységekből áll. Ez lehet egy szubteszt (rész-teszt), amely a teszt kisebb, önálló feladatokból álló részét jelenti. (Nem feltétlenül szükséges, hogy a tesztet szubtesztekre bontsuk, másrészt a szubteszteket tekinthetjük önálló teszteknek is.) A tesztek feladatokból állnak, a feladatok legkisebb, önállóan értékelhető részeit itemeknek nevezzük. A pedagógiai, ezen belül a tudásszintmérő tesztekben hasznos ennek a négy szintnek a megkülönböztetése. A legtöbb pszichológiai teszt azonban nem bomlik fel ilyen sok szintre, a klasszikus tesztelmélet feltevése szerint elegendő csak a tesztek és az azokat alkotó legkisebb önálló egységeket, az itemeket értelmezni.

Az itemek, mivel a legkisebb önállóan értékelhető egységek, már nem bonthatók fel további részteljesítményekre, éppen ezért az itemeket már csak kétféleképpen minősíthetjük: a tanuló megoldotta, az item megoldása jó, a tanuló nem oldotta meg, az item megoldása nem jó. A klasszikus tesztelmélet szerint 1 pontot ér az az item, amelyet a tanuló jól oldott meg, 0 pontot természetesen az, melyet a tanuló nem jól oldott meg. (Tudásszintmérő tesztek esetében az itemeket súlyozhatjuk is, azaz nemcsak 1 pontos, hanem az item súlyától függően többpontos itemeket is definiálhatunk, l. később.) A tesztek összpontszámát, vagyis a teszt eredményét az itemekre kapott pontok összege adja. Ha tehát 0-val és 1-gyel pontozunk, a teszt összpontszáma a jól megoldott itemek számával lesz egyenlő.

A tesztelmélet célja az, hogy a tesztekkel kapcsolatban felmerülő kérdéseket matematikai úton fogalmazza meg, és ezáltal lehetővé tegye, hogy kérdéseinkre egzakt választ kapjunk. Az egyik legtermészetesebben felmerülő gyakorlati kérdés az, hogy tesztünk mennyire jól méri azt, aminek a mérésére kidolgoztuk. Példaként nézzük meg, hogy erre a kérdésre milyen módon keresi a választ a klasszikus tesztelmélet. Fel kell tételeznünk, hogy annak a tulajdonságnak, amelyet mérni szeretnénk, minden személyre jellemzően van egy meghatározott valódi értéke. Tehát minden vizsgálandó személyhez hozzátartozik például valamilyen intelligenciaszint, a kreativitásnak valamilyen mértéke, vagy témánk fogalmait használva egy adott tárgy valamilyen tudásszintje. Ezt nevezzük az adott tulajdonság, – például a tudásszint – valódi értékének, amit V betűvel

jelölünk. Fel kell azonban azt is tételeznünk, hogy amikor ezt a tulajdonságot mi a tesztünkkel meg akarjuk mérni, akkor a tesztel való mérés során, – mint ahogy mindenféle mérés során – hibát követünk el. Tesztünk tehát nem pontosan tükrözi ennek a tulajdonságnak a valódi értékét. Ezt a klasszikus tesztelmélet úgy veszi figyelembe, hogy egy bizonyos mértékű hibát definiál. Ezt a hibát H betűvel jelöljük. Ily módon, amikor egy tulajdonságot megmérünk, az általunk mért értékben a valódi érték, és a hiba játszik szerepet. Ez a hiba a valódi értéket felfele vagy lefele eltérítheti, felfele például úgy, hogy a tanuló véletlenszerűen ráhibáz a jó megoldásra, lefele úgy, hogy figyelmetlen, nem a tudása legjavát nyújtja. Ezt az alapvető feltételezést a klasszikus tesztelmélet alapegyenletében foglalhatjuk össze, következőképpen:

$$M = V + H$$

A képletben M a mért értéket, a V a tulajdonság valódi értékét, H pedig a hibát jelöli. Ehhez az egyenlethez a klasszikus tesztelmélet bizonyos feltételezéseket kapcsol. Olyan feltevéseket, amelyeket igaznak fogadunk el, és amely feltevések igazsága esetén az összes azokból levezetett összefüggések is mind igazak. Itt nem soroljuk fel tételelesen a klasszikus tesztelmélet axiómáit, csupán utalunk arra, hogy ezek milyen jellegűek. Feltesszük például azt, hogy a hiba várható értéke nulla, vagyis a valódi értéktől ugyanolyan valószínűséggel térnek el a tanulók vagy a megvizsgált személyek negatív, mint pozitív irányban, tehát összességében a sokféle hiba statisztikusan kiegyenlíti egymást. Feltesszük továbbá azt is, hogy a méréskor jelentkező hiba nem függ attól, hogy milyen az adott tulajdonság valódi értéke. Statisztikailag ezt úgy fogalmazhatjuk meg, hogy a valódi érték és a hiba közötti korreláció 0 . Az ehhez hasonló további alapfeltevések rendszeréből a matematikai levezetések szabályait figyelembe véve a tesztelmélet tételeit vezethetjük le. Visszatérve kiinduló kérdésünkhöz – tehát hogy mennyire jól méri a teszt azt, amit vele mérni kívánunk – e kérdést a klasszikus tesztelméletet felhasználva matematikai formában fogalmazhatjuk meg. Azt mondhatjuk, hogy tesztünk annál jobban méri a mérni kívánt tulajdonságot, minél jobban összefügg a valódi érték a mért értékkel. Statisztikai kifejezéssel: minél jobban korrelál M és V .

A tesztnek ezt a tulajdonságát, az így értelmezett jóságát a teszt megbízhatóságának, idegen szóval reliabilitásának nevezzük. A reliabilitás tehát a klasszikus tesztelmélet definíciója értelmében a mért érték és a valódi érték közötti összefüggés. Pontos matematikai meghatározása szerint a reliabilitási koefficiens, amit r_{tt} -vel jelölünk, a mért és a valódi érték közötti korrelációs együttható négyzete:

$$r_{tt} = r^2$$

Szokás még az úgynevezett reliabilitási indexről beszélni, ami nem más, mint az előbbieken definiált korrelációs együttható.

Kérdésünket ezzel matematikai pontossággal fogalmaztuk meg, gyakorlati problémánk azonban ezzel még nincs megoldva. Nem tudjuk ugyanis meghatározni a korrelációs együtthatóban szereplő egyik mennyiséget, vagyis az adott tulajdonság valódi értékét. Bármilyen tesztet használunk, minden tesztben a

valódi érték mellett valamilyen hibát is mérünk, tehát mindig csak bizonyos mért értékekkel számolhatunk. Ezen a ponton derül ki, hogy mi az értelme az egész formalizált tesztelméletnek, a matematikai apparátus felhasználásának. Lehetőségünk van ugyanis arra, hogy ismert matematikai, statisztikai tételek felhasználásával olyan formulákat dolgozzunk ki – a már előbb említett axiómákra alapozva –, amelyekben már csupa mért értékek szerepelnek, és amely formulákkal becslést adhatunk a reliabilitásra.

Mivel ez az egyik legfontosabb gyakorlati probléma (azaz, hogy tesztjeink megbízhatóságát jellemezzük), a klasszikus tesztelmélet keretében számos reliabilitási mutatókat dolgoztak ki. Ezek a reliabilitási mutatók olyan formulák, amely formulákba csupa mért értéket kell behelyettesíteni, és segítségükkel a reliabilitási értéket lehet kiszámítani. Pontosabban a reliabilitásra lehet egy alsó becslést adni. Olyan becslést, amelynél a reliabilitás valódi értéke mindig csak nagyobb lehet. Ezeket a becsléseket a megfelelő statisztikai törvényszerűségekre alapozhatjuk. A reliabilitás fogalmával, gyakorlati értelmezésével és a formulák felhasználásával a következő fejezetben még részletesebben foglalkozunk. Itt csupán a klasszikus tesztelmélet matematikai apparátusának működését illusztráljuk. További problémákat a klasszikus tesztelmélet hasonló formulákkal, levezetésekkel old meg.

A 60-as évek végére, mikorra a klasszikus tesztelmélet végleges formáját elérte, az is kiderült, hogy a gyakorlatban felmerülő számos problémára nem tud választ adni, illetve, hogy azok a feltételek, amelyeket az axiómákra rögzített, a pszichológiai tesztek használata során többnyire nem teljesíthetők, vagy nem szigorúan teljesíthetők.

Még több probléma jelentkezett a pedagógiai tesztek esetében, különösen a tudásszintméréssel kapcsolatban. Itt az a helyzet, hogy azok az értékek, amelyeket a pszichológiai tesztek értéknek tekintenek, nem feltétlenül értékek a pedagógiai, a tudásszintmérő tesztek használatánál. Csak egy példát említve: a pszichológiai tesztek akkor működnek jól, ha azokon a megvizsgált személyek közel normális eloszlású teljesítményt produkálnak. A tesztek különböző technikai megoldásokkal, úgy készítik el, hogy a végső eredményeik már normális eloszlást adjanak. Ezzel szemben egy pedagógiai tesztnél (pl. egy tudásszintmérő tesztél) nem feltétlenül kell elérnünk azt, hogy normális eloszlásúak legyenek az eredmények, hiszen elképzelhető az is, hogy a tanulók nagy többsége, vagy szinte mindenki tökéletesen megoldja a tesztet. Ha a tanítás során jól dolgoztunk, akkor éppen erre törekszünk. Szélsőséges esetben, ha mindenki 100%-os, tökéletes megoldást nyújt, akkor a tesztnek nem is lesz szórása, tehát a klasszikus tesztelmélet összes formulája használhatatlanná válik. Ezek a problémák vezettek el oda, hogy különböző, más megfontolásokon alapuló tesztelméleteket dolgozzanak ki. Az egyik fejlődési irány, amely a probablisztikus (valószínűségelméleti) tesztelméletek kidolgozásához vezetett, elsősorban a felmerült tesztelméleti, matematikai problémákat kívánta megoldani. A másik irány, a kritériumorientált tesztelmélet a tudásszintmérés gyakorlati követelményeinek jobban megfelelő tesztek kidolgozásának megalapozásához vezetett.

1.3. VALÓSZÍNŰSÉGI TESZTELMÉLETEK

A tesztemlétekek újabb generációját képezik a modern, ún. probabulisztikus (valószínűségi) tesztemlétekek. Ezek a tesztemlétekek a teszt használata során elkövethető hibát más módon veszik számba, mint tette azt a klasszikus tesztemléte. Mivel ezeknek a tesztemlétekeknek meglehetősen bonyolult a matematikai apparátusuk, csupán az alapgondolatukat mutatjuk be. (Részletesebb leírásokat illetően I. Rasch, 1960; Rost, 1988; Horváth, 1985, 1991.)

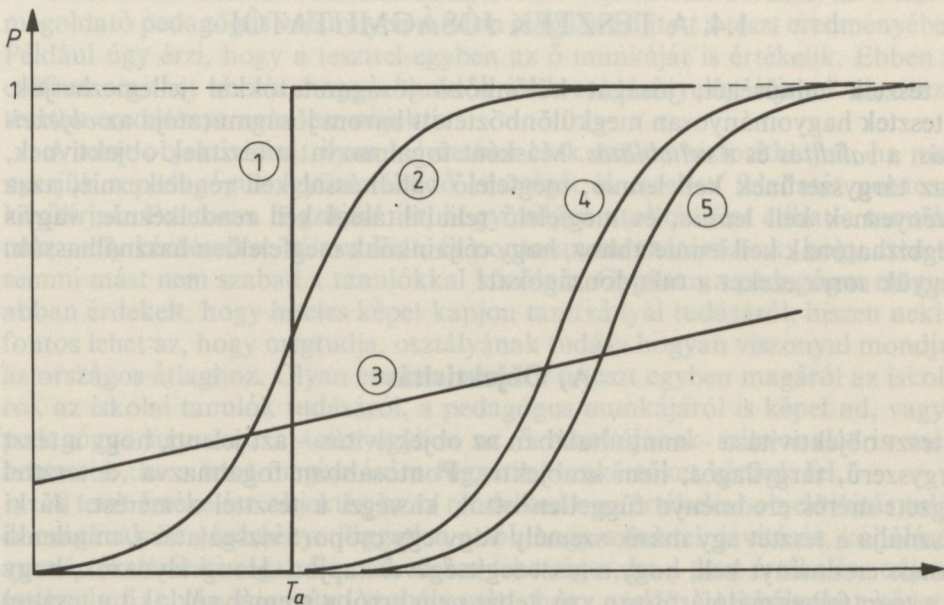
Az elmélet a tesztemlétekek tulajdonságait valószínűségelméleti eszközökkel jellemzi. Ebben a tesztemléteben az, hogy egy tanuló megold egy tesztemléteket, vagy nem old meg, nem determinisztikus, hanem valószínűségi természetű. Ez azt jelenti, hogy ha különböző tudásszintű tanulókat veszünk, akkor annál nagyobb annak a valószínűsége, hogy az adott tesztet a tanuló jól fogja megoldani, minél nagyobb a tanuló tudásszintje. Az összefüggés azonban nem determinisztikus, hanem valószínűségi, azaz a gyengébb tudású tanuló kisebb valószínűséggel, a jobb tudású tanuló nagyobb valószínűséggel oldja meg az adott tesztet. Így az item tulajdonságait jól kifejezi az itemhez rendelhető valószínűségeloszlás. A valószínűségeloszlás és az itemek tulajdonságai közötti összefüggéseket a 10.1 ábrán szemléltetjük.

Az ábrán öt különböző item valószínűségi függvényét ábrázoltuk. Az x tengelyen a tanuló tudását tüntettük fel, az y tengelyen pedig annak valószínűségét, hogy a tanuló az itemet jól oldja meg. A valószínűségeloszlást ebben az esetben jellegzetes S alakú görbék tükrözik. A görbék alakja jól kifejezi azt, hogy milyen tulajdonságú itemmel állunk szemben.

Az 1. számmal jelölt görbe például egy olyan itemnek a karakterisztikáját jelöli, amelyet senki nem tud megoldani, aki nem érte el legalább a T_a tudásszintet, de mindenki meg tud oldani, aki ennél a T_a -nál nagyobb tudásszinttel rendelkezik. Ez tehát egy olyan item, ami nagyon pontosan két csoportra osztja a tanulókat aszerint, hogy a tanuló elérte-e az adott itemmel kapcsolatos tulajdonság egy bizonyos szintjét, vagy nem érte el. Természetesen ez a szögletes görbe csak egy határeset az S alakú görbéknek, amit a valóságban meg lehet közelíteni, de elérni soha nem lehet. Az ilyen típusú görbékkel jellemezhető itemek nagy differenciáló erővel rendelkeznek, azaz nagyon érzékenyen különbséget tesznek a tanuló között. Az itemek többsége azonban nem rendelkezhet ilyen szélsőségesen magas differenciáló erővel.

Nézzük például a 2. számmal jelölt item görbét. Ezt átlagosan ugyancsak a T_a tudásszinttel rendelkező tanuló oldja meg, azonban bizonyos, kis valószínűséggel megoldják azok a tanuló is, akik ezt a tudásszintet nem érték el. Majd a tanuló tudásszintjének növekedésével fokozatosan növekszik annak valószínűsége, hogy a tanuló megoldja az itemet, de a T_a fölötti tudásszinttel rendelkező tanulóknak sem mindegyike oldja meg az itemet hibátlanul, hanem csak egyre nagyobb valószínűséggel. Azonban még az igen magas tudásszinttel rendelkező tanulókról sem mondhatjuk biztosan, hogy megoldják az itemet, hanem csak azt, hogy ezek esetében a megoldás már nagyon valószínű. Ez az item láthatóan kisebb differenciáló erővel rendelkezik, mint az 1. item.

