DOI 10.15773/EKE.HABIL.2018.008

Nahalka István

Ellentmondások a pedagógiai mérés és értékelés elméleteiben

Tartalom

Előszó

Hosszú ideig írtam ezt a könyvet. Írás közben végig azon töprengtem, hogy mi a megfelelő módja annak, hogy viszonylag nehéz matematikai megfontolásokat tárjak elsősorban pedagógia szakos bölcsészek, illetve többségükben nem matematika szakos pedagógusok elé. A könyv a mérésekről szól, szándékom szerint úgy, ahogyan a matematika tárgyalja e témát, miközben nem a matematika belügyéről, hanem a pedagógia (a pszichológia, és több más humán tudomány) egyik legfontosabb kérdéséről írok. Ha úgy akarnám megmondani, hogy miről szól ez a könyv, ahogy az a médiában szokásos, vagyis figyelemfelhívóan, blikkfangosan, akkor azt is mondhatnám: a mű azt taglalja, lehet-e mérni a pedagógiában, és arra jutok, hogy nem. De ez így nem fedné teljes mértékben a valóságot. Tény, hogy a matematika méréselméleti alaptételeiből kiindulva rendkívül súlyos problémákat írok le a pedagógiai (és pszichológiai,…) mérésekkel kapcsolatban. De megoldásokat is felvázolok, igaz, ezek közül a reményt keltőbbek nagyrészt ma még csak szűk körben ismertek.

Mielőtt bárki azt hinné, hogy e területen abszolút originális kutatást végeztem, és mint első szakember bemutatom a világnak, hogy mi a baj azzal, amit mérések címszóval a kutatásainkban és napi gyakorlatunkban teszünk, sietek leszögezni, hogy a könyv egy nagyon széles, hazánkon kívül jól ismert szakirodalmi bázist használ. Bár néhány saját eredményt is közlök az itt következő lapokon, de ennél fontosabb volt számomra, hogy bemutassam a hazai olvasóknak, szakembereknek, hogy milyen fogalmi alapokra épül a mérés, milyen követelményekkel kell szembesülniük a társadalmi, emberi folyamatokkal kapcsolatban méréseket tervezőknek.

A könyvben bemutatom a méréselméletek (nem a mérés!) történetét. A 19. század utolsó harmadában kezdődő történet azért nagyon érdekes, mert nagymértékben a pszichológiai mérések szükségletei formálták. Ennek az az oka, hogy mindenekelőtt a humán tudományok, s ezen belül is elsősorban a pszichológia művelése során vetődtek fel olyan problémák, amelyek időről időre újabb matematikai megfontolásokat igényeltek. Erre nem sok példa akad a tudományok történetében. Az új matematikai területek feltárását e tudomány saját fejlődési logikáján túl elsősorban a természettudományok ösztönzik, a humán tudományok a matematikai elméleteket többnyire csak használják. A méréselméletek esetében azonban e könyv írásakor már kb. 80-90 éve a pszichológia sajátos problémái szolgáltak hajtómotorként a matematikai elgondolások fejlődéséhez. Ezért is tanulságos számunkra a méréselméletek története.

Bemutatom azokat az ismeretelméleti problémákat, amelyek a mérések értelmezéséhez feltétlenül átgondolandók, ha le akarjuk küzdeni a fogalmi eredetű nehézségeket. A tudományokban az ismeretelméleti alapok áttekintése nem pusztán valamilyen függeléke a téma tárgyalásának, nem a filozófusok kenyéradó foglalatossága, hanem az egyik legfontosabb „eszköz” a kutatók kezében, hogy fogalmilag következetes rendszert alkossanak. Az ismeretelméleti kérdések alapos vizsgálata fontos összefüggésekre mutat rá. Többek között segít eligazodni a modellek világában, a választott ismeretelméleti paradigma tereli a kutató gondolkodását abszolút szakmai kérdésekben is, és figyelmeztet a gondolkodási, logikai hibákra, ha elvesznénk a rengetegben. Az ismeretelméleti kérdések különösen izgalmasak egy olyan területen – mint a humán méréselméleti problémák is –, ahol a megfontolások elméleti, paradigmatikus keretei a „legkeményebb” empirizmustól a radikális konstruktivizmusig terjednek, ahol a valóság megismeréséhez való viszony szélsőségesen különböző tudományos attitűdjei jutnak szerephez.

Bemutatom a méréselmélet ma leginkább elfogadott paradigmáját, a reprezentációs elméletet, és ez ennek a könyvnek az elsődleges célja. A reprezentációs elmélet akár szálka is lehetne a humán tudományokkal foglalkozók szemében, mert ebben szerepelnek azok a mérésekkel szembeni, matematikailag megalapozott elvárások, amelyeket a mai mérések a mi pedagógiai és pszichológiai kutatásainkban, fejlesztéseinkben és napi gyakorlatunkban igen sok esetben nem teljesítenek. Ma a reprezentációs méréselmélethez képest is erősebb követelményeket megfogalmazó alternatívákat ismerünk csak, bajban van az a tudomány, amely még a reprezentációs méréselméletben megfogalmazott elvárásokat is képtelen teljesíteni (vagy pontosabban: gyakorlatának egy része nem teljesíti).