Különböző paraméterekkel rendelkező itemek karakterisztikus görbéi



10.1. ábra

Ha megnézzük a 3. számú item görbét, azt látjuk, hogy ezt az itemet csaknem ugyanolyan valószínűséggel oldják meg a gyenge tanulók, mint a magas tudásszinttel rendelkező tanulók. Ez egy lapos karakterisztikus görbe, ami azt jelenti, hogy ez az item nem túlságosan differenciál a tanulók között. Ha tehát olyan tesztet készítünk, amelynek az a lényege, hogy a tanulókat tudásuk szerint jól megkülönböztesse egymástól, akkor az ilyen jellegű itemeket ki kell hagyni a tesztből. Természetesen más a helyzet, ha a célunk alapvetően nem az, hogy a tanulók között igen érzékenyen különbséget tegyünk. A 4. és az 5. itemek összehasonlítása két párhuzamosan futó karakterisztikus görbét mutat. Az 5. számú görbe olyan, mintha a 4-et a tanulók tudásszintjét jellemző tengely mentén a pozitív irányba eltoltuk volna. Ez azt jelenti, hogy a két item nehézségében van csak különbség, többi jellemzőjük egyébként megegyezik. A probabilisztikus, vagyis valószínűségi tesztelméletek tehát az itemek karakterisztikus görbéinek a jellemzőit vizsgálják, annak alapján jellemzik az egyes itemek tulajdonságait, és ezeken keresztül a tesztek jóságát.

A görbéket valószínűségi függvényekkel adják meg, ezek a függvények két vagy három paramétert tartalmaznak. A pedagógiai tesztelés egyik legfonto-

sabb és leggyakrabban használt valószínűségelméleti modellje a kétparaméteres *Rasch modell* (Rasch, 1960), illetve annak különböző továbbfejlesztett változatai.

1.4. A TESZTEK JÓSÁGMUTATÓI

A tesztek minőségét, jóságát különböző jóságmutatókkal jellemezhetjük. A tesztek hagyományosan megkülönböztetett három jóságmutatója az *objektivitás*, a *validitás* és a *reliabilitás*. Másként fogalmazva: a tesztnek objektívnek, azaz tárgyyszerűnek kell lennie; megfelelő validitással kell rendelkeznie, azaz érvényesnek kell lennie; és megfelelő reliabilitással kell rendelkeznie, vagyis megbízhatónak kell lennie ahhoz, hogy céljainknak megfelelően használhassuk. Vegyük sorra ezeket a tulajdonságokat!

A) Objektivitás

A teszt objektivitása – mint általában az objektivitás – azt jelenti, hogy a teszt tárgyyszerű, tárgyilagos, nem szubjektív. Pontosabban fogalmazva: a teszttel végzett mérés eredménye független attól, ki végzi a teszttel a mérést. Bárki használja a tesztet egy másik személy vagy egy csoport vizsgálatára, mindenki azonos eredményt kell, hogy a teszt segítségével kapjon. Hangsúlyozzuk, hogy itt a teszt felhasználójáról van szó, tehát nem arról a személyről, akit a teszttel mérünk, akit tesztelünk. Másként fogalmazva tehát ez azt is jelenti, hogy az elért eredményeket, a tesztpontokat csakis a vizsgált személy tulajdonságai határozzák meg, és függetlenek attól, hogy a tesztet ki használja, ki értékeli, ki végzi el a vele kapcsolatos számításokat.

Az objektivitásnak további formáit megkülönböztetjük meg, attól függően, hogy a tesztelés különböző fázisaiban milyen személyek kerülnek a teszttel kapcsolatba. Ennek megfelelően az objektivitás három alapformája (alesete) az *adatfelvételi objektivitás*, az *értékelési objektivitás* és az *interpretációs objektivitás*.

Az előző definíció értelmében az adatfelvételi objektivitás azt jelenti, hogy a teszteredmény független az adatfelvevő személyétől. Adatfelvételen azt a folyamatot értjük, amikor a teszttel az adatokat összegyűjtjük, tehát amikor a tesztet a vizsgált tanulók megoldják. Egyszerű hétköznapi példára lefordítva: bármelyik pedagógus oldatja meg a tanulókkal a tesztet, ugyanazok a tanulók minden pedagógus felügyelete, irányítása alatt ugyanolyan eredményt érnek el. A teszt eredménye tehát nem függhet attól, hogy ki az a pedagógus, aki a teszt megoldásában közreműködik. Első pillantásra úgy tűnik, hogy ezt a feltételt nagyon egyszerű teljesíteni, hiszen nem kell mást tenni, mint egyértelműen előírni a tesztet megoldató pedagógus feladatát. Ha az utasításoknak pontosan eleget tesz, akkor a teszt megoldása máris független attól, hogy ki az, aki ezeket az utasításokat pontosan követi. Tudjuk azonban, hogy a gyakorlatban néha nehéz ezt megvalósítani. Egy teszt megoldása során a tanulók a legkülönfélébb

kérdésekkel fordulhatnak a pedagógushoz, és így a különböző pedagógusok beállítódásuktól, feladatuk értelmezésétől függően más és más válaszokat adhatnak a tanulóknak. Egy-egy mondat, egy figyelmeztetés, egy orientáló megjegyzés egy egész osztály teljesítményét tolhatja el valamilyen irányba.

Különösen kiélezetté válik az adatfelvételi objektivitás kérdése, ha a tesztet megadató pedagógus valamilyen módon maga is érintett a teszt eredményében. Például úgy érzi, hogy a teszttel egyben az ő munkáját is értékeli. Ebben az esetben segíthet abban, hogy a tanulók jobb eredményeket érjenek el, azaz a tesztek eredménye nem lesz objektív.

A teszt objektivitását ilyen helyzetben csak azáltal biztosíthatjuk, ha megnyerjük a pedagógus együttműködő készségét és emellett feladatát pontosan körülírjuk. Pontosán rögzítjük azt, hogy milyen utasításokat adhat a tanulóknak, mely kérdéseikre válaszolhat, és pontosan meghatározzuk, hogy ezen túl semmi mást nem szabad a tanulókkal közölnie. Gyakran a pedagógus maga is abban érdekelt, hogy hiteles képet kapjon tanítványai tudásáról, hiszen neki is fontos lehet az, hogy megtudja, osztályának tudása hogyan viszonyul mondjuk az országos átlaghoz. Olyan esetben, amikor a teszt egyben magáról az iskoláról, az iskolai tanulók tudásáról, a pedagógus munkájáról is képet ad, vagy a pedagógus úgy érzi, a teszt egyben az ő munkájának színvonalát is méri, helyesebb, ha a teszt megíratására független szakértőket kérünk fel.

A teszt értékelése objektív, azaz rendelkezik az értékelési objektivitás tulajdonságával, ha eredménye független attól, hogy a tesztek javítását, kódolását, azaz értékelését ki végzi. Ez csak akkor lehetséges, ha a teszt értékelése teljesen egyértelmű, olyan szigorúan szabályozott, hogy ezeket a szabályokat követve bárki ugyanarra az eredményre jut. A tesztek értékelési objektivitását különböző eljárásokkal fokozhatjuk. A legjobb értékelési objektivitást akkor kapjuk, ha feleletválasztós tesztet alkalmazunk, hiszen ekkor a tanuló a feleletek közül egyszerűen csak választ, a javítókulcsok alapján pedig bárki ellenőrizni tudja, hogy a tanuló a megfelelő, azaz a helyes választ jelölte-e meg. Ilyenkor az értékelés objektivitását csak az értékelő figyelmetlensége csorbíthatja. Sajnos éppen ezzel a feladattípussal kapcsolatban (mint később látni fogjuk) más problémák merülhetnek fel, ezért biztosítanunk kell a feleletalkotó kérdéstípusok értékelési objektivitását is. A feleletalkotó kérdéseknél a javítókulcs precíz megadásával, a helyes válasz pontos, egyértelmű rögzítésével biztosíthatjuk az értékelés objektivitását.

A teszt értékelési objektivitását viszonylag könnyen ellenőrizhetjük, azaz tapasztalati úton is megvizsgálhatjuk, mennyire objektív az általunk alkalmazott teszt és a hozzá tartozó értékelési utasítás (azaz javítókulcs). Nem kell mást tennünk, mint ugyanazokat a tesztek különböző pedagógusokkal értékeltetni. A pedagógusok az értékelést a tesztől független külön lapon végzik el, az eredményeket azután számítógépbe visszük. Ugyanazokat a tesztek két, esetleg három vagy négy különböző személy is értékeli. Ha a teszt értékelése objektív, akkor minden egyes pedagógusnak, minden egyes értékelőnek azonos eredményre kell jutnia. Ha olyan kérdést találunk, amelynél az értékelők bizonytalankodnak, azaz nem mindegyikük azonosan minősíti a megoldást, alapos gyanúnk lehet arra, hogy a javítási utasítással valami probléma van, nem

egyértelműen írja körül a helyes megoldást, tehát a feladat javítókulcsán (vagy a feladaton is) változtatnunk kell.

Az objektivitás harmadik formája az interpretációs objektivitás. Az interpretáció azt jelenti, hogy a tesztek eredményeit hogyan hasznosítjuk, abból milyen következtetéseket vonhatunk le, hogyan értelmezzük a teszt eredményét. Az interpretáció kérdése többnyire csak a bonyolultabb személyiségtesztek, vagy a tudományos kutatásban használatos tesztek esetében merül fel, így ezekkel a kérdésekkel itt részletesebben nem foglalkozunk.

B) Reliabilitás

A teszt egyik legfontosabb tulajdonsága a megbízhatóság, idegen szóval reliabilitás. A teszt megbízhatóságának mérésére, számszerű jellemzésére az ún. *reliabilitásmutatók* szolgálnak. Elsőként a reliabilitás számításának egyik alapfogalmát mutatjuk be.

Képzeljük el, hogy van egy tesztünk, amelyik csupa olyan feladatból áll, amely feladatok mindegyike ugyanazt a tulajdonságot méri. Például készítünk egy tesztet, amelynek mindegyik feladata azt vizsgálja, hogy a tanulók hogyan tudnak törtet törttel osztani. Ha a feladatok nagyon hasonlítanak egymásra, azaz lényegében csak a számértékek különbözőek, akkor azt várhatjuk, hogy aki az egyik feladatot megoldja, az nagy valószínűséggel megoldja a másik feladatot is, aki pedig egyik feladatot elrontja, az elrontja a másikat is. Természetesen vannak határesetek vagy középső esetek, amikor a tanuló a feladatok egy részét meg tudja oldani, más részét nem. A tipikus azonban mégiscsak az, hogy vagy tud törtet törttel osztani, vagy nem, és ha tudja, akkor a feladatok többségét megoldja, ha nem tudja, akkor a feladatok nagy részét nem oldja meg. Ha tehát a tesztünk jól mér, akkor azt várhatjuk, hogy az egyes feladatok megoldása között magas korrelációt kapunk. Ha azonban a teszt nem megbízható, nem jól mér, például véletlenszerűen, találgatással is helyes választ lehet az egyes kérdésekre adni, akkor az egyes feladatok eredményei között, az egyes feladatok megoldásának jósága között semmiféle összefüggés nem lesz. Ez a gondolatmenet felhasználható a reliabilitás vizsgálatának egyik nagyon egyszerű módjára.

Vegyünk egy tesztet, amely viszonylag sok (10–15) itemből áll. Osszuk két részre a kérdéseket, sorszámozzuk be, és vegyük a páros és páratlan itemeket külön-külön. Képezzünk így egy képzeletbeli féltesztet a páros, és egy másik képzeletbeli féltesztet a páratlan itemekből. Adjuk most össze külön a páros és külön a páratlan számú itemek eredményét, és számítsuk ki a korrelációt az így kapott két félteszt között. Töltsünk ki képzeletben egy tesztet véletlenszerűen, például kockadobással. Ekkor a két félteszt eredménye között semmiféle összefüggés nem lesz: semmi sem befolyásolja azokat olyan irányba, hogy ha az egyik félteszt az átlagosnál jobb, akkor a másik félteszt is az legyen. A két félteszt eredményeit ugyanis csak a tanuló tudása kapcsolhatja össze. Azt várhatjuk, hogy ha a teszt jól méri azt, aminek a mérésére használni akarjuk, akkor azok, akik a teszt egyik felét jól oldották meg, jól fogják megoldani a másik felét is,

akik az egyik felét rosszul oldották meg, rosszul oldják meg a másik felét is. Minél szorosabb a két felteszt között a kapcsolat, annál inkább a tanuló tudása határozza meg a teszteredményt, és nem valami más külsődleges tényező, például találgatás.

A reliabilitás vizsgálatára más alapelveket, például a varianciák összehasonlítását is felhasználhatjuk. A gyakorlatban a reliabilitás kiszámításának konkrét módját általában attól tesszük függővé, hogy milyen adatok állnak rendelkezésünkre. A reliabilitás a teszt alapvető jellemzője. A tesztet bárminek a mérésére csak akkor használhatjuk, ha magas reliabilitási mutatókkal rendelkezik.

C) Validitás

Validitáson a teszt érvényességét értjük: azt a tulajdonságát, hogy a teszttel valóban azt mérjük-e, aminek a mérésére a tesztet kidolgoztuk, amit a teszttel mérni kívánunk. A kérdés megint túlságosan is egyszerűnek tűnik, hiszen ha egy tesztet elkészítettünk valaminek a mérésére, szinte természetesnek tűnik, hogy a teszt valóban azt is méri. Ez azonban egyáltalán nem biztos. A problémát megint egy egyszerű példával szemléltethetjük.

Tételezzük fel, hogy egy tesztet készítünk, amivel az általános iskolai tanulók matematika tudásának egy speciális részét, például a szöveges feladatok megoldásának készségét kívánjuk vizsgálni. Elvégeztük a felméréseket, és az eredmények értékelésénél tartunk. Tegyük fel továbbá, hogy rendelkezésünkre állnak a tanulók korábbi tesztjei, így például az írásbeli szorzás és osztás végzésével kapcsolatos tesztek, és ugyancsak rendelkezésünkre állnak a tanulók olvasásmegértési tesztjeinek eredményei. Az értékelés során furcsa dolgot tapasztalunk: azt látjuk, hogy azok a gyerekek, akik a matematikában jók, rendszeresen rossz megoldást adnak a teszten, míg mások, akik korábban nem jeleskedtek, egészen jól megoldják a feladatokat. A furcsa jelenség feltárása érdekében számítógépbe tápláljuk a korábbi matematika teszteredményeket, az olvasás teszt eredményeit és a szóban forgó szöveges feladat megoldásával kapcsolatos teszt eredményeit is. Kiszámítjuk a tesztek eredményei közötti korrelációt, és azt tapasztaljuk, hogy a szöveges feladat tesztünk eredményei sokkal szorosabban korrelálnak az olvasás teszt eredményeivel, mint a korábbi matematika eredményekkel. A jelenség oka tehát rögtön egyértelművé válik, kissé pontosabb elemzéssel rá is jövünk, hogy a szöveges feladat teszt nem azt méri, aminek a mérésére kidolgoztuk: a szövege ugyanis túlságosan bonyolult. A tanulók egy része el sem jut a feladat megoldásáig, vagy legalábbis nem annak a feladatnak a megoldásába fog bele, amit mi kitűztünk, mert nem érti meg a feladat szövegét. Olvasási készsége nem elég fejlett ahhoz, hogy a feladat bonyolult szövegét pontosan értelmezze. Hiába tudná a szöveg mögött rejlő, a számára esetleg egyszerű matematikai feladatot megoldani, ha a bonyolult szöveg megakadályozza abban, hogy ezt a feladatot megértse, és a matematikai megoldást megadja.

Tesztünk tehát valóban mér valamit, mégpedig azt, hogy a tanulók mennyire jól értik meg a feladat szövegét, azaz milyen az olvasási készségük. A teszt

eredményeiben e készség fejlettsége tükröződik, nem pedig az, aminek a mérésére kidolgoztuk, azaz a szöveges feladatok megoldásának matematikai készségei. Ebben az esetben azt mondjuk, hogy a teszt validitása nem megfelelő, nem azt méri, aminek a mérésére készítettük. Emellett a teszt rendelkezhet még magas reliabilitási mutatókkal, azaz a teszt valamit jól mérhet, csak nem azt, amit mi vele mérni gondoltunk.

Érdemes utalni a feladat reliabilitásának és validitásának összefüggéseire. Ahhoz, hogy a tesztnek a validitásáról egyáltalán beszélhessünk, a tesztnek megbízhatónak kell lennie, tehát jó reliabilitási mutatóval kell rendelkeznie. Az összefüggés fordítva nem áll fenn, ha egy tesztnek rossz a validitása, attól még a reliabilitása lehet igen magas. Láttuk, hogy egy teszt validitása még viszonylag egyszerű esetekben is kétséges lehet. Szinte természetes, hogy olyan bonyolult pszichikus tulajdonságoknál, mint például az intelligencia, a kreativitás, a problémamegoldó gondolkodás, a megfelelő tesztek validitása még sokkal nehezebb kérdéseket vet fel.

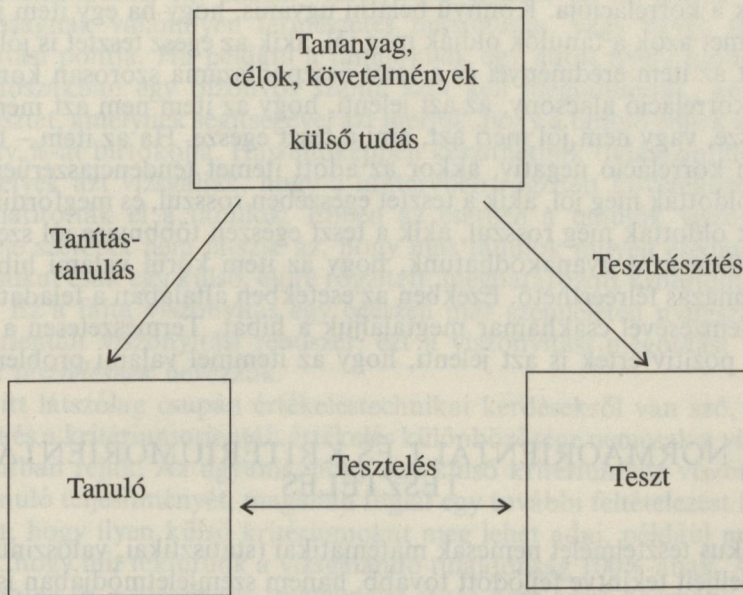
Sok pszichológiai teszt validitása, validitásának igazolása csak igen nehezen megoldható problémát jelent, általában már validnak elfogadott tesztekhez viszonyítják az újonnan kidolgozott teszteket. A pedagógiai tesztek, közelebbről a tudásszintmérő tesztek esetében a kérdés kissé egyszerűbb, vannak olyan támpontjaink és olyan technikai eljárások, amelyek segítségével a teszt készítése során elérhetjük, hogy a teszt megfelelő validitással rendelkezzen.

A tudásszintmérő tesztek validitásának kérdéseit a 10.2. ábra alapján tekinthetjük át. Az egyszerűség kedvéért induljunk ki abból az alapmodellből, hogy a célokban, követelményben, tananyagban rögzített külső tudásból a tanítástanulási folyamatai során a tanuló belső tudása jön létre: ez az egész tanítástanulási folyamat célja és értelme. A tudásszintmérő teszttel azt kívánjuk ellenőrizni, hogy ez a leképezési folyamat valóban megfelelően ment-e végbe, azaz a tanulóban kialakult-e az a belső tudás, amely szerkezetét, összefüggéseit, igazságtartalmát tekintve megfelel a tananyagban rögzített külső tudásnak. Ennek megvizsgálására használjuk a teszteket.

A tesztkészítés a tanítás-tanulás folyamataival analóg folyamat: a célokat, a követelményeket, a tananyagot ugyancsak le kell képeznünk, de ebben az esetben objektívált, rögzített formában. A teszt készítése során ugyancsak el kell érünk azt, hogy a teszt pontosan képezze le a célok, követelmények, tananyag formájában rögzített tudásrendszert.

Az ábrán a tananyagtól kétféle úton juthatunk el a tanuló tudásához: az egyik a tanulás, ez az az út amelyen a tudás létrejött. A másik út a tesztkészítés-teszt-tesztelés: ez pedig a kerülőút, az ellenőrzés, amelynek során összehasonlítjuk a tanuló tudását a tananyaggal. Belátható, hogy ez az összehasonlítás csak akkor végezhető el hitelesen, csak akkor tudjuk a tesztünkkel valóban azt vizsgálni, hogy a tanulás eredményeként milyen belső tudás jött létre, illetve hogy ez a belső tudás megfelel-e a céloknak, követelményeknek, ha a tesztünk pontosan a célok, követelmények, illetve a tananyag alapján készült, pontosan tükrözi azt. Ha tesztünk elegendően részletes és megbízható, akkor a teszteredmények valóban azt mutatják meg, hogy a tanuló tudása mennyiben egyezik meg a tananyag formájában rögzített tudással.

A tudásszintmérő tesztek validitása



10.2. ábra

A tudásszintmérő tesztek validitását tehát azáltal biztosíthatjuk, ha olyan technikát, olyan tesztkészítési eljárást használunk, amelynek segítségével a tananyagot pontosan leképezzük a tananyag mérésére, tesztelésére szolgáló mérőeszközbe. A következőkben a tesztek készítésének azokat a technikai eljárásait mutatjuk be, amelyek az ilyen módon elérhető validitást helyezik a középpontba.

1.5. AZ ITEMÉK JELLEMZŐI

A tesztek mutatóihoz hasonlóan az itemek jellemzésére is használhatunk bizonyos paramétereket. Korábban már megismertedtünk az itemek néhány tulajdonságával, most az itemek jellemzésére szolgáló paramétereket gyűjtjük össze, és röviden megadjuk ezek értelmezését.

Az item nehézsége, vagy *nehézségi indexe* azt mutatja meg, hogy egy tetszőlegesen kiválasztott tanuló milyen valószínűséggel oldja meg az itemet. Gyakorlatilag ez nem más, mint a jó megoldások száma osztva a feladatot megoldó tanulók teljes számával. Az item nehézségi indexe tehát 0 és 1 közé eső szám. Minél nagyobb ez a szám, annál könnyebb az item.

Az item *differenciáló ereje* vagy megkülönböztető képessége azt mutatja meg, hogy az item mennyire érzékeny a tanulók tudására, mennyire élesen tesz különbséget a különböző tudásszintű tanulók között.

Beszélhetünk az *itemek megbízhatóságáról*, reliabilitásáról is. Az itemek reliabilitását is sokféle mutatóval jellemezhetjük. A legegyszerűbb mutató, amiből az itemnek reliabilitására következtethetünk, az itemek és a teszt összpontszámának a korrelációja. Könnyű belátni ugyanis, hogy ha egy item jó, akkor azt az itemet azok a tanulók oldják meg jól, akik az egész tesztet is jól megoldják, tehát az item eredményei és a teszt összpontszáma szorosan korrelálnak. Ha ez a korreláció alacsony, az azt jelenti, hogy az item nem azt méri, amit a teszt egésze, vagy nem jól méri azt, ami a teszt egésze. Ha az item – tesztösszpontszám korreláció negatív, akkor az adott itemet tendenciaszerűen azok a tanulók oldották meg jól, akik a tesztet egészében rosszul, és megfordítva, azok a tanulók oldották meg rosszul, akik a teszt egészén többnyire jól szerepeltek. Ilyen esetben arra gyanakodhatunk, hogy az item körül valami hiba van, a megfogalmazás félreérthető. Ezekben az esetekben általában a feladat, vagy az itemek elemzésével csakhamar megtaláljuk a hibát. Természetesen a már nagyon kis pozitív érték is azt jelenti, hogy az itemmel valami probléma lehet.

1.6. NORMAORIENTÁLT ÉS KRITÉRIUMORIENTÁLT TESZTELEÉS

A klasszikus tesztelmélet nemcsak matematikai (statisztikai, valószínűségelméleti) modelljeit tekintve fejlődött tovább, hanem szemléletmódjában is, amit az ún. kritériumorientált tesztelés kialakulása jelez. Már az előző részben foglalkoztunk azzal, hogy sok pszichológiai teszt esetében nem tudunk külső kritériumokat találni, amelyek segítségével a mérendő tulajdonság maximális értékét megadhatnánk, és nehézséget okoz annak pontos meghatározása is, hogy mit mér a tesztünk. Ezzel szemben a tudásszintmérő teszteknel a cél- és követelményrendszer megfelelő támpontot ad a validitás biztosításához. Ez a különbség vezetett (többek között) a kritériumorientált tesztelés gyors fejlődéséhez (Srittmatter, 1973; Fricke, 1974; Popham, 1978; Berk, 1980; Csapó, 1987).

A kritériumorientált és a klasszikus tesztelés különbségeinek bemutatása érdekében először a klasszikus tesztelmélet és értékelési eljárás egyik lényeges vonását kell tovább elemeznünk. Ahhoz, hogy a teszt összpontszámának, a teszteredménynek valamilyen jelentést tulajdonítsunk, ezt a pontszámot valamihez viszonyítanunk kell. Önmagában például nem tudjuk, mit jelent az, hogy egy tanuló 60 pontot ért el egy teszten. Más ennek a 60 pontnak a jelentése akkor, ha a minta – a tanulóknak az a csoportja, amely a tesztet megoldotta –, 50 pontos átlagteljesítményt ért el. Megint más a helyzet akkor, ha az átlag 70 pont volt. Ebben az esetben a kiszemelt tanulóat annak a csoportnak a teljesítményéhez viszonyítottuk, amelyik a tesztet megoldotta. A viszonyítási szint megadásával nyert értelmet a 60 pontos teljesítmény. Ugyanezt az alapelvet megtartva tovább finomíthatjuk eljárásainkat. Nemcsak azt mondhatjuk meg, hogy ez a 60 pont az átlagnál jobb, vagy rosszabb, de egész pontosan megmondhatjuk azt is, hogy a tanulók hány %-a ért el ennél gyengébb vagy jobb eredményt. Így elkészítettünk egy olyan viszonyítási rendszert, amelyben minden egyes megvizsgált személyt a populáció jellemzőihez viszonyítunk. Mivel itt

a populáció normáihoz viszonyítunk, ezt a viszonyítási eljárást *normaorientált*, vagy *normavonatkozású viszonyítás*nak nevezzük.

Alkalmazhatunk egy másik viszonyítási eljárást is, ha létezik a vizsgált tulajdonságnak valamilyen természetes maximuma, vagy egyéb természetes viszonyítási pontja. Ha például a tanulóknak egy adott nyelv tanulása során adott időszakban egy bizonyos számú szót kell megtanulniuk, készíthetünk olyan tesztet, amelynek segítségével azt mérjük, hogy a tanuló ennek a szókincsnek hány %-át birtokolja. Hasonlóképpen készíthetünk tudásszintmérő teszteket, amelyek azt vizsgálják, hogy a tantervben rögzített követelmények hány %-át sajátították el a tanulók. Ebben az esetben a tanulók viszonyításánál semmiféle szerepet nem játszik az, hogy más tanulók milyen eredményt értek el, tudásukat csak egy külső, előre rögzített, meghatározott kritériumhoz viszonyítjuk. Ez a fajta viszonyítás egy egészen más szemléleten nyugszik, mint a normaorientált viszonyítási rendszer, ezt a viszonyítási megoldást *kritériumorientált tesztelésnek* nevezzük.

Bár itt látszólag csupán értékeléstechnikai kérdésekről van szó, a normaorientált és a kritériumorientált értékelés különbözősége nemcsak a viszonyítási szempontban rejlik. Az ugyanis, hogy egy külső kritériumhoz viszonyítjuk az adott tanuló teljesítményét, magában foglal egy további feltételezést is, nevezetesen azt, hogy ilyen külső kritériumokat meg lehet adni, például meg tudjuk adni azt, hogy mit tekintünk a vizsgálandó tulajdonság 100%-ának. A pszichológiai tesztek és a klasszikus tesztelmélet tehát nem azért használja a normavonatkoztatású viszonyítási rendszert, mert ezt tartja a legelőnyösebbnek, hanem azért, mert sok pszichikus tulajdonság tekintetében nem tudunk megadni természetes maximumot. Nem lehet például megadni azt, hogy mennyi az intelligencia 100%-a, vagy ki a 100%-osan kreatív ember, mi az a maximum, ami fölé nem mehetünk.

A tudásszintmérő tesztekhez lehet maximumokat, külső kritériumokat kitűzni. Ha arra vagyunk kíváncsiak, hogy a tanulók milyen mértékben teljesítettek a követelményeket, akkor ezeket a kitűzött követelményeket a tanulók különböző mértékben teljesíthetik. Teljesíthetik 20, 30, 40, 60, vagy akár 100 százalékos mértékben. Ebben az esetben természetesen nem a tanuló teljes, különböző forrásokból megszerzett tudását értékeljük. Lehet, hogy bizonyos elemeiben a tanuló tudása messze meghaladja azokat a követelményeket, amelyeket mi számára kitűztünk. A dolog természeténél fogva ilyenkor is csak azt tudjuk a teszt eredményei alapján mondani, hogy a tantervi követelményeknek 100 százalékosan megfelel.

A kritériumorientált tesztelés mindezeknek megfelelően sajátos tesztkészítési technikákat igényel. E tesztkészítési technikák középpontja, alapja a tananyag megfelelő elemzése. Csak a tananyag elemzése, szerkezetének pontos feltárása révén készíthetünk megfelelő teszteket, vagyis csak így érhetjük el azt, hogy tesztünk pontosan leképezze a tananyagot. A megfelelő teszt készítéséhez azonban nemcsak a tananyagot kell ismernünk, hanem pontos képpel kell rendelkez-nünk arról is, hogy milyen természetű az a tudás, milyen jellegűek azok a rendszerek, amelyek a tanulók pszichikumában kialakulnak. A tesztelés során ugyanis nem egyszerűen a tanuló által elsajátított külső tudás meglétét vizsgál-

jük, hanem egészen más minőséget: pszichikus tulajdonságokat, amelyek alapvetően más sajátosságokkal rendelkeznek, mint az egyébként elemezhető, tanönyvek, tananyagok formájában megjelenő külső tudás.

2. FELADATÍRÁS

2.1. A TANANYAG ELEMZÉSE, A FELMÉRENDŐ TUDÁS KATEGORIZÁLÁSA

Egy-egy tesztfeladattal, teszt-ítemmel az elsajátított tudásnak csak meglehetősen kis elemeit tudjuk vizsgálni. Ahhoz, hogy a tudásnak már egy ilyen kis eleméhez is hozzáférjünk, és ugyanakkor a tudás elemenkénti megközelítésével a vizsgálandó tudás teljes épületét is megfelelően építsük fel, szükség van arra, hogy a tananyagot megfelelő módon elemezzük.