Tárgyalom a klasszikus tesztelméletet, amely ma még mindig a pedagógiai, pszichológiai kutatások meghatározó „szerszámosládája”. Elsősorban a kritika dominál majd a könyv e fejezetében. Igyekszem kimutatni, hogy a klasszikus tesztelmélet egyáltalán nem teljesíti a reprezentációs méréselméletben megfogalmazott követelményeket. Keményen fogalmazva: ez az elemzés – ha reális, de annak tűnik, hiszen a nemzetközi szakirodalomban is szereplő alapelveken nyugszik – bizonytalanná teszi a pszichológia és a pedagógia méréseken alapuló számtalan eredményét.

Magam fontos fejezetnek gondolom a modern tesztelméletről szólót, mert olyan elméleti alapjait írom le elsősorban a legegyszerűbb Rasch-modellnek, amelyek ilyen tárgyalásban – tudtommal – hazai publikációban még nem kerültek ismertetésre. Bár még „receptet” is nyújtok a modern tesztelméleti modellek alkalmazásához, célom mégsem egy kézikönyv összeállítása volt, nem állt szándékomban a modern tesztelméleti módszerek ismertetése. Ennek ellenére remélem sikerült úgy tálalnom e fontos témát, hogy az segítheti a további, részletesebb, már a technikára is az itteninél jobban figyelő ismerkedést és alkalmazást. A modern tesztelmélet problémáit is bemutatom, azt, hogy miközben matematikai szempontból már megfelelő, teljesíti a reprezentációs méréselméletben megfogalmazott igényeket, eközben egy jelentős korláttal is számolnunk kell: a modern tesztelméleti modellek nem alkalmasak összetettebb, vagyis például a pedagógiát jobban érdeklő pszichikus rendszerek vizsgálatára.

Tudtommal először jelenik meg Magyarországon ismertetés az együttes mérés (conjoint measurement) elméletéről. A pszichológiában (és a pedagógiában) ma még szinte egyáltalán nem alkalmazott gondolkodásmód és módszeregyüttes egyrészt arra való, hogy elméleti háttérül szolgáljon, s ezzel mintegy legitimálja a modern tesztelméletet, másrészt viszont alkalmas korrekt, tehát a reprezentációs méréselméletnek megfelelő mérések koncipiálására. Az együttes mérés azonban ugyanazokkal a korlátokkal rendelkezik, mint a modern tesztelmélet.

A könyv legtöbb fejezete negatív eredményeket sorol. Az utolsó részben azonban összegyűjtöm azokat a pozitívumokat, amelyek lehetőségeket teremtenek a reprezentációs méréselméletnek megfelelő mérésekhez. A klasszikus tesztelmélet problémái abból fakadnak, hogy az elmélet nem képes olyan struktúrák fogalmi alapjainak lerakására, amelyek lehetővé tennék, hogy magas szintű, nagyon hatékony statisztikai vizsgálatokat végezzünk a mért adatokkal. Sokan ismerik a terminust, bevezető kutatásmódszertani kurzusok anyagát is alkotja: nem tudunk intervallumskálákat létrehozni. De ha megelégszünk ordinális skálák kialakításával, akkor egyes vizsgálatainkat méréselméleti szempontból korrektté tehetjük. Ha egyszerűbb pszichikus struktúrákat vizsgálunk, akkor pedig a modern tesztelmélet alkalmazásával még intervallumskálák is létrehozhatók, és bevethetjük a statisztikai nagyágyúkat (regressziószámítás, faktoranalízis, klaszterelemzés, stb.). Ugyanez érvényes az együttes mérés elméletére is.

És van itt valami, ami igazán reményt keltő a pedagógiai kutatók és a pszichológusok számára: ha lemondunk arról, hogy képességek fejlettségét, attitűdöket, és számtalan más pszichológiai konstruktumot egyetlen számmal jellemezzünk, és megnyitjuk a kaput az összetettebb matematikai struktúrák alkalmazása előtt (pl. gráfok, hálók), akkor nagyon izgalmas lehetőségek tárulnak fel. A tudás struktúrájának kutatása intenzíven zajlik a világban, egyik produktuma, a tudástér elmélet (Knowledge Space Theory) fejlesztése már gyakorlati alkalmazásokat is eredményezett. Annyira jelentős fejlemény, hogy ezen írás keretei közé már nem is fért, csak jelzem a mű végén, miről is van szó e teória esetében.