Mind a tananyag elemzésére, mind a tananyag elsajátítása nyomán kialakuló pszichikus rendszerek számbavételére sokféle megoldás született, sokféle technológia ismeretes. Az oktatás céljainak, és így az értékelés objektumainak a számbavételére is alkalmazható egyik legelterjedtebb taxonómiai rendszer Benjamin Bloom nevéhez fűződik. Bloom, majd később munkatársai értékelési taxonómiájukat a személyiség három nagy területére, a kognitív, az affektív és a pszichomotoros szférára dolgozták ki.

A bennünket közelebről érdeklő kognitív taxonómia rendszerezte az elsajátított tudás különböző elemeit, és ezek között az elemek között hierarchikus szinteket állapított meg. Bloom az összegyűjtött elemeket 6 hierarchikus szintbe foglalta össze, ezek a következők: ismeret, megértés, alkalmazás, analízis, szintézis, értékelés (Bloom 1956). Bloom elmélete hamarosan népszerűvé vált, és széles körben alkalmazták az oktatás céljainak rendszerezésére, és az értékelés, a tesztek kidolgozásának megalapozására is (Bloom, Hastings és Madaus 1971). Bár a taxonómia fogalomrendszere sokáig igen nagy hatást gyakorolt a tesztkészítés fejlődésére, mára már sokan bírálják a koncepcionális megalapozás hiányát, a taxonómiába felvett elemek esetlegességét. A Bloom által feltételezett hierarchiát sok vizsgálat nem igazolta. Más hasonló, vagy bonyolultabb modellek is születtek, amelyek ismertetésére nem vállalkozhatunk. A következőkben a pedagógia elterjedt szóhasználatánál maradva egy leegyszerűsített megoldást vázolunk fel. (A tananyag illetve a tudás szerkezetével, elemzésével részletesen foglalkozik Ágoston, Nagy és Orosz, 1979; Kádárné és Joó, 1977; Nagy, 1972, 1979, 1985; Orosz, 1977).

A tanulók által elsajátított, a tesztekkel megvizsgálandó tudást két fő szférára osztjuk fel. Az egyikbe tartoznak az ismeret-jellegű tudáselemek, így például a képzetek, a fogalmak, a tények, a definíciók, a szabályok, a leírások, a törvények és az elméletek. Ezekre a tudáselemekre, különösen a képzetekre, tényekre, egyszerű fogalmakra, definíciókra az jellemző, hogy azok viszonylag rövid idő alatt elsajátíthatók, megfelelő számú ismétlés és rögzítés után tartós tudássá válhatnak. A tudásnak ebbe a szférájába tartoznak például a következő tények:

a vas vegyjele Fe, a mohácsi csata 1526-ban volt, az asztal németül der Tisch stb.

A tudás másik nagy csoportjába, a képesség-jellegű tudás körébe tartoznak a készségek, a jártasságok és az általános képességek. Készség például az íráskészség, az olvasás elemi készségei, az alapműveleti számolási készségek. Jártasságot alakíthatunk ki például a tananyag egy meghatározott témakörében való feladatmegoldásra, így a gáztörvényekkel kapcsolatos számításokra vagy a nyelvtanban a mondatelemzésre, vagy jártassá válhatunk szövegek idegen nyelvre történő fordításában. Általános képesség például a problémamegoldó gondolkodás, az intelligencia, a kreativitás, vagy a kommunikáció különböző képességrendszerei.

A képesség-jellegű tudást az jellemzi, hogy hosszú fejlődési folyamaton keresztül jut el a kezdetektől egy viszonylag állandósult szintre. Gondoljunk például az íráskészségre, amelynek elemeit, az írásmozgás-koordinációt már az óvodában elkezdik a gyermekek tanulni, a betűírást az első osztály végére megtanulják, tűrhető sebességű írással rendelkeznek az általános iskola végére, de az állandósulás jegyei csak a kamaszkor vége felé jelentkeznek, és a kiirt írás, amely állandóvá válik, csak a felnőttkor küszöbén jelenik meg. Számolási készségeink is viszonylag hosszú időn keresztül fejlődnek és alakulnak ki, bizonyos általános képességeink pedig esetleg egész életünkön keresztül is fejlődhetnek. Ezeknek a képesség-jellegű tudáselemeknek az elsajátítása többnyire nem rendelhető az oktatási folyamat egy viszonylag rövid szakaszához, mint ahogy azt bizonyos ismeretek esetében megtehetjük.

Az ismeret-jellegű és a képesség-jellegű tudás még sok más sajátosságban is különbözik, amely a tesztelés szempontjából fontos lehet számunkra. Ezek közül itt most csak néhányat sorolunk fel.

Az ismeret-jellegű tudás (mivel annak elsajátítása az oktatási folyamat egy rövidebb szakaszán történik) felmérése többnyire egy rövidebb időszakasz eredményeiről ad képet. Ezzel szemben bármikor is történik a képesség-jellegű tudás vizsgálata, annak eredményében mindig az oktatási folyamat egy hosszabb szakasza van jelen. Előfordul például, hogy a 8. osztályosok olvasási színvonalát mérjük fel. Ekkor természetesen nem a megelőző 1–2 hónap iskolai munkájának eredményességét vizsgáljuk, hanem a megelőző 6–8 év eredményei tükröződnek az olvasásszint-felmérésben. Így tehát mások lehetnek a teszt eredményeiből levont következtetések, az azok alapján megfogalmazható feladatok is. A tanulók képességeiben talált hiányosságokat csak hosszú, néha több éves kemény munkával lehet pótolni.

Fontos különbség az ismeret-jellegű és a képesség-jellegű tudás között az is, hogy az ismeret-jellegű tudást, annak elemeit a maguk konkrét formájában tehetjük a tesztelés tárgyává. Például konkrétan rákérdezhetünk, hogy mi a vas vegyjele, mikor volt a mohácsi vész stb. Ezzel szemben a képesség-jellegű tudást a működés felől, a szerkezet felől közelíthetjük meg. Például az olvasás készségét többféle tartalom is vizsgálhatjuk, sokféle szöveget olvastathatunk az olvasási készség felméréseként. Hasonlóképpen különböző szövegekkel vizsgálhatjuk az íráskészséget, fizika feladatokat különböző számértékekkel adhatunk és így tovább. A képesség-jellegű tudásnak tehát a szerkezetbeli és működésbeli sajátosságait kell leírni és elemeznünk.

A tudás e két formája többnyire meghatározza az alkalmazható tesztszerkesztési technikát, a tesztelésre használható feladattípusokat, tesztelési eljárásokat.

2.2. FELADATTÍPUSOK

A tesztfeladatok készítésének óriási irodalma van. A mérés céljaitól függően sokféle tesztet és tesztfeladatot lehet készíteni. Ezek közül a következőkben csak olyan feladatokkal foglalkozunk, amelyek a tudásszint mérésére alkalmasak, még szűkebben csak azokkal, amelyekből papír-ceruza teszteket lehet összeállítani. Természetesen még ezen a csoporton belül is csak vázlatos áttekintésre vállalkozhatunk.

A tudásszint mérésére szolgáló feladatok kiindulásként két nagy csoportra oszthatók, az egyik csoportba a feleletválasztó, a másikba a feleletalkotó feladatokat soroljuk.

Mindkét feladattípusnak megvannak a maga relatív előnyei és hátrányai, továbbá nem egyformán alkalmazhatók a tudás különböző komponenseinek a mérésére.

Az iskolai használat során a feleletválasztó feladatok néhány típusának nagy hátránya az, hogy hamis, nem igaz, ámde hihető válaszalternatívák nagy tömegének az elolvasására készítetik a tanulót, és gyenge, bizonytalan tudás esetén van bizonyos valószínűsége a hibás válaszok rögzítésének. Ezért pedagógiai szempontból a hibás válaszokat nem tartalmazó típusokat, illetve az aktív válaszadást előnyben részesítjük.

A) Feleletválasztó feladatok

A feleletválasztó kategóriába azokat a feladatokat soroljuk, amelyeknek a megoldása során a tanuló kész, előre megadott válaszok közül választja ki a megfelelőt (helyeset, jót, igazat stb.), és azt valamilyen módon megjelöli. A feleletválasztó feladatok javítása a legegyszerűbb, kidolgozásuk azonban általában hosszabb időt vesz igénybe. A nehézségek különösen a többszörös választást tartalmazó feladatoknál jelentkeznek, mivel formailag azonos, a megfelelő kontextusba egyaránt illő (egyaránt hihető, a szükséges tudással nem rendelkezők számára jónak tűnő) válaszalternatívákat kell kidolgozni. A feladatok értékelő rendszerének, javítókulcsának és pontozásának elkészítése viszont könnyű.

A feleletválasztó feladatok elsősorban a tudás egyszerűbb elemeinek, az ismeret-jellegű tudásnak a mérésére alkalmasak. Kevésbé, vagy csak igen elmélyült tesztelés révén használhatók a képességeknek, a gondolkodás rugalmasságának, a tudás minőségének, a megértésnek, a szintézisnek a vizsgálatára. Ez azonban nem zárja ki azt, hogy akár az általános képességek vizsgálatára is feleletválasztó tesztet szerkesszünk.

A feleletválasztó tesztfeladatok sokféle típusa közül itt csak néhányat muta-

tunk be. (A közölt minták csak az illusztrációt szolgálják, egy részük kitalált példa, más részük a JATE Pedagógiai Tanszékén készült tesztekben származik.) A helytel való takarékoskodás érdekében csak a feladatok lényeges elemeit emeljük ki, tehát nem tüntetjük fel mindazokat a formai tartozékokat, amelyekkel egy valódi tesztbe illesztett feladatnak rendelkeznie kell.

a) Alternatív választás

Ennél a feladattípusnál csak két válaszlehetőség adott, és ezek közül az egyik mindig jó, a másik rossz megoldás. A véletlen találat valószínűsége itt a legmagasabb, 50%. Fontosabb típusai a következők:

(1) Állítás, amelyről el kell dönteni, hogy igaz, vagy hamis.

Ha egy zárt térben levő gáz hőmérsékletét növeljük, a nyomása csökken.
igaz hamis

(2) Feladat, amelyről el kell dönteni, hogy a megadott megoldás jó-e. Például jó-e az alábbi angol mondat?

The boy have a book.
jó rossz

(3) Eldöntendő kérdést tartalmazó kérdőmondat.

Voltak-e Mátyás király híres könyvtárában Amerikáról szóló útleírások?
igen nem

b) Többszörös választás

A feladatban több válaszlehetőség van megadva, ezek közül lehet egy, de több is helyes. A feladatnak mindig közölnie kell, hogy egy vagy több választ kell a teszt megoldójának megjelölnie. A véletlen találat valószínűségét matematikailag a válaszlehetőségek száma és a megjelölendő válaszok száma határozza meg, gyakorlatilag az is számít, hogy a válaszlehetőségek mennyire „hasonlítanak” egymásra. Ha ugyanis egy rossz válaszról a vizsgálandó tudás birtoklása nélkül is nyilvánvaló, hogy rossz, akkor valójában csak a többi jön számításba, mint reális válaszalternatíva. Ezzel a megoldással nagyon sokféle feladatot lehet szerkeszteni, itt csak néhány alaptípust sorolunk fel.

(1) Egy jó válasz

Mi az Amerikai Egyesült Államok fővárosa?

- (a) Montreal
- (b) New York
- (c) Washington
- (d) Los Angeles

c

- (2) Több jó válasz
Kik voltak az első triumvirátus tagjai?
(a) Antóniusz
(b) Cézár
(c) Lepidusz
(d) Krasszus
(e) Pompejus
(f) Oktáviusz

b, d, e

- (3) A legjobb válasz kiválasztása
Melyik szó adja vissza legjobban magyarul a „kompetencia” jelentését?
(a) tudás
(b) szaktudás
(c) műveltség
(d) szakértelem
(e) képzettség

d

- (4) A rossz válasz kiválasztása
Melyik az a város, amelyik nem az Egyesült Államokban van?
(a) Boston
(b) Chicago
(c) San Francisco
(d) Canberra
(e) Santa Barbara

d

c) Válaszok illesztése

Ennél a kérdéstípusnál két halmaz elemei között kell a kapcsolatot megtalálni. A két halmaz nagyon sokféle módon állhat tartalmi kapcsolatban egymással, például dolgok és tulajdonságaik, szerzők és művek, fogalmak és általánosabb fogalmak stb. A hozzárendelés típusa lehet egy az egyhez és egy a többhöz (ritkábban több a többhöz is). Természetesen a feladatban mindig pontosan rögzíteni kell, hogy hogyan várjuk a választ.

- (1) Egy az egyhez hozzárendelés.
Kapcsoljuk össze a szerzőket és a műveket!
(a) A fekete város (A) Ady Ende
(b) A fekete kolostor (B) Mikszáth Kálmán
(c) A fekete zongora (C) Kuncz Aladár
a-B b-C c-A

(2) Egy a többhöz hozzárendelés (osztályozás).

Melyik szó melyik szófajhoz tartozik?

- | | |
|---------------|-------------|
| (a) ige | (A) Petőfi |
| (b) főnév | (B) alszik |
| (c) melléknév | (C) kicsi |
| (d) számnév | (D) sok |
| (e) névmás | (E) magas |
| | (F) enyém |
| | (G) olvas |
| | (H) írás |
| | (I) szépség |

a: B, G b: A, H, I c: C, E d: D e: F

A feleletválasztó feladatoknak a felsoroltakon kívül még nagyon sok változata használatos. Mivel a számítógépes tesztelést is egyszerűbb feleletválasztó feladatokkal megoldani, a számítógépes oktatás korai szakaszában ez a forma terjedt el. Ma azonban, a programok „intelligensebbé” válásával már a számítógépek is képesek a tanulók által alkotott bonyolult válaszokat is értékelni.

B) Feleletalkotó feladatok

A feleletalkotó feladatok megszerkesztése viszonylag könnyű munkának tűnik, hiszen látszólag általában elegendő egy felszólító vagy kérdőmondatot megfogalmazni. Azonban e feladatok értékelő rendszerének megalkotása, az egyértelmű, kellő objektivitással használható javítókulcs kidolgozása többnyire időigényes feladat.

A feleletalkotó feladatokkal a tudás bonyolultabb összetevőinek a mérését is megoldhatjuk, azonban ez gyakran speciális értékelő eljárást is igényel. Kellő gyakorlattal azonban a feleletalkotó feladatok javítókulcsát annyira egyértelművé lehet tenni, hogy az elfogadható válaszlehetőségek, illetve az azokból értékelhető mozzanatok olyan kategóriákba sorolhatók be, amilyeneket a feleletválasztó feladatok esetében is kapunk.

a) Kiegészítés

Ez a feladattípus többnyire egy kijelentő mondat, amelyből bizonyos fogalmat vagy fogalmakat kihagytunk.

(1) Egy kihagyott fogalom.

Az cselekvést, történést, létezést jelentő szó.

(ige)

(2) Több kihagyott fogalom.

A forgatónyomaték megegyezik az és az szorzatával.

(erő, erőkar)

b) Rövid válasz

Ennél a feladattípusnál egy kiegészítendő kérdést úgy fogalmazzunk meg, hogy arra röviden, lehetőleg egyetlen szóval, névvel vagy számmal lehessen válaszolni. Értékelése teljesen egyértelművé tehető, ha a kérdést úgy fogalmazzuk meg, hogy csak egyetlen jó megoldást lehessen adni.