Végül a köszönetnyilvánítás. Köszönöm volt munkahelyem, az ELTE Pedagógiai és Pszichológiai Kar Neveléstudományi Intézete oktatóinak, kutatóinak, hogy egy korábbi változatában egy intézeti szakmai vita, majd számtalan beszélgetés keretében kifejtették véleményüket az elgondolásaimról, és biztattak a könyv megírására. Hallgatóimnak, akiknek – ahogy a kurzusok tematikái ezt lehetővé tették – igyekeztem gondolatokat megjeleníteni a könyv anyagából, köszönöm, hogy türelmesek voltak, hajlandók voltak együtt gondolkodni velem, és elviselték, hogy ezek az alkalmak is hozzájáruljanak ahhoz, hogy az én fejemben is rendeződjenek a gondolatok.

A szerző
Csömör, 2018. június

1. Bevezetés

Ahogy a pedagógia is „egyre inkább tudománnyá válik”, s ahogy egyre erősebb az igény, hogy tételei kellően megalapozottak legyenek, úgy válik egyre fontosabbá a pedagógiai mérések státuszának, mondanivalójának kérdése, egyáltalán: a pedagógiai mérések elvégezhetőségének problémája. A kérdés minduntalan felmerül a szakmai és szakmapolitikai vitákban. Egyre erősebbé válik sokakban a meggyőződés, hogy a pedagógia csak akkor válhat hiteles tudománnyá, olyanná, amely megfelelően tudja segíteni a nevelés gyakorlatát, ha a kutatásainak legtöbbje empirikus vizsgálódás, *objektív méréseken alapszik*, érvényes és megbízható mérőeszközöket alkalmaz az objektív pedagógiai jelenségek objektív leképezésére. Ezt a tudományos attitűdöt gyakran jellemzik úgy, hogy szemben áll egy inkább a *filozófiai-történeti megközelítéseket*, illetve a *kvalitatív kutatás* módszereit előnyben részesítő kutatási beállítódással. Talán senki nincs, aki a két véglet valamelyikét „tisztán” képviselné, természetesen működik az „értelmiségi hozzáállás”, hogy tudniillik a kettő egyensúlyára van szükség, mégis létezik egy látens vita, egy szembenállás az inkább az egyik attitűdöt és az inkább a másikat preferáló szakemberek tábora között.

Talán e ponton lenne a legjobb megmagyarázni a kétségkívül bizarr címet: *„Mit mondanak a vajákosok?”*. Magam részese voltam egy olyan szakmapolitikai vitának, amelyben mintha az előbb csak futólag jellemzett két irány, két eltérő szakmai attitűd körül alakult volna ki érzelmektől sem mentes polémia. A konkrét vita során kétségtelenül erősebbnek bizonyult a pedagógiai kutatásokat korrektnek, objektívnek mondott mérésekre alapozni szándékozók, vagy pontosabban magukat ide sorolók tábora, amely rendkívül éles kritikát fogalmazott meg a „másik oldallal” szemben, vagyis azokkal szemben, akik sokkal erősebben bíznak a történeti, filozófiai megalapozásban, a kvalitatív kutatásokban.

Hogy egészen pontos legyek, egy alaposabb értelmezés szerint a vita foglalata nem is e két attitűd ellentéte volt. Az objektív méréseket favorizáló „oldal” által kialakított képben – ahogy a másik „oldalról” gondolkodott – valahogy összemosódott kétféle tudományos magatartás, kétféle minőség. Az egyik az elméletibb hajlam, a történetiséget, a filozófiai megalapozást előtérbe állító megközelítés, a másik viszont maga a *színvonaltalanság*, a nemzetközi tudományosságtól távol álló, a nem éppen a magas szintű – értsd: a nemzetközi tudományos színtéren kiválónak tartott – tudományos eredményeket felmutatni képes kutatóknak a tevékenysége. Úgy tűnt, a tudományos színvonal áll szemben a színvonaltalansággal. Nos, ebben a vitában – annak egyik túlfűtött szakaszában – hangzott el ezzel a feltételezetten a „színvonaltalanságot reprezentáló” táborral kapcsolatban, hogy képviselőinek munkája, amit „tudományosnak” lehetne csúfolni, valójában béljóslás, sarlatánság, vajákosság, hiszen annyira távol áll a korrekt tudományos igazolásokat alkalmazó, korszerű, és a nemzetközi standardoknak is jól megfelelő, kvantitatív, empirikus alapú kutatásoktól.

Innen vettem a címbe a „vajákosság” szót. Nem írom le részletesen, mi volt ennek a vitának a kimenetele, talán nem is volt eredmény. De azt jelezni szeretném, hogy miközben természetesen egyetértek azzal, hogy az empirikus alapokon szerveződő tudományosság óriási értékeket tud felvonultatni, eközben nem szabad elfeledkezni arról, hogy *az empirikus tudomány is csak elméletekbe ágyazottan működhet*. Ez az egész könyv valójában erről szól. Azt szeretném megmutatni, hogy amikor empirikus kutatásokról, jelesül pedagógiai mérésekről van szó, és lényegében vakon megyünk, no, nem az orrunk, hanem az adatok után, akkor nagyon is szükség van azokra, akiket egy egyszerűsítő szemléletmód egy kalap alá vesz az igazi vajákosokkal, de akik valójában egy nagyon magas szintű tudományosságot képviselnek. Olyat, ami nélkül az empirikus vizsgálódások, a képesség- vagy az attitűdmérések, és minden hasonló „igazán hitelesnek tartott" tudományos tevékenység egyszerűen lehetetlen. Szeretném megmutatni ebben a könyvben, hogy *a mérések abszolutizálására való törekvés, az elméleti megalapozás lekicsinylése, az elméleti alapon felvethető kérdéseknek az ignorálása súlyos problémákhoz vezet*.

Túl a tudományos kutatások irányainak, értékrendjének kijelölésén, a kérdést azért is érdemes alaposabban megvizsgálni, mert a pedagógia gyakorlatában ma a mérés fontosabb szerepet játszik, mint eddig bármikor. Talán nem volt még a magyar oktatáspolitikai közéletben olyan téma, amely gyakrabban lett volna szakmai diskurzusok tárgya a 2000-et követő időkben, mint a *PISA vizsgálatok*. De gondolhatunk az *Országos kompetenciamérésre* (OKM) is, és még számtalan olyan mérésre, amelyek legfeljebb csak másodlagosan szolgáltak tudományos kutatások tárgyaiként, elsődleges funkciójuk az országok iskolarendszerei bizonyos területeken mutatott teljesítőképessége felmérésének, illetve összehasonlításának segítése volt.

Mindehhez tegyük hozzá a tanítási-tanulási folyamatok során formálódóban lévő, *új értékelési kultúrát*. A sokszor becslésnek mondott, a tanulók teljesítményének alapvetően csak kvalitatív megítélését lehetővé tevő (vagy még arra sem jó), de mindenképpen sok szubjektivitással terhelt osztályozás további alkalmazása mellett ma már nem is nagyon lehet felhozni érveket. Át kell alakulniuk a pedagógiai értékelési folyamatoknak, s ebben az egyik nagyon fontos területként értékelődik a korrekt, bemért, standardokkal ellátott tesztek alkalmazása. A tesztekkel végzett értékelő munka már mérés – legalábbis a ma uralkodó általános tudományos megítélés szerint.

A mérés tehát kiemelt kérdéssé vált az utóbbi időben Magyarországon a pedagógiában. Ehhez képest a folyamat elméleti hátterével, alapvető problémáival vajmi keveset foglalkoztunk. (És most persze provinciális szemléletről teszek tanúbizonyságot, hiszen az állításom csak a hazai tudományosságra vonatkozik.) Teljesen természetes a számunkra, hogy legyen szó tudományos kutatásról, vagy tanulói (iskolai) teljesítmények felméréséről (akár egy „nagy vizsgálat" részeként, akár a hétköznapi pedagógiai munka, a pedagógiai értékelés eszközeként), bármely esetben alkalmazhatunk megfelelő mérőeszközöket, rendszerint teszteket. Elég összeállítani néhány odaillőnek látszó kérdést, feladatot, be kell gyűjteni a gyerekektől a válaszokat és a megoldásokat, ezeket adatokká kell konvertálni, beírni a megfelelő statisztikai programba, és alkalmazni néhány jól kidolgozott algoritmust az adatok elemzésére, majd az eredmények értelmezésére. Mintha ezzel a tevékenységgel kapcsolatban semmilyen probléma nem merülne fel, legfeljebb olyan technikai jellegű gondok, amelyek megoldása nem igényel túl nagy erőfeszítést, amely gondok az úgynevezett tesztfejlesztés keretében valóban megoldhatók. Pedig amióta létezik pszichológiai mérés (a pedagógiai mérések nagy része pszichológiai mérés, vagy teljes mértékben követi azok logikáját), azóta vitatott a státusa, és egyáltalán nem mondhatjuk, hogy a tudománynak sikerült minden alapvető problémát megoldania. Nyilván nem véletlen, hogy azok a folyóiratok, amelyek kifejezetten a mérések témájával foglalkoznak (pl. a *Measurement*, a *Psychometrika*, vagy a *Journal of Mathematical Psychology*), de gyakorta más orgánumok is, nem is beszélve a tanulmánykötetekről és a monográfiákról, időről időre teret szentelnek a témának, és felvetnek alapkérdéseket is (ld. pl. Tendler 2009; Rossi 2007; Borsboom 2005; Narens 2002a; Barrett 2000, 2003, 2008; Borsboom és mts. 2003; Michell 1999, 2000, 2005; Mari

1997; Luce 1996; Luce és Narens 1994; Goldstein 1980). Hiszen nincs minden kérdés megnyugtatóan lezárva, talán soha nem is lesz.