- (1) Válasz tulajdonnévvel
Ki a Csongor és Tünde szerzője?
(Vörösmarty Mihály)
- (2) Válasz számmal
Mikor volt a Mohácsi csata?
(1526-ban)
- (3) Válasz egyéb szóval
Melyik névelő tartozik a Tisch főnévhez?
(der)

c) Hosszú válasz

A feladat olyan kérdést tartalmaz, amely hosszabb, általában egész mondatos választ vagy egy felsorolást igényel. A válasz értékelése bonyolultabb, mivel ugyanazt a tartalmat többféle mondattal is meg lehet fogalmazni. Ekkor azt kell megítélnünk, hogy a válasz kifejezi-e a lényegét. Ezt pedig úgy tehetjük meg, hogy megadjuk azokat a kifejezéseket, amelyeknek elő kell fordulniuk a válaszban, és azt, hogy ezeknek milyen kapcsolatban kell állniuk egymással.

- (1) Válasz egy mondatral (definíció).
Mi az ige?
(szó; és cselekvést, vagy történést, vagy létezését jelent)
- (2) Válasz felsorolással.
Melyek a halogén elemek?
(fluór, klór, bróm, jód, asztácium)

d) Esszé típusú válasz

A legbonyolultabb értékelési problémákat az egy mondatnál hosszabb összefüggő válaszokat igénylő feladatok jelentik. Az ilyen feladatok alkalmasak az önálló ítéletalkotás, az összefüggések felismerésének, a szintetizálás, a lényegkiemelés képességének a vizsgálatára. Az értékelés objektivitását általában azoknak az állításoknak a felsorolásával lehet elérni, amelyeknek elő kell fordulni a válaszban. Ilyen kérdések lehetnek a következők:

(1) Elemezze a mohácsi csatavesztés okait!

(2) Foglalja össze a kénsavgyártás folyamatait!

C) A képesség-jellegű tudás mérésére alkalmas feladatok

A korábban bemutatott feladattípusok elsősorban az ismeret-jellegű tudás vizsgálatára alkalmasak. Természetesen megfelelő technikával néhány típusuk a képességek mérésére is alkalmas, az ismeret- és a képesség-jellegű tudás közötti különbségek miatt azonban szükség van más, speciális feladattípusok alkalmazására is. A képesség-jellegű tudás vizsgálatára szolgáló feladatok nagyon sokfélék lehetnek, néha egészen egyszerűek, máskor bonyolult értékelési technikákra van szükség. Ezeknek a feladattípusoknak a részletes bemutatására területi okok miatt nem vállalkozhatunk, de az érdeklődő olvasó gazdag példátárat találhat az ajánlott irodalomként felsorolt munkákban.

Az egyszerű alpműveleti számolási készségek tesztelésére jól használhatók azok a feladatok, amelyeket az iskolai gyakorlatban egyébként is használunk. Többnyire elegendő kitűzni magát a feladatot, az eredményről egyértelműen eldönthető, hogy jó vagy rossz. Például:

$$27 + 16 = \dots \quad 48 - 17 = \dots \quad 33 + \dots = 51$$

Hasonlóképpen könnyen teszté alakíthatjuk a szöveges feladatok megoldásának jártasságait felmérő szokásos iskolai feladatokat. Mivel az ilyen feladatok már több önállóan is értékelhető részfeladatból állnak, mindig pontosan meg kell határozni a feladatot felépítő itemeket és az azokra adható pontszámokat. Ilyen jellegű feladatot lehet készíteni a fizikában és a kémiában kialakítandó feladatmegoldó jártasságok vizsgálatára is.

Speciális technikákat igényel az anyanyelvi kommunikáció képességrendszerének vizsgálata. Az írás készségeit az egyszerű írásmozgás-koordinációtól kezdve a helyesírás és mondatalkotás készségein és jártasságain keresztül a fogalmazástechnika képességeiig vizsgálhatjuk, és erre ma már sokféle tesztelési technika áll rendelkezésünkre.

A képességek kialakulása hosszú fejlődési folyamat eredménye. Amikor a képességeket (készségeket, jártasságokat) vizsgáljuk, azt akarjuk megállapítani, hogy az egyes tanulók ebben a fejlődési folyamatban hol tartanak. Ezért a képesség-jellegű tudás fejlettségét két olyan mutatóval is jellemezhetjük, amelyek az ismeret-jellegű tudással kapcsolatban nincs lényeges szerepe: beszélhetünk a képességek (készségek, jártasságok) működésének a sebességéről és hibátlanúságáról.

A készségek és jártasságok sebessége a kezdeti megjelenésüktől az automatizáció és a begyakorlottság növekedésével a teljes kifejlődésig egyre fokozódik. A sebességet általában az időegység alatt megoldott feladatok száma jellemezzük. Például az alpműveleti számolási készségek esetében használhatjuk a művelet/perc dimenziót.

A képességek működésének fejlődését, összehangolódását az is mutatja, hogy csökken a hibázások száma. A hibátlanúsági arányt jellemezhetjük például

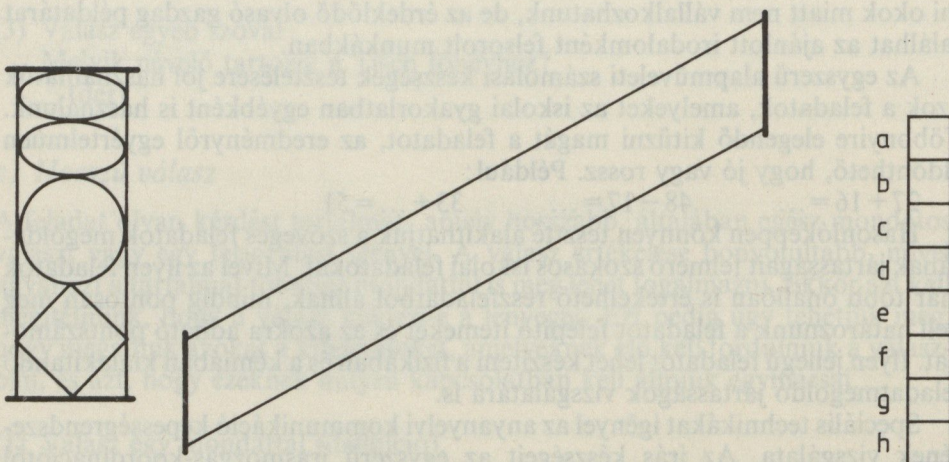
azzal, hogy a jól megoldott feladatelemek számát az összes megoldott feladatelem számának százalékában fejezzük ki.

A készségek és a képességek esetében mind a sebesség, mind pedig a hibátlanági arány (minőség) tekintetében használhatjuk külső viszonyítási pontként (kritériumként) a felnőtt szakember megfelelő mutatóit.

A képesség-jellegű tudás mérése néha egészen speciális megoldásokat igényel. Ezekre a speciális technikákra három példát mutatunk be.

A 10.3. ábrán bemutatott feladat a rajzkészség egyik összetevőjének, a térszemléletnek az elemeit vizsgálja (Csapó és Varsányi, 1985). A feladat megoldásához szükséges transzformációs műveletek, a szem-kéz koordináció, a kéz-ügyesség természetesen nemcsak a rajzolásban játszanak szerepet.

A rajzkészség összetevőit vizsgáló teszt egyik feladata



10.3. ábra

A feladat értékelése a 10.4 ábrán (302. oldal) látható átlátszó fóliával történik, melyet a tanuló rajzára helyezünk, és leolvassuk, hogy megoldása mennyire pontos. Ha az egyes érintkezési pontok a megadott intervallumba esnek, mind-egyikre egy pontot lehet kapni.

A következő feladat a gondolkodás egyik műveleti képességének, a nyelvi-logikai műveletek kialakultságának a vizsgálatára szolgál. Nemcsak arra alkalmas, hogy segítségével eldöntsük, vajon a tanuló az egyszerű kijelentésekből alkotott összetett kijelentést a logikai konvencióknak megfelelően értelmezi-e, hanem azt is megállapíthatjuk, hogy ha nem így érti, akkor mit ért helyette.

Gondosan hasonlítsd össze a kijelentést a felsorolt tényekkel, és állapítsd meg, hogy melyik ténnyel való összehasonlításban igaz, melyikben hamis a kijelentés! Minden tényt egyenként hasonlítsd össze a kijelentéssel. Amennyiben igaznak találsz, a tény előtti betűjelzést karikázd be! Ha megítélésed szerint hamis, akkor a tény előtti betűjelzést húzd át!

János kijelentése: MA NEM FELELTEM SEM MATEMATIKÁBÓL,
SEM MAGYARBÓL.

Tények:

A Matematikából is és magyarból is felelt.

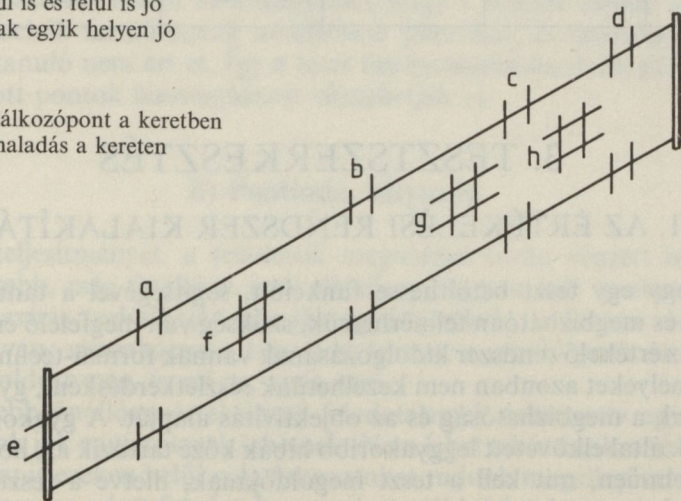
B Matematikából felelt, de magyarból nem.

C Magyarból felelt, de matematikából nem.

D Sem magyarból, sem matematikából nem felelt.

A rajzkészség összetevőit vizsgáló tesztfeladat javítására szolgáló fólia rajza

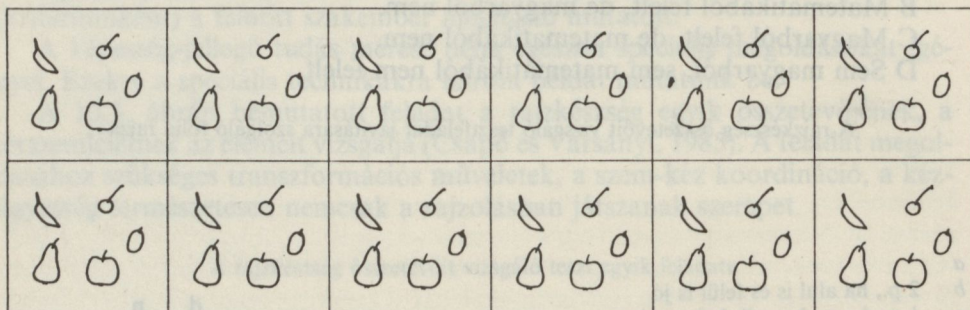
- a
- b 2 p., ha alul is és felül is jó
- c 1 p., ha csak egyik helyen jó
- d
- e 1 p., ha jó
- f 2 p., ha találkozópont a keretben
- g
- h 1 p., ha áthaladás a kereten



10.4. ábra

Az alábbi feladat egy másik művelési képesség, a kombinatív képesség értékelésére alkalmas (Csapó, 1988). A tanuló teljesítményének jellemzésére alkalmazhatunk mennyiségi mutatókat (hány konstrukciót alkotott meg, ezek közül hány a jó, hány a feleslegesen ismétlődő), és jellemezhetjük a gondolkodás minőségét is annak elemzésével, hogy a felsorolás során a tanuló milyen gondolkodási stratégiát használt.

Karikázz be egy zárt vonallal három gyümölcsöt minden lehetséges módon! Minden kis ábrán különböző összeállításokat keretezz be! Vigyázz, több ábra van, mint amennyi különböző lehetőséget találni fogsz!



3. TESZTSZERKESZTÉS

3.1. AZ ÉRTÉKELÉSI RENDSZER KIALAKÍTÁSA

Ahhoz, hogy egy teszt betölthesse funkcióit, segítségével a tanulók tudását objektíven és megbízhatóan felmérhessük, szükség van megfelelő értékelő rendszerre is. Az értékelő rendszer kidolgozásának vannak formai-technikai követelményei, amelyeket azonban nem kezelhetünk részletkérdésként, gyakran éppen ezek képezik a megbízhatóság és az objektivitás alapját. A gyakorlatlan tesztszerkesztők által elkövetett leggyakoribb hibák közé tartozik az, hogy nem írják elő egyértelműen, mit kell a teszt megoldójának, illetve a teszt javítójának tennie.

A) A javítókulcs és az értékelés technikája

A tesztelés alapvető feltétele, hogy a teszt megoldója pontosan tudja, hogy mit kell tennie, milyen módon kell a feladatot megoldania, a kérdést megválaszolni. Ezért minden tesztnek, vagy ha a tesztben különböző típusú feladatok vannak, akkor minden kérdésnek vagy feladatnak pontosan tartalmaznia kell, hogy mi az, amit a megoldótól elvárunk. Ügyelni kell a kérdések megfogalmazására, helyes, ha a kérdés nyelvtani formája is kérdőmondat, és pontosan kifejezi, hogy eldöntendő, vagy kiegészítendő kérdésről van szó. Ha válaszalternatívákat adunk, pontosan közölni kell, hogy mi a teendő, egy vagy több jó válasz van, a jót kell bekarikázni, vagy a rosszat áthúzni. Ha azt akarjuk, hogy a tanulók nagyobb figyelemmel olvassanak el minden egyes válaszlehetőséget, megkövetelhetjük egyidejűleg mindegyiket, vagyis a jók bekarikázását és a rosszak áthúzását is (l. például a korábban bemutatott logika feladatot). Így mód van

annak megkülönböztetésére is, hogy a tanuló rosszul válaszolt, vagy hozzá sem fogott a feladat megoldásához.

Hasonló szigorúsággal kell előírunk a tesztet javító pedagógusok feladatát is. Erre a célra a tesztekhez külön javítókulcsokat kell készítenünk, amelyek leírják a javítás módját, és megadják a jó megoldásokat. A feladat típusától függően közlik a jó megoldás kódját, vagy megadják azokat a szavakat, tényeket, amelyeknek a válaszban elő kell fordulniuk ahhoz, hogy azokat jónak fogadhassuk el.

A tesztet el kell látnunk az értékeléshez szükséges tartozékokkal. A tudás mérésére szolgáló tesztek esetében célszerű a feladatok vagy feladatelemek mellett feltüntetni a megoldásukkal elérhető pontszámot. Ez már a teszt megoldása során orientálhatja a tanulókat, a kijavított tesztet visszakapva pedig tájékoztat a megszerzett és elvesztett pontokról. Mindemellett a javítást is egyszerűvé teszi. Jól bevált az a megoldás, hogy a feladat mellett a lap szélén külön kódkockák tartalmazzák az elérhető pontokat, és javításkor áthúzzuk azt, amit a tanuló nem ért el. Így a teszt összpontszámának megállapítását az át nem húzott pontok összeadásával végezhetjük el.

B) Pontozás, súlyozás

A tanulók teljesítményét, a feladatok megoldása során végzett munkáját a lehető legkisebb, még önállóan értékelhető egységekre kell bontanunk. Ezzel nemcsak a javítás technikáját tehetjük egyértelművé, de fontos visszacsatoló funkciója is van: visszakapva a kijavított tesztet, a tanuló pontosan tudja, mi az, amit jól oldott meg, és mi az, amit nem.