Mégis, milyen problémákról van szó? Azért is kell nagyon tisztességesen válaszolni erre a kérdésre, mert a problémák nem nyilvánvalók. A mérési gyakorlat a felszínen szinte problémamentes. Ma Magyarországon a széles tudományos közösségben szakmai szempontból leginkább kiérleltként nyilvántartott, a tudományos kutatásokkal szemben támasztott, nemzetközi szakmai követelményeknek leginkább megfelelő kutatások, vagyis a kompetenciák, képességek, készségek kutatása (elsősorban a *Szegedi Egyetem Neveléstudományi Tanszékén*) töretlenül halad előre a *Nagy József* által még évtizedekkel ezelőtt kijelölt úton, és sikert sikerre halmoz. A tanulmánykötetekben, folyóiratokban, konferenciákon megjelenő munkák középpontjában az emberi kognitív és szociális működések hátterét alkotó pszichikus rendszerek vizsgálata áll, a módszerek közt alapvető szerepet szánva a tesztekkel való mérésnek. Lenyűgöző módszertani apparátus áll rendelkezésre, a tesztelemzés, tesztértékelés, tesztfejlesztés legmodernebb eljárásai, az alkalmazott statisztikai módszerek közt jelentős szerepet kapnak a legnagyobb hatásúak, lélegzetelállító eredményeket felmutatni képes masinéria kezeli a tesztelés során nyert, sokszor igen nagy mintáktól származó adatokat. *Lehet itt egyáltalán probléma*?

Vagy itt vannak a nemzetközi- és a hazai „nagy vizsgálatok" (PISA, IALS, TIMSS, PIRLS, OKM, stb.). A hatalmas adatrendszerek kezelésére alkalmas szoftverek jóvoltából ömlenek az információk, ezek értő kezelésével olyan tudást szerezhetünk az iskolarendszerünkről, amilyenről mondjuk a 20. század végén is még csak álmodoztunk. Megvalósulhat, illetve lényegében már megvalósult egy rendkívül fontos terv: mérhetjük a pedagógiai hozzáadott értéket, szinte beláthatatlan, mennyi jótéteménye lehet ennek a fejleménynek. *Lehet itt egyáltalán probléma*?

De említhetjük az iskolai értékelési kultúra átalakítását is. Éppen kikecmeregni akarunk egy olyan világból, amelyben az osztályozás nevű, már régen túlhaladottá vált metódus uralkodott az értékelésen, és felfedezzük a hiteles, jól bemért tesztek egyéni pedagógusi szinten való alkalmazásának előnyeit. Ezzel lényegesen megnöveljük az egyik fontos pedagógiai tevékenység objektivitását, mintegy „rendbe tesszük" a területet, szakmailag hitelessé, sokkal biztosabb tudományos alapokon állóvá tesszük. *Lehet itt egyáltalán probléma*?

Lehet. Szerintem van is. Méghozzá *súlyos problémák* vannak. A tesztek, a statisztikák, a számítógépes programok csinnadrattái elnyomják a kritikus kérdéseket. Hajlamosak vagyunk az impozáns eszközök láttán – észlelve bizonyos tudományos metodológiai szabályrendszerek érvényesítésében a következetességet – azt hinni, hogy itt semmilyen komolyabb, elvi jellegű problémáról nem lehet szó. Ebben az írásban én a mellett szeretnék érvelni, hogy az itt szóba kerülő mérések

1. ismeretelméleti szempontból egy irányban, és a kutatások során egyre kevésbé elfogadottá váló irányban elkötelezettek, és ezért ismeretelméleti státuszuk jelentős mértékben problematikussá vált,
2. nem tették minden esetben világossá a viszonyukat a tudományos mérések alapjait rögzítő méréselmélethez,
3. egyes mérési eljárások (hazánkban a konkrét kutatásokban és egyéb, pedagógiai tevékenységekben zajló mérések döntő többségéről van szó) nem is felelnek meg a ma e téren a tudományos közösségben mércének tekintett reprezentációs méréselmélet követelményeinek,
4. azokban az esetekben sem történik meg a méréselméleti megalapozás, amelyekben ez megtehető lenne,

5. rendkívül szűk azoknak a képességeknek a köre, amelyekben méréselméleti értelemben korrekt mérések végrehajthatók, az összetett, és ezért a pedagógiát inkább érdeklő képességek esetén egyáltalán nem létezik méréselméleti megalapozás,
6. a képességmérések nem teszik világossá a „mit mérünk?” kérdésére adható választ,
7. a mérések más, a tudományos, vagy oktatásfejlesztési céloktól eltérő szándékok teljesülését is szolgálják.

Ezek nagyon kemény szavak. Már előre szeretném jelezni, hogy elemzésem elsősorban a tudományos kutatások mérésekkel kapcsolatos mozzanataival, valamint a korrektnek mondott iskolai mérési gyakorlattal összefüggésben vet fel súlyos problémákat, miközben arra az eredményre jutok, hogy a „nagy mérésekkel” kapcsolatban (PISA, OKM, stb.), vagyis a nagyon szélesen értelmezett „felkészültséget” mérő vizsgálatokkal összefüggésben kisebbek (de ott sem elhanyagolhatók) a gondok.