A legkisebb önállóan értékelhető feladatelemek (alternatív egységek, itemek) már csak jók vagy rosszak lehetnek. Nem lehet tehát azokra részpontokat adni, ha ugyanis ezeken belül még fokozatokat tudunk megállapítani, akkor ez azt jelenti, hogy az adott feladatelemet még tovább lehet bontani, és akkor ezt meg is kell tennünk.

További kérdés az, hogy az egyes feladatelemekre hány pontot adjunk. Azt az eljárást, amelynek során a feladatelemekhez különböző pontszámokat rendelünk, *súlyozásnak* nevezzük, a feladatelemekhez tartozó pontszám a feladatelem *súlya*.

Nyilvánvaló követelmény, hogy ha a tudásnak vannak fontosabb és kevésbé fontos elemei, akkor egy fontosabb elemnek a tudásával vagy nem tudásával több pontot lehessen szerezni, illetve veszíteni, mint egy jelentéktelenebb elemmel. Azt is elvárhatjuk, hogy egy nehezebb feladat megoldásával több pontot lehessen szerezni, mint egy könnyebb. De hogyan lehet megállapítani, hogy melyik elem fontosabb, melyik nehezebb? A kérdés megválaszolására többféle megoldás is létezik, elméletileg azonban egyik sem teljesen kifogástalan. Mind-egyik tartalmaz valamilyen szubjektív vagy önkényes mozzanatot, ez azonban gyakorlati szempontból többnyire nem okoz problémát.

A tudáselemek relatív fontosságának meghatározására a szakértők közmegegyezése szolgálhat: valamilyen módon meg kell szavaztatni a szakembereket

(esetünkben a szaktanárokat), hogy melyik tudáselemet mennyire tartják fontosnak. Ez történhet például úgy, hogy megkérjük, rendezzék sorba az elemeket fontosságuk szerint (kevés elemnél), vagy értékeljék az elemek fontosságát egy (például öt vagy tíz fokozatú) skálával. Kellő számú szaktanár véleményét begyűjtve, statisztikai átlagolással állapíthatjuk meg az elemek egymáshoz viszonyított súlyát.

A feladatelemek nehézségének megállapítását empirikusan is elvégezhetjük: kellő számú tanulóval megoldatjuk a tesztet, majd (ha a feladatok jóságát egyébként már bizonyítottuk), a feladatot annál nehezebbnek tekintjük, minél kevesebb tanuló oldotta meg.

Még mindig kérdés marad, hogy mekkora a legfontosabb és a legkevésbé fontos elemek pontszámainak az aránya. Itt is csak gyakorlati megfontolásokból indulhatunk ki (amit valószínűségi számításokkal is alátámaszthatunk): nem lenne helyes, ha egyetlen feladattal túlságosan sokat lehetne nyerni vagy veszíteni. Praktikusan tehát azt lehet ajánlani, hogy ha az itemek legkisebb pontértéke az 1 pont, akkor a legmagasabb súlya se legyen 5-nél nagyobb. Célszerű, ha ez a terjedelem már magában foglalja a fontosságból és a nehézségből fakadó különbségeket is.

Ha egy pedagógus saját használatra készít tesztek, az előző szempontokat mérlegelve alakíthatja ki a feladatok súlyát. A feladatelemek fontosságának megítélésében csak saját szakmai kompetenciájára támaszkodhat, a nehézség mérlegeléséhez felhasználhatja a tanulók korábbi eredményeit.

Ugyancsak a súlyozással függ össze az is, hogy néha kényelmi szempontból célszerű a pontszámokat úgy megállapítani, hogy a teljes tesztre összesen 100 pontot lehessen kapni. Ez azt jelenti, hogy az eredeti (nyers-) pontokat át kell számítani százalékpontra. A számítást egyszerűen elvégezhetjük, ha minden egyes nyerspontot elosztunk a teszt összpontszámával, és megszorozzuk százzal. Így viszont többnyire nem egész számot, hanem tizedes törtet kapunk az egyes itemek százalékpontjaként. Ezen ismét úgy segíthetünk, hogy a pontszámokat kerekítjük, de olyan módon, hogy összességében a teszt összpontszáma 100 maradjon. Látható, hogy a százalékpontra való áttérés kismértékben megváltoztatja a feladatok egymáshoz viszonyított súlyát. Ezért mindig a gyakorlati szempontok alapján kell eldöntenünk, mit tartunk fontosabbnak, a kifejezőbb (az összehasonlítás lehetőségét is kínáló) százalékpontot vagy a pontosabb súlyozást.

3.2. EKVIVALENS TESZTVÁLTOZATOK KÉSZÍTÉSE

Több elméleti és gyakorlati szempont is indokolja, hogy ugyanannak az átfogó tudásterületnek a felmérésére azonos módon használható, egymással egyenértékű tesztek álljanak rendelkezésünkre. Például célszerű, ha az egymás mellett ülő tanulók nem pontosan ugyanazokat a feladatokat oldják meg, vagy esetleg szükség van arra, hogy ugyanannak a tanulócsoporthoz a tudását egy adott témakörben többször is megvizsgáljuk. Ilyen esetekben használhatjuk az *ekvivalens tesztváltozatokat*.

Más szempont is indokolja, hogy egy adott témakörben több ekvivalens tesztváltozatot is összeállítsunk. A tudásszintmérő tesztek készítése során arra törekszünk, hogy egy témakör teljes anyagát tesztfeladatokká fogalmazzuk át. Ez rendszerint azt jelenti, hogy több feladatunk van, mint amennyit egy tanulónak feladhatunk, hiszen nincs annyi idő a tesztelésre, hogy mindenkitől mindent kikérdezzünk. De egyben azt is szeretnénk, hogy a teljes tananyag elsajátításának színvonaláról képet kapjunk legalább az osztály szintjén, vagy ha szélesebb körű felmérésről van szó, akkor iskolai, megyei vagy országos szinten.

Két tesztet akkor tekintünk ekvivalensnek, ha a mérést a tesztekkel elvégezve a két teszt eredményei minden tanuló esetében megegyeznek. Ezt az ideális határesetet a gyakorlatban természetesen nem lehet elérni, esetleg csak többé-kevésbé meg lehet közelíteni. Ahhoz, hogy a különböző tesztváltozatok ekvivalensek legyenek, többek között az kell, hogy mindegyik teszt megközelítően ugyanazt a tágabb tudásterületet vizsgálja, a tesztekben hasonló arányban legyenek könnyű és nehéz feladatok.

Az ekvivalens tesztváltozatok szerkesztésekor az elkészített feladatokat arányosan osztjuk szét a különböző változatokba. Az arányosság többféle szempont egyidejű figyelembevételét igényli. Szükség van arra, hogy a tananyag elejéről, közepéről és végéről egyaránt kerüljenek feladatok mindegyik tesztváltozatba. Emellett a feladatoknak a bennük megtestesített tudás fontosságát és a nehézséget tekintve is arányosan kell megjelenniük az egyes változatokban.

Az ekvivalens tesztváltozatok készítése során másként járunk el a tudás ismeret-jellegű és képesség-jellegű elemeit mérő feladatokkal. Az ismeret-jellegű elemek mindig a maguk konkrétságában jelennek meg, egy adott tény vagy definíció mindig ugyanaz a tény vagy definíció lesz. Például a mohácsi csata évére többféleképpen rákérdezhetünk, többféle tesztfeladatot készíthetünk, a helyes megoldás mindig az 1526 lesz. Ezzel szemben a tudás képesség-jellegű összetevői sok, többé-kevésbé különböző konkrét tevékenységben megnyilvánulhatnak. Például a szöveges feladatok megoldásának tudását végtelenül sok pontosan azonos szerkezetű, de különböző szövegű és különböző számértékeket tartalmazó feladattal lehet megvizsgálni. Hasonlóképpen az írás- vagy olvasáskészség felmérésére is sokféle szöveget használhatunk, és joggal feltételezhetjük, hogy mindegyik nagy pontossággal ugyanazt méri. Feltételezhetjük, hogy ha egy tanuló egy fizika feladatot meg tud oldani, akkor meg tud oldani egy ettől csak számértékekben különböző másik feladatot is. Nem ennyire szükségszerű azonban az, hogy ha valaki ismer egy adott történelmi tényt, akkor egy másikat is ismerni fog. E különbségeket figyelembe véve a képesség-jellegű tudásnál mód van arra, hogy pontosan (vagy legalábbis nagyon jó közelítéssel) azonosan mérő feladatokat készítsünk, az ismeret-jellegű tudáselemekkel viszont csak annyit tehetünk, hogy azokat arányosan szétosztjuk a különböző tesztváltozatok között.

Az ekvivalens tesztváltozatok készítése a felsorolt alapelvek figyelembe vételével történhet, végső soron azonban csak többszöri kipróbálással, a feladatok cserélésével, javításával lehet jó közelítéssel azonosan mérő tesztváltozatokat készíteni, és ezek ekvivalenciáját csak tapasztalati úton lehet bizonyítani (Nagy, 1972).

3.3. A TESZTEK KIPRÓBÁLÁSA ÉS STANDARDIZÁLÁSA

A tesztek elkészítése nem ér véget azzal, hogy a feladatokat megfogalmazzuk, belőlük változatokat állítunk össze, ezeket oldalakra tördeljük, sokszorosításra alkalmas formában legépeljük. Szükség van arra is, hogy a gyakorlatban kipróbáljuk azokat, és a tapasztalatok alapján korrigáljuk az esetleges hibákat.

A kipróbálás és a korrekció során elvárható alaposság attól függ, hogy ki, milyen célokkal készítette a tesztek. A pedagógusok saját használatra szánt tesztjeiket ritkán próbálják ki, esetleg az ismételt alkalmazott tesztek az előző tapasztalatok alapján átdolgozzák. Ma már szinte minden középiskola, de sok általános iskola is rendelkezik akkora (vagy nagyobb) számítástechnikai kapacitással, mint amilyennel tíz évvel ezelőtt csak a kiemelt kutatóintézetek rendelkeztek. Nincs tehát akadálya annak, hogy a tanárok tesztjeiket kipróbálják és igényes elemzéseknek vessék alá.

Szélesebb körű felhasználásra szánt, szakértők által készített teszteknek mindenképpen át kell esniük a kipróbáláson, és az eredmények statisztikai elemzésén alapuló korrekción. Az igazán jó tesztek csak hosszabb fejlesztő folyamat eredményeként készülhetnek el. Ennek részleteivel a következő fejezetben foglalkozunk.

A tesztek kipróbálásán túl szükség van a végleges, tartós használatra szánt tesztek „bemérésére”, különböző paramétereik megállapítására. Ennek egy sajátos formája a tesztek standardizálása, melynek során azt vizsgáljuk meg (felmérjük, vagy statisztikai eszközökkel becsüljük), hogy a tesztelni szándékozott személyek teljes köre (a teljes populáció) milyen teljesítményeket ért el a teszten.

A tesztek standardizálása a norma-orientált teszteknel hozzátartozik a pontozási rendszer kialakításához, hiszen ebben az esetben az egyes személyek teljesítményeit a populáció normáihoz viszonyítjuk. Például az intelligencia-teszteknel a populáció átlaga lesz a 100 IQ pont, a szórásnak pedig (általában) 15 pontot feleltetnek meg.

A tudásszintmérő teszteknel más a standardizálás funkciója. Ezeknel a teszteknel ki lehet alakítani a csoport normáitól független értékelő rendszert is. Ilyen megoldás lehet például az, hogy az eredményeket a maximális pontszám százalékában adjuk meg, tehát az egyes tanulók eredményeit a csoport eredményeitől függetlenül is értelmezhetjük. A standardizálásnak ebben az esetben az a szerepe, hogy a pedagógusok illetve a tanulók munkájuk értékeléséhez külső szempontokat kapjanak, saját eredményeiket összehasonlíthassák az országos eredményekkel.

A standardizáláshoz országos reprezentatív mintát kell választani. Figyelembe véve a szükséges pontosságot, erre a célra tesztváltozatoként elegendő 300 körüli tanuló. A megfelelő módon kiválasztott tanulók megoldják a teszteket, majd eredményeik alapján kiszámítjuk a tesztek összpontszámának országos átlagát. Hasonlóképpen minden egyes itemre kiszámítjuk, hogy a tanulók hány százaléka oldotta meg helyesen. Ezeket az adatokat célszerű a tesztekkel együtt eljuttatni a pedagógusokhoz, így azok saját osztályuk, iskolájuk eredményeit az országos színvonalhoz viszonyíthatják.

A standardizálás lehetőséget adhat arra is, hogy a tanulók osztályozását országosan egységesítsük. Mivel a tudásszintmérő tesztek elkészítése a kritérium-orientált szemléletmódon alapszik, vagyis a tesztek a tananyagot képezik le, lehetne az osztályozást úgy is egységesíteni, hogy például minden teszten a 85% feletti teljesítményre adunk ötöst, 70% felett négyest stb. Ilyen értékelő rendszert alkalmaznak a megtanító stratégiák, ahol csak azok a tanulók haladhatnak tovább, akik a tananyagot egy előre meghatározott szinten elsajátították. (Többnyire csak kétféle minősítés van: elsajátította, nem sajátította el.) A mai iskolákban azonban a tanulók többsége nem éri el a teljes elsajátítás szintjét, hanem egy részük nagyon is hiányos tudással halad tovább. Az elsajátítás különbségei az osztályzatok különbségeiben jelennek meg. Az egyes témakörök elsajátításában is nagy különbségek lehetnek: egyik témakört a tanulók többsége jól elsajátítja, a másikat kevésbé. Az értékelés gyakorlatában azonban nagy hagyománya van annak a szemléletmódnak, amelyik a tanulókat egymáshoz viszonyítva osztályozza, a jelesek, jók, közepesek stb. egymáshoz viszonyított aránya nagyjából mindig ugyanolyan, függetlenül attól, hogy könnyű vagy nehéz tananyagról van szó, hogy az elsajátítás szintje átlagosan jó vagy gyenge.

Az iskola létező mechanizmusaira, elvárására reagál a *standard osztályzat*. Ez a százalékpontoknak egy olyan osztályzattá alakítási rendszere, amely az országos eredményeket figyelembe véve úgy jelöli ki az egyes érdemjegyek pontszám-határait, hogy országos szinten mindig a tanulók egy meghatározott hányada kapjon ötöst, négyest stb. A nehezebb teszteknel tehát a határok alacsonyabbra, a könnyebbeknél magasabbra kerülnek. (A standardizálást illetően I. Nagy, 1972; példák: Báthory, 1973).

4. TESZTELEMZÉS, TESZTFEJLESZTÉS

Jó tesztek csak hosszabb fejlesztés eredményeként lehet készíteni. Egy széleskörű használatra szánt tesztet elterjesztése előtt legalább kétszer ki kell próbálni ahhoz, hogy megfelelő színvonalú mérőeszközzé váljon. Az oktatásban huzamosan használt tesztek pedig érdemes időről időre felülvizsgálni, átdolgozni, továbbfejleszteni. A tesztelemzés és tesztfelkészítés ma már sokféle matematikai, számítástechnikai eszközt használhat. A következőkben ezekből csak a legismertebb, személyi számítógéppel is megvalósítható eljárásokat ismertetjük. (A standardizált témazáró tesztek fejlesztésével kapcsolatban I. Nagy, 1975).

4.1. A FELADATOK, ITEMEK JÓSÁGÁNAK VIZSGÁLATA

A feladatok, itemek jóságának elemzésére, a hibás itemek kiszűrésére csak két módszert mutatunk be, egyet a klasszikus tesztelméletre és egyet a kritérium-orientált tesztelméletre alapozott eljárásokból.