Méréselmélet

A mérésről általánosan

A pedagógiai mérések is olyan mérések, vagy legalábbis azt szeretnénk, ha olyanok lennének, mint a más tudományokban és gyakorlati cselekvési formákban alkalmazottak. Mások persze azok a tulajdonságok, amelyeket a pedagógiában mérések segítségével szeretnénk jellemezni, egy sor sajátosság fakad a terület szakmai tartalmából. A szigorúság, egzaktság iránti igény azonban a vágyak szintjén ugyanaz: szeretnénk, ha *az összefüggések kvantifikálását* segítenék a mérések. Jó lenne, ha hasonlítanának a hosszúság, a tömeg, vagy a térfogat méréséhez. Megpróbálom kibontani, hogy ez a vágy milyen konkrét igényekben fogalmazható meg.

A mérés a legtöbb ember számára szinte hétköznapi tevékenység, legalábbis bizonyos mérésekkel szinte mindenki kapcsolatba kerül. Talán a legtöbbször alkalmazott mérés, amit naponta sokszor elvégzünk, az idő mérése. Ha vásárolunk, gyakran kerül szóba a termék mennyisége, súlya, hossza, térfogata. 102 cm-es képátlójú televíziót szeretnénk vásárolni, az ünnepi ebédhez legalább 80 dkg mennyiségű pulykahúsra van szükségünk, új kerítést szeretnénk barkácsolni a nyaraló telekre, ki kellene számolni, hogy ehhez mennyi vasra, fára, festékre van szükség, ám ehhez mérni kell – ezek a példák is azt mutatják, hogy mértékekkel, mértékegységekkel, és gyakran konkrét méréssel is lépten, nyomon kapcsolatba kerülünk.

Azon azonban biztosan nagyon kevesen szoktak elgondolkodni, hogy vajon min alapszanak az ilyen feladatok részét jelentő mérések. A hétköznapi méréseinkkel kapcsolatban nem találkozunk olyan problémákkal, amelyek alaposabb átgondolást igényelnének. Rendkívül adaptív, igencsak gyakorlatias tudásunk van a legegyszerűbb mérésekkel összefüggésben, a legtöbben e tudásukat ügyesen tudják alkalmazni.

Van természetesen a mérésekről alkotható tudásnak egy olyan szintje, amelyet elsősorban a természet-, és műszaki tudományokkal, a mérnöki feladatokkal foglalkozók használnak. Ők olyan mennyiségek mérését, így elsősorban a hosszúság, az idő, a tömeg, a felszín, a térfogat mérését, amelyek a hétköznapi tevékenységek során is gyakran előfordulnak, a hétköznapi igényekhez képest sokkal pontosabban, és részben ezért sokkal rafináltabb eszközökkel kell, hogy elvégezzék. De ők mérnek olyasmit is, amit mi, „egyszerű halandók" általában nem szoktunk, például radioaktivitást, elektromos feszültséget, reakciósebességet, pH-t, és hadd ne soroljak további, sokak számára egyáltalán nem is ismert mennyiségeket. Ezek a szakmai tevékenységek természetesen a hétköznapiaknál sokkal összetettebbek, alaposabb tudást igényelnek. A természettudományok története azt mutatja, hogy e mérésekkel, a mérések elvével kapcsolatban is felmerültek mélyebb, akár filozófiainak is nevezhető problémák. E problémák megoldására elképzelések születtek, amelyek körül nagyszerű viták zajlottak. Mégis jogosnak érzem azt a megállapítást, hogy alapvető, egész nagy tevékenységi területek munkáját megkérdőjelező problémákkal a természettudományi kutatások és a mérnöki gyakorlat területén nem kell számolnunk.

Nem így van ez a pedagógiával és a pszichológiával. A pszichológia önálló tudománnyá válása idején, vagyis a 19. század második felében éppen a *mérhetőség*, az emberi mentális működésekkel összefüggő folyamatokban az egyes tulajdonságokkal kapcsolatos *kvantitatív értelmezések lehetősége* volt az egyik fő hajtóerő, amelyet az új tudomány képviselői felhasználtak. A korabeli elképzelések szerint – pontosabban azok szerint, akik az „igazi tudomány” egzaktsága legfontosabb zálogának a tudomány által vizsgált tulajdonságok kvantitatívvá, mérhetővé tételét tartották – csak az a diszciplína tarthat számot a tudomány megjelölésre, amely ezen a folyamaton végig megy, *fogalmaihoz mérhető mennyiségeket kapcsol, és törvényeit mértékek, változók közti összefüggésekként képes megfogalmazni*. Ezzel megnyílt egy máig tartó, rendkívül izgalmas tudományos fejlődés, amely egyáltalán nem csak sikerek sorozata, és még ma is a kutató tevékenység alapjait érintő, nem lezárt fejezet a tudományok történetében. Olyan a tudománytörténetből többségében jól ismert folyamatok tartoznak ide, mint a pszichofizika felvirágzása, az értelmi képességek mérésére kidolgozott technikák formálódásának több, mint 100 éves története, az intelligenciával kapcsolatos, egyáltalán nem csak tudományos viták, a pszichometria szinte önálló tudománnyá válása, egy „háttéripar” kialakulása a tesztelés feladataival kapcsolatban, a gyerekek iskolai szegregálását is gyakran szolgáló mérési eljárások kidolgozása, a klasszikus- és a modern tesztelmélet konstrukciója és gyakorlati alkalmazásai, az együttes mérés (conjoint measurement) elméletének megszületése, a pedagógiában a képességfejlesztés alapvető céllá válásával együtt a képességek fejlettsége diagnosztizálásának feladata, és a sor természetesen nem zárható le. Mély tudományos felfedezések, konstrukciók, alapvető kérdéseken, sokszor definíciókon zajló, mára sem lezárt viták, a társadalmi demokráciát mélyen befolyásoló folyamatok, politikai csatározások – igazán nem mondhatja senki, hogy a tudományok fejlődésének ez a jelenségegyüttese unalmas lenne.