Ha feltételezzük, hogy a tesztünk homogén, vagyis minden feladat hozzávetőlegesen ugyanazt a tudást méri, akkor elvárhatjuk, hogy a feladatok eredmé-

neyi szorosan korreláljanak egymással. Ki kell tehát számítanunk a korrelációs mátrixot, azaz minden feladatnak minden más feladattal vett korrelációját. Ha olyan feladatot találunk, amelynek a korrelációi többnyire alacsonyak, akkor arra gyanakodhatunk, hogy az adott feladat nem azt a tudást méri, amit a többi feladat. Például a megoldásában nem azok az ismeretek, készségek dominálnak, amit az adott témakörben a tanulók megszereztek, hanem a másutt, korábban megszerzett tudás.

Szélsőséges esetekben előfordulhatnak a korrelációs mátrixban negatív értékek is. Ha sok a negatív korreláció, akkor szinte biztosra vehetjük, hogy az adott feladattal valami baj van. Ez ugyanis azt jelenti, hogy a jó tanulók (akik a feladatok többségét jól oldották meg, ezt a feladatot elrontották. Ilyen esetekben rendszerint rövid elemzéssel ki lehet deríteni, hogy a feladat rosszul van megfogalmazva, vagy a javítókulccsal van valami baj.

Ugyanezt az információt tömörebben tartalmazza a feladat-tesztösszpontszám korreláció. Ha egy feladat a teszt összpontszámával alacsonyan korrelál, az jelzi, hogy a megoldásához szükséges tudáselem idegen a témakör egészétől, a negatív korreláció pedig a feladat hibájára utal. Hasonlóan elemezhetjük a feladatok részeit, az itemeket is.

A feladatok korrelációs mátrixa és a feladat-összpontszám korrelációkban rejlő információk között vannak finom különbségek, amelyeket ugyancsak felhasználhatunk a hibásnak tűnő feladatok vizsgálatánál. A korrelációs mátrix egésze, sok korrelációs együttható együttesen ugyanis kialakít egy képet arról, hogy milyen magas a korrelációk többsége. Ha a korrelációs együtthatók többnyire magasak, az azt jelzi, hogy a teszt egésze homogén tudást vizsgál. Ebben az esetben egy feladat alacsonyabb együtthatói már jelezhetnek valamit. Ha azonban a korrelációk többsége alacsony, akkor a teszt inhomogén tudást vizsgál, és csak a többségnél sokkal kisebb, vagy kifejezetten negatív korrelációk utalnak hibás feladatra.

A tudásszintmérő tesztektől soha nem várhatjuk el, hogy teljesen homogének legyenek, vagyis minden itemjük ugyanazt a tudást mérje, ezért a klasszikus tesztelméletre alapozott elemzések csak korlátozott érvényűek lehetnek. Nem lenne helyes a feladatokat addig változtatgatni, amíg végül csupa magas korrelációt kapunk, ezzel ugyanis esetleg megfosztjuk a tesztet minden, az adott tananyagrészeire jellemző specifikumától, és csak olyan feladatok maradnak benne, amelyek megoldásához már csak valamelyik általános képességre van szükség.

Más jellegű technikákat javasol az itemek elemzésére a kritérium-orientált tesztelmélet. Az itemek jóságának megítélésénél a hangsúly inkább azon van, hogy mire akarjuk a tesztet használni, és arra a célra megfelele-e az adott item.

Mivel a tudásszintmérő tesztek tipikus alkalmazása az oktatási folyamat hatékonyabb irányítása, a kritérium-orientált tesztelmélet az elemzések többségét erre a célra dolgozta ki. Például a tesztelés célja az lehet, hogy megállapítsuk, a tanulók mennyit tanultak egy adott időszakban. Nem az tehát a célunk, hogy felmérjük, általában mit tudnak a szóbanforgó kérdésekről, hanem arra vagyunk kíváncsiak, hogy az adott tanítási periódus mennyire volt számukra hatékony. Ilyenkor azokat az itemeket akarjuk a tesztből kiszűrni, amelyek nem ilyen

tudást mérnek. Erre a célra felhasználhatjuk az oktatással szembeni érzékenység mutatóját.

Az *itemek oktatással szembeni érzékenysége* azt jelenti, hogy az adott item megoldásának esélyét hogyan befolyásolja az adott tanítási-tanulási periódus. Nyilvánvalóan azok az itemek, amelyeket a tanulók már az oktatás előtt is jól meg tudnak oldani, nem az adott időszakban megszerzett tudást mérik. Hasonlóképpen azok az itemek sem, amelyeket a tanulók az adott oktatás hatására sem tudnak megoldani.

E gondolatmenet felhasználásával megkonstruálhatjuk az oktatással szembeni érzékenység mutatóját úgy, hogy képezzük az item megoldásának oktatás előtti és oktatás utáni átlagát. E mutató kiszámításához a tesztet kétszer kell a tanulónak megírniuk: az adott témakör tanítása előtt és után. Ekkor 1-gyel értékelve minden jó és 0-val minden rossz megoldást, az itemek átlageredményei 0 és 1 közé fognak esni. Minden egyes itemre képezzük az oktatás előtti és az oktatás utáni eredmények különbségét. Az így kapott mutatók -1 és $+1$ között változhatnak. Minél nagyobb ez az érték, annál inkább az adott oktatás hatására elsajátított tudást méri az item. Legmagasabb értéke 1 lehet, ebben az esetben az oktatás megkezdése előtt senki sem tudta a feladatot megoldani, utána viszont mindenki. Akkor adódik 0, ha az oktatásnak nincs hatása az eredményekre, tehát előtte és utána pontosan ugyanannyian oldják meg jól az itemet. A negatív érték azt jelzi, hogy az oktatás után kevesebben oldották meg az itemet, mint előtte. Ezt (hacsak nem tételezzük fel az oktatás butító hatását) már mindig valamilyen hiba jelzésének kell tekintenünk.

Ha el akarjuk kerülni azt, hogy ugyanazzal a tanulócsoporttal kétszer megírassuk a tesztet, hasonló megfontolások alapján végezhetünk elemzéseket egy-egy, a szóbanforgó oktatásban részt vett és részt nem vett csoporttal. Bár gyakorlatilag nehéz lehet két egyenértékű, egy oktatott és egy nem oktatott csoportot találni, az oktatott és a nem oktatott csoport eredményeinek különbségei is jól jellemzik azt, hogy mely itemek mérik az oktatás hatására kialakult tudást.

Ez utóbbi elemzésekre a klasszikus megoldásokkal teljesen ellentétes hatású tesztfejlesztést alapozhatunk. Ha azokat az itemeket hagyjuk ki a tesztből, illetve javítjuk ki, amelyeknek alacsony az oktatással szembeni érzékenységét jellemző indexük, akkor a tesztünk egyre inkább azt a speciális tudást méri, amit az adott oktatási szakasz eredményeként szereztek meg a tanulók.

4.2. A RELIABILITÁSMUTATÓK KISZÁMÍTÁSA

A tesztfejlesztés egyik célja a tesztek megbízhatóságának, reliabilitásának javítása. Ezért a tesztfejlesztés során annak minden fázisában ki kell számítani a reliabilitásmutatókat, hogy képet kapjunk a tesztfeladatokon végzett változtatások hatásáról, hogy lássuk, mennyit kell még a feladatokon csiszolni, hogy megbízható tesztet kapjunk. Ezen kívül a tesztek felhasználóját is tájékoztatnunk kell arról, hogy a teszt milyen jól mér, mennyire bízhatunk meg az általa szolgáltatott eredményekben.

A reliabilitásmutatók kiszámítására sokféle formula áll rendelkezésünkre. Mint korábban már kifejtettük, ezek mindegyike alsó becslést ad a teszt reliabilitásának valódi értékére. A különböző formulák különböző értékeket adnak a reliabilitásra, de bármelyik formulával is számolunk, a reliabilitás valódi értéke mindegyiknél csak nagyobb vagy vele megegyező lehet.

A formulák között azt mérlegelve válogathatunk, hogy milyen adatok állnak rendelkezésünkre, és milyen eszközzel végezzük a számításokat. Kézi vagy zsebszámológéppel dolgozva választhatunk kevesebb munkát igénylő, de gyengébb becslést adó formulát. Számítógépes adatfeldolgozás esetén érdemes számításigényesebb, de jobb becslést adó formulát használni.

Ha a tesztfelezéses módszerrel akarjuk a teszt reliabilitását meghatározni, külön számítjuk ki a teszt páros és páratlan sorszámú itemjeinek az összpontszámát. Majd kiszámítjuk a két érték közötti korrelációs együtthatót. Ezt jelöljük r -rel. Az r felhasználásával a Sperman–Brown formula segítségével számíthatjuk ki a reliabilitásmutatót (ezt r_{tt} -vel jelöljük).

$$r_{tt} = \frac{2r}{(1+r)}$$

A tesztek belső konzisztenciájának jellemzésére szolgáló legszemléletesebb mutató a *Cronbach-féle alfa koefficiens*. Ez az itemek és a teszt varianciáiból indul ki, ha a szórások rendelkezésünkre állnak, kiszámítása viszonylag egyszerű.

$$\alpha = \left(\frac{n}{n-1} \right) \left(1 - \frac{\sum s_i^2}{s_t^2} \right)$$

ahol n az itemek számát, s_i az itemek szórását, s_t a tesztösszpontszám szórását jelöli.

Ha rendelkezésünkre állnak az itemek átlagai és az item-tesztösszpontszám korrelációk, használjuk a *Gulliksen formulát*.

$$r_{tt} = \left(\frac{n}{n-1} \right) \left(1 - \frac{\sum pq}{\sum r_{it}^2 / (pq)^2} \right)$$

Itt p az itemek megoldásának arányát jelöli, 1-gyel a jó, 0-val a rossz megoldásokat értékelve ez megegyezik az átlaggal. A $q=1-p$, az r_{it} pedig az item-tesztösszpontszám korreláció.

A *Kuder–Richardson 20 formula* (a szakirodalomban gyakran *KR20*-ként hivatkoznak rá) kiszámításához elegendő az itemek (súlyozás nélküli) átlagát és a tesztösszpontszám szórását ismerni. Az előző jelöléseket felhasználva:

$$r_{tt} = \left(\frac{n}{n-1} \right) \left(1 - \frac{\sum pq}{s_t^2} \right)$$

A tesztek reliabilitását a folyamatos fejlesztő munkával igen nagy mértékben javítani lehet. Már néhány hibás item kijavításával, kicserélésével is jelentős javulást lehet elérni.

A tanárok által készített tesztek reliabilitásmutatói ritkán haladják meg a 0,6 értéket. A gyakorlatban használt egyszerűbb tesztekől megkövetelhetjük, hogy reliabilitásmutatójuk legalább a 0,8 legyen. A 0,9 már a standardizált tesztek körében is elfogadható érték, a hosszas fejlesztő munka eredményeként elkészült tesztek reliabilitása pedig elérheti 0,95–0,98 értékeket is.

5. TUDÁSSZINTMÉRÉS ÉS PEDAGÓGIAI KUTATÁS

A tudásszintmérő tesztek nemcsak a mindennapi iskolai gyakorlatban játszanak szerepet, hanem a pedagógiai kutatásnak is rendkívül fontos, mással nem helyettesíthető eszközei. A korábban már elemzett tulajdonságokon (vagyis, hogy a tesztekkel objektív, érvényes és megbízható, számszerű adatokhoz jutunk) túl ezúttal érdemes arra is utalnunk, hogy a tesztelés eredményeit megőrizhetjük, a méréseket tetszőleges alkalommal megismételhetjük. Ez lehetővé teszi, hogy összehasonlítsunk egymással *időben és térben egymástól távoli teljesítményeket*, továbbá statisztikai módszerekkel elemezzük, hogy bizonyos feltételek milyen teljesítményeket eredményeznek.

Már a reprezentatív mintákon végzett tudásszintmérések korai szakaszában elemezték a *teljesítményeket különböző háttérváltozók szerinti bontásban*. Ilyen háttérváltozók például a tanulók családjára, a szülők foglalkozására, iskolázottságára, anyagi helyzetére, a tanulók lakóhelyére, az iskolára vonatkozó adatok. Ezekből az elemzésekből megismerhetjük, hogy mely feltételek mellett milyen átlagos teljesítményt nyújtanak a tanulók (l. például: Csáki–Nagy, 1976).

Egy adott ország vagy oktatási rendszer hatékonyságának (azaz az ott oktatott tanulók teljesítményeinek) változását *rendszeresen megismételt felmérésekkel* időben is követhetjük. A világ számos országában jöttek létre olyan központok, amelyek tesztfejlesztő, szolgáltató tevékenységük mellett a rendszeresen elvégzett felmérések eredményeit tudományos elemzésekre is felhasználják. A világ legnagyobb ilyen központja a Princetoni (USA) székhelyű Educational Testing Service, kiadványai évről évre rendszeresen elemzik, hogyan változik az USA tanulóinak teljesítménye az egyes iskolai tantárgyakban. Európában az egyik legismertebb szolgáltató-fejlesztő tesztközpont a hollandiai CITO.

Magyarországon az Értékelési Központ végzett hasonló, monitor típusú méréseket (Báthory és mtsai, 1983, 1985; Vári, 1989). Az első és mindezekig egyik legátfogóbb standardizált tesztrendszer pedig a JATE Pedagógiai Tanszéken készült a 70-es években (Standardizált ... 1972–1976), és ott alakult meg az 1991-ben az Alapműveltségi Vizsgaközpont. Az 1990-es évek elejére Magyarországon is kialakult és – elsősorban a személyi számítógépeknek köszönhetően – széles körben elterjedt a tesztkészítés és tesztfejlesztés kultúrája. A két kutatófejlesztő intézmény (Értékelési Központ, Alapműveltségi Vizsgaközpont) mellett elsősorban a megyei pedagógiai intézetek váltak a tesztkészítés és -fejlesztés helyi centrumaivá.

A sok országot átfogó tudásszintmérés többségére az IEA társaság (International Association for the Evaluation of Educational Achievement: az oktatási-nevelési eredmények értékelésének nemzetközi társasága) szervezésében került sor. Magyarország 1968 óta kilenc jelentősebb programban vett részt. Többek között ezek eredményeként vált világszerte ismertté a magyar tanulók nemzetközi viszonylatban is kiemelkedő teljesítménye a matematika és a természettudományok terén. (Az IEA vizsgálatok jelentőségét részletesen elemzi Báthory, 1992) Az IEA felmérések részletes eredményeit közreadó közel harminc vaskos kötet mellett rendszeresen megjelennek egyes speciális kérdésekre koncentráló másodelemzések is. Keeves (1992) könyve például a természettudományos tudás változását elemzi 1970 és 1984 között huszonhárom országban. A magyarországi IEA vizsgálatokat más tudásszintmérő vizsgálatokkal együtt mutatja be a Horánszky (1991) szerkesztésében megjelent tanulmánykötet. Míg a nemzetközi összehasonlító vizsgálatok és a rendszeresen megismételt országos reprezentatív felmérések révén elsősorban a rendszer szintű fejlesztésekhez (pl. tantervi fejlesztés), oktatáspolitikai döntések megalapozásához kapunk hasznos információt, az oktatás közvetlen, helyi, iskolai folyamatainak fejlesztését is hatékonyan segíti a *diagnosztikus értékelés* (Ingenkamp, 1985, Vidákovich 1987, 1990).