Vagyis fontos kérdésről van szó. Minden ilyen esetben mi jut eszébe először egy bölcsésznek? Hát az, hogy nézzük meg a téma történetét. És valóban, én, mint „átvedlett bölcsész” (eredetileg matematika-fizika szakos tanári diplomával rendelkezem), szintén hasznosnak tartom, hogy megvizsgáljuk, hogyan alakult ki a pszichológiában és a pedagógiában a mérés. Látjuk majd, hogy itt kifejezetten *méréselméletekről* van szó, sőt, a méréselméletek világa – ezzel most biztos sokakat elijesztek – *matematika*. A méréselméletek – már a legelső próbálkozásokra is igaz ez, vagyis közelítőleg *Hermann von Helmholtz* és teljes mértékben *Otto Hölder* elméleteire – matematikai axiómarendszerek voltak, s a maiak is azok. Ezért van az, hogy e könyv nagy része vegyesen pedagógia, pszichológia és matematika. Mivel olvasóim – várhatóan – elsősorban bölcsészek lesznek, igyekszem a legalaposabban megmagyarázni a matematikai részleteket.

Miről szólnak a méréselméletek?

Mint elvileg bárminek, a mérésnek is lehet, és természetesen van is elmélete, sőt, van több elmélete. Ezek alapvető tartalmukat tekintve matematikai alkotások, axiomatikusan felépített teóriák, olyan *struktúrák* leírásai, amelyekben nem arról van szó, hogy konkrétan mit mérünk, hanem arról, hogy maga a mérés mit jelent egy elvont értelemben.

Az elemzés során többször használok majd analógiákat a természettudományok köréből. Ez biztos sokak számára ellenszenves eljárás. Magam sem gondolom, hogy az embert körülvevő természet megismerésének tudományos metodológiája „egy az egyben” minta a humán tudományok számára is. Viszont jó „kapaszkodónak” tartom: mint tájékozódási pontot, vagy ötletadó kiindulási helyet, esetleg, mint bizonyos értékek megfogalmazását segítő háttértudást érdemes felhasználni azt, amit tudunk arról, hogy a természettudományok mit produkáltak a mérési folyamatokkal, azok értelmezésével kapcsolatban. Érdemes azt is figyelembe vennünk, hogy a pszichológiának (ha pontosabban akarok fogalmazni, akkor azt kell mondanom, hogy számos pszichológusnak) törekvése, ideálja, hogy a pszichológia egyre inkább hasonlítson a természettudományokhoz, sőt, egyesek szerint *a pszichológia valójában természettudomány*. És e törekvés éppen a pszichológiai mérések kérdésében nyilvánul meg a legerőteljesebben. Ahogy már említettem, a pedagógiai mérések – kis túlzással – valójában pszichológiai mérések, a képességek fejlettségi szintjének meghatározása, amire e könyvben koncentrálni szeretnék, egyértelműen. Mindezek talán kellőképpen indokolják, hogy érdemes a kifejtés során használni a természettudományi analógiákat.

A méréselméletek formálódásának történetével foglalkozó kutatók e történetet különböző módokon írják le. Ha a publikációkból ki szeretnénk olvasni egy egységes klasszifikációt a méréselméletekkel összefüggésben, akkor nincs is nagyon nehéz dolgunk. A történeti munkákban a legtöbbször idézett tudósok *Hermann von Helmholtz, Otto Hölder, Bertrand Russell, Percy Williams Bridgman, Norman Robert Campbell, Stanley Smith Stevens, Patrick Suppes, Duncan Luce, Louis Narens, Jean-Claude Falmagne és Jean-Paul Doignon* tettek sokat azért, hogy a méréselmélet, mint matematikai terület mára elérje azt a fejlettségi szintet, amely lehetővé teszi, hogy egy-egy tudomány a sajátos problémáinak megvilágítására használja a felhalmozott tudást.