Gyakran kerül sor tesztek használatára a tudás valamely összefüggő, komplex területének, egy készség- vagy képességrendszer belső *szerkezetének* a feltárására. Nagy József (1980) az iskolakészültség részletes szerkezetének vizsgálatára dolgozott ki egy tesztrendszert, de tesztekkel végzett vizsgálat tárta fel a szövegesfeladat-megoldókészség (Csáki-Nagy, 1976), a fogalmazástechnika (Orosz, 1974a), a kombinatív képesség (Csapó, 1988) és a rajzkészség (Csapó és Varsányi, 1985) szerkezetét is. Ezek a vizsgálatok többnyire nem csak egy időpontban, statikusan mutatják be az adott képességek szerkezetét, hanem különböző életkorú tanulókkal végzett (ún. keresztmetszeti) mérések révén bemutatják a *fejlődés folyamatát* is. (A fejlődési folyamat felmérésére további példák: Nagy, 1971, 1973, Orosz, 1974b.)

A tudásszintmérés szinte mindig megjelenik mint a *kísérleti eredmények dokumentálásának eszköze* abban az esetben, ha azt vizsgálják, hogy valamilyen kísérleti hatás eredményeként hogyan változik meg a tudás valamely összetevője. Például egy új oktatási módszer hatékonyságát úgy tudjuk meghatározni (az általunk kítűzött célokhoz, vagy egy másik eljárásához, általában a hagyományos gyakorlathoz viszonyítva), hogy felmérjük az eredményeket. A legegyszerűbb esetben a kísérleti csoport eredményeit a kontrollcsoportéhoz viszonyítjuk. Ebben az esetben a kísérleti beavatkozás előtt felmérjük a tudás szintjét mind a kísérleti, mind pedig a kontrollcsoportban, így ellenőrizzük, hogy a kiinduláskor azonos-e a két csoport tudásszintje. Ezt követi az egyik csoportban a kísérleti beavatkozás, majd a záró szint felmérése mindkét csoportban. Kiszámítjuk a kezdő és a záró mérés különbségét a két csoportban, majd a különbségek különbsége lesz az az eredmény, amit a kísérleti beavatkozás hatásának tulajdoníthatunk (ilyen kísérleti elrendezésre példa: Csapó, 1991).

ELLENŐRZŐ KÉRDÉSEK

1. Mire szolgálnak a tesztelméletek?
2. Melyek a klasszikus tesztelmélet alapvető feltevései?
3. Hogyan jellemzik az itemeket a valószínűségi tesztelméletek?
4. A teszt mely tulajdonságát jellemzi az objektivitás?
5. Hogyan értelmezzük a tesztek reliabilitását?
6. Hogyan lehet biztosítani a tudásszintmérő tesztek validitását?
7. Mit jelent egy item nehézségi indexe, differenciáló ereje, reliabilitása?
8. Mit jelent az, hogy a tesztelés normaorientált vagy a kritériumorientált?
9. Melyek a feleletválasztó feladatok fontosabb típusai?
10. Melyek a feleletalkotó feladatok fontosabb típusai?
11. Milyen sajátosságai vannak a képesség-jellegű tudás mérésére szolgáló feladatoknak?
12. Mi a javítókulcs?
13. Mi a súlyozás?
14. Miért van szükség ekvivalens tesztváltozatokra?
15. Hogyan készülnek a tudásszintmérő tesztek ekvivalens változatai?
16. Mit jelent a standardizálás?
17. Melyek a tesztelezés fontosabb lépései?
18. Mi a tesztfelvezéses reliabilitásmutató-számítás alapelve, mi a Spearman-Brown formula?
19. Mit fejez ki a Cronbach-féle alfa koefficiens és hogyan lehet kiszámítani?
20. Mi a KR-20 formula, hogyan számítjuk ki?
21. Milyen jellegű pedagógiai kutatások alkalmazzák a tudásszintmérést?

FOGALMAK

Differenciáló erő: az item megkülönböztető képessége. Azt mutatja meg, hogy az item mennyire érzékeny a tanulók tudására, mennyire élesen tesz különbséget a különböző tudásszintű tanulók között.

Ekvivalens tesztváltozatok: olyan tesztek, amelyek különböző feladatokat tartalmaznak, de arányosan válogatva ugyanabból a feladathalmazból. Meghatározott valószínűséggel egymás helyettesítésére használhatók.

Érvényesség: 1. validitás.

Feleletalkotó feladatok: a választ a tanuló alkotja meg, nem pedig a felkínált lehetőségek közül választ.

Feleletválasztó feladatok: megoldásuk során a tanulók kész, előre megadott válaszok közül választják ki a megfelelőt (helyeset, jót, igazat stb.).

Item: a teszt legkisebb önállóan értékelhető eleme.

Javítókulcs: annak leírása, hogy milyen feltételek mellett lehet az egyes tesztitemek megoldását jónak elfogadni.

Jóságmutatók: a teszt jóságát kifejező adatok. A három leggyakrabban használt jóságmutató az objektivitás (tárgyszerűség), a reliabilitás (megbízhatóság) és a validitás (érvényesség).

Klasszikus tesztelmélet: az elsőként kidolgozott matematikai tesztelmélet. A tesztek megbízhatóságát a teszttel mért érték és a valódi érték összefüggésével jellemzi.

Kritériumorientált tesztelés: a tesztek nyerspontszámait előre rögzített külső kritériumokhoz (pl. a lehetséges maximumhoz) viszonyítjuk.

Megbízhatóság: 1. reliabilitás.

Nehézségi index: az item egyik jellemző tulajdonsága. Azt mutatja meg, hogy egy tetszőlegesen kiválasztott tanuló milyen valószínűséggel oldja meg az itemet.

Normaorientált tesztelés: a tesztek nyerspontszámait a megvizsgált minta adataihoz (pl. átlagához) viszonyítjuk. A normaorientált tesztek a mintához képest mutatják meg az egyes személyek tulajdonságait.

Nyerspont: a teszt felvétele során közvetlenül kapott pontszám. A nyerspontokból különböző transzformációkkal (pl. súlyozás vagy standardizálás) kifejezhető pontszámokat származtathatunk.

Összpontszám: a tesztet alkotó itemekre adott pontszámok összege.

Reliabilitás: a teszt megbízhatósága. Azt fejezi ki, hogy a teszt mennyire jól méri azt, amit mér.

Reliabilitási koefficiens: a tesztekkel mért tulajdonság és a tulajdonság valódi értéke közötti korrelációs együttható négyzete.

Standardizálás: a teszt felvétele statisztikailag reprezentatív mintán, majd az eredmények alapján a nyerspontok átszámítása standard skálára.

Szubteszt: a teszt valamely egységként kezelhető része. Itemek, feladatok elkülöníthető csoportja.

Tesztelméletek: a tesztek készítését és használatát megalapozó egységes matematikai modellek.

Tesztfeladat: egy gondolatmenettel megoldható, összetartozó itemek együttese.

Validitás: érvényesség, azt fejezi ki, hogy a teszt valóban azt méri-e, aminek a mérésére felhasználjuk.

Valószínűségi tesztelméletek: a tesztitemek megoldásának valószínűségét a tudás (képesség) függvényének tekintik. Az itemek tulajdonságait a valószínűségelmélet eszközeivel jellemzik.

IRODALOM

Ágoston György–Nagy József–Orosz Sándor (1979): Mérések módszerei a pedagógiában. Tankönyvkiadó, Budapest.

Báthory Zoltán (1973): 7 standardizált tantárgyvizsga. Országos Pedagógiai Intézet, Budapest.

Báthory Zoltán (1992): Tanulók, iskolák, különbségek. Tankönyvkiadó, Budapest.

Báthory Zoltán és mtsai (1983): Az iskolai nevelés néhány összetevőjének vizsgálata egy felmérés tükrében TOF–80 felmérés. Pedagógiai Szemle, 2. sz. 135–185. o.

Báthory Zoltán és mtsai (1985): Monitor típusú felmérések a közoktatás rendszerében. Az Értékelési Központ kiadványai 2. OPI–ÉK, Budapest.

- Berk, R. A.* (1980, szerk.): Criterion-referenced measurement: The state of the art
The Johns Hopkins Press Ltd., London.
- Bloom, B. S.* (1956): Taxonomy of Educational Objectives: Cognitive Domain
Mc. Kay, New York.
- Bloom, B. S., Hastings, J. T., és Madaus, G. F.* (1971): Handbook of Formative and Summative
Evaluation of Student Learning
McGraw Hill Co.
- Csáki Imre–Nagy József* (1976): Alsó tagozatos szöveges feladatbank
Acta Paedagogica, Ser. Spec., Szeged.
- Csapó Benő* (1987): A kritériumorientált értékelés
Magyar Pedagógia, 3. sz. 247–266. o.
- Csapó Benő* (1988): A kombinatív képesség struktúrája és fejlődése
Akadémiai Kiadó, Budapest.
- Csapó Benő* (1991): A gondolkodás műveleti képességeinek fejlesztése – A kísérlet eredményei
Új Pedagógiai Szemle, 4. sz. 31–40. o.
- Csapó Benő–Varsányi Zoltán* (1985): A rajzkészség fejlettségének vizsgálata középiskolai tanulóknál
Acta Paedagogica, Ser. Spec., Szeged.
- Fricke, R.* (1974): Kriteriumsorientierte Leistungsmessung
Verlag W. Kohlhammer, Stuttgart.
- Guilford, J. P.* (1965): Fundamental Statistics in Psychology and Education
McGraw Hill, New York.
- Horánszky Nándor* (1991): Jelzések az elsajátított műveltségről
Közoktatási Kutatások sorozat. Akadémiai Kiadó, Budapest.
- Horváth György* (1985): Tesztelmélet: problémák és perspektívák
Pszichológia, 1. sz. 53–78. o.
- Horváth György*: (1991): Az értelem mérése
Tankönyvkiadó, Budapest.
- Ingenkamp, K.* (1985): Lernbuch der Pädagogischen Diagnostik
Beltz Verlag, Weinheim und Basel.
- Joó András* (1980): A feladatképzítés kérdései
Országos Oktatástechnikai Központ, Veszprém.
- Kádárné Fülöp Judit–Joó András* (1977): Beszámoló a strukturális elemzés pedagógiai alkalmazásá-
nak néhány módszeréről
OPI dokumentumok 8. OPI, Budapest.
- Keeves, J. P.* (1992): Learning Science in a Changing World. Cross-national Studies of Science
Achievement: 1970 to 1984.
IEA, The Hague.
- Kozéki Béla* (1984): Személyiségfejlesztés az iskolában
Békés Megyei Pedagógiai Intézet, Békéscsaba.
- Lienert, G. A.* (1967): Testaufbau und Testanalyse
Julius Beltz, Weinheim und Berlin.
- Lord, F. M. and Novick, M. R.* (1968): Statistical Theories of Mental Test Scores
Addison-Wesley, Reading, Massachusetts.
- Nagy József* (1971): Az elemi számolási készségek
Tankönyvkiadó, Budapest.
- Nagy József* (1972): A témazáró tudásszintmérés gyakorlati kérdései
Tankönyvkiadó, Budapest.
- Nagy József* (1973): Alapműveleti számolási készségek
Acta Paedagogica, Ser. Spec., Szeged.
- Nagy József* (1975): A témazáró tesztek reliabilitása és validitása (STT 18. kötet)
Acta Univ. Szeg. de A. J. nom. Sectio Paed. Ser. Spec., Szeged.

- Nagy József* (1979): Köznevelés és rendszerszemlélet
Országos Oktatástechnikai Központ, Veszprém.
- Nagy József* (1980): 5–6 éves gyermekeink iskolakészültsége
Akadémiai Kiadó, Budapest.
- Nagy József* (1985): A tudástechnológia elméleti alapjai
Országos Oktatástechnikai Központ, Veszprém.
- Orosz Sándor* (1974a): A fogalmazástechnika mérésmetodikai problémái
Tankönyvkiadó, Budapest.
- Orosz Sándor* (1974b): A helyesírás fejlődése
Tankönyvkiadó, Budapest.
- Orosz Sándor* (1977): A tananyag elemzése
Országos Oktatástechnikai Központ, Budapest.
- Popham, W. J.* (1978): Criterion-referenced measurement
Prentice Hall, Englewood Cliffs, N. J.
- Rasch, G.* (1960): Probabilistic models for some intelligence and attainment tests
Kopenhagen.
- Rost, J.* (1988): Quantitative und qualitative probabilistische Testtheorie
Verlag Hans Huber, Bern–Stuttgart–Toronto.
- Srittmatter, P.* (1973, szerk.): Lernzielorientierte Leistungsmessung
Beltz, Weinheim.
- Standardizált Témazáró Tesztek (1972–1976)
17 kötet teszt az általános iskola felsőtagozatának tananyagához
Acta Univ. Szeg. de A. J. nom. Sectio Paed. Ser. Spec., Szeged.
- Thorndike, R. L.* (1971, ed.): Educational Measurement
American Council on Education, Washington.
- Vári Péter* (1989): A MONITOR '86 ismertetése
Pedagógiai Szemle, 12. sz. 1123–1130.
- Vidákovich Tibor* (1987): Innovatív célú diagnosztikus pedagógiai értékelés
Közoktatási Kutatások Titkársága, Budapest.
- Vidákovich Tibor* (1990): Diagnosztikus pedagógiai értékelés
Akadémiai Kiadó, Budapest.

IRODALOM

TARTALOM

Előszó	5
I. <i>A pedagógiai kutatás metodológiai kérdései</i>	7
1. fejezet <i>A pedagógiai kutatás metodológiai kérdései</i> (Falus Iván)	9
II. <i>A szakirodalom feltárása és feldolgozása</i>	35
2. fejezet <i>A szakirodalom tanulmányozása</i> (Falus Iván)	37
3. fejezet <i>Korszerű információkereső rendszerek a kutatás szolgálatában.</i> (Tóthné Környei Márta)	68
III. <i>Kutatási stratégiák</i>	87
4. fejezet <i>A pedagógiai kísérlet</i> (Bábosik István)	90
5. fejezet <i>Deduktív (analitikus) jellegű kutatások</i> (Szabolcs Éva) ...	106
IV. <i>A kutatás módszerei, eszközei</i>	123
6. fejezet <i>A megfigyelés</i> (Falus Iván)	125
7. fejezet <i>A kikérdezés</i> (Nádasi Mária)	171
8. fejezet <i>A szociometriai módszer</i> (Bábosik István)	212
9. fejezet <i>Pszichológiai tesztek</i> (Réthy Endréné)	235
10. fejezet <i>Tudásszintmérő tesztek</i> (Csapó Benő)	277
11. fejezet <i>Dokumentum és tartalomelemzés</i>	317
11/a fejezet <i>Dokumentumelemzés</i> (Nádasi Mária)	317
11/b fejezet <i>Tartalomelemzés</i> (Szabolcs Éva)	330
V. <i>Pedagógiai vizsgálatok leíró és matematikai statisztikai módszerei</i> ...	341
12. fejezet <i>A statisztikai módszerek pedagógiai alkalmazásának indokai, statisztikai alapfogalmak</i> (Nahalka István)	343
13. fejezet <i>Az adatok elemi vizsgálata.</i> (Nahalka István)	357
14. fejezet <i>Statisztikai összehasonlítások</i> (Nahalka István)	401
15. fejezet <i>A változók rendszerének struktúrája</i> (Nahalka István) ...	440
Mellékletek <i>a statisztikai fejezetekhez</i> (Nahalka István)	492
VI. <i>A kutatási eredmények értelmezése, közlése</i>	515
16. fejezet <i>A kutatási eredmények értelmezése, közlése</i> (Falus Iván) ...	517