Miről szólnak a méréselméletek? Arról, hogy miről szólnak a mérések. Minden mérés – itt még nincs különbség az elméletek között – matematikai objektumoknak „dolgokhoz" (tárgyakhoz, fizikai testekhez vagy anyagi rendszerekhez, emberekhez, folyamatokhoz, jelenségekhez, tulajdonságok értékeihez stb.) való hozzárendelését jelenti. Könnyen leírtam ezt a mondatot, pedig a méréselméleti kérdésekben kirobbant egyik legérdekesebb vita éppen ahhoz kapcsolódik, vajon definiálhatjuk-e a mérést úgy, hogy az nem más, mint *matematikai objektumoknak bizonyos „dolgokhoz" való hozzárendelése*. A helyzet ugyanis nem ilyen egyszerű. Olyannyira, hogy valójában ilyen definíciót a szakirodalomban nem is találunk. Kis gondolkodás után rájöhetünk miért: matematikai objektumokat, mondjuk számokat egy adott vizsgálódási terület objektumaihoz, jelenségeihez, stb. tetszőleges módon hozzárendelhetünk. Biztos, hogy nem lehet akármilyen hozzárendelés mérés. Vagyis amit itt a mérés első megközelítéseként leírtam, az nem definíció. Egy kis kiegészítéssel azonban olyan definícióvá válik, amely a méréselméletek történetében fontos szerepet játszott, sőt, játszik még ma is az erről szóló diskurzusban. *Stanley Stevens* ugyanis – később még más vonatkozásokban is fontos szerephez jut a történet leírásában ez az amerikai pszichológus – nagyon hasonlóan definiálja a mérést, azonban a hozzárendelés tényét megtoldja azzal, hogy a matematikai objektumok hozzárendelésének valamilyen *szabályok* szerint kell történnie (Stevens 1946). Szándékosan ilyen határozatlan a fogalmazás, hiszen *Stevens* csak egyetlen „szabályt" zár ki a hozzárendelések tekintetében, ez az az eset, amikor nincs szabály, vagyis véletlenszerű a hozzárendelés. Amúgy bármilyen szabály megfelelhet.

Joel Michell – akiről szintén sok szó esik még ebben az írásban – élesen bírálja *Stevens* definícióját (Michell 1999). *Michell* szerint a hozzárendelés szabályai közt csak olyanok fogadhatók el, amelyek megfelelnek a mért tulajdonság *kvantitatív jellegének*. Az utóbbi jelentése: a mérés során azt állapítjuk meg, hogy egy „dolog" adott tulajdonságának értéke egy jól definiált értelemben hányszorosa az egységnyi tulajdonságúnak választott „dolog" tulajdonsága értékének. Egy házfal hossza, a hossz mértékszáma az a szám, amely azt mondja meg, hogy a házfal hossza hányszorosa az egységnyinek választott (centiméteresnek nevezett) hosszúságnak. Az a mennyiség kvantifikálható, amellyel összefüggésben definiálható egység, és meghatározható, hogy az egyes tulajdonság értékek hányszorosai ennek az egységnek. *Stevens* definíciója – így *Michell* – semmilyen megkötést nem tartalmaz a hozzárendelésre (túl a random hozzárendelés kizárásán), nem alkalmas kvantitatív tulajdonságok jellemzésére, nem lehet a mérés hiteles definíciója.

A matematikai objektumok, amelyeket a mért „dolgokhoz" rendelünk hozzá, rendszerint számok, esetleg számokból felépülő összetettebb struktúrák elemei (pl. vektorok, függvények, mátrixok), itt nem sértjük jelentős mértékben az általánosságot, ha számokról beszélünk majd. Számokat rendelünk hozzá a testek (pl. autók) által megtett utakhoz, az eltelt időkhöz, a testek tömegéhez, számokat rendelünk hozzá az előítéletesség értékéhez, egy gyermek induktív gondolkodási képességének fejlettségéhez, egy állítás igaz volta elfogadásának értékéhez. Nyilván az a kérdés, hogy *ez a bizonyos hozzárendelés hogyan, milyen szabályok szerint történik, milyen típusai vannak*. Erről szólnak a különböző méréselméletek, s természetesen a válaszaik többé vagy kevésbé különbözők. A méréselméletek formálódásának története van, és ez a történet is, mint oly sok minden másé, alapos vizsgálódások tárgya volt. Nézzük meg közelebbről!

A formálódás szakasza: klasszikus méréselmélet

Egy nem túl régen született munka, egy spanyol kutató, *José Díez* tanulmánya (1997) a *formálódás* és az *érett méréselmélet* egymás utáni szakaszaiban igyekszik leírni a történetet. A formálódás szakaszában, amely nagyjából a 20. század közepéig, kicsit pontosabban *Patrick Suppes* alapvető munkájának megjelenéséig tartott (Suppes 1951)[1], a témában kutatók kidolgozták a részleteket. Ezt követte az érett elmélet fejlődésének időszaka, amely az előző korszak két nagy irányzatának szintéziseként is értelmezhető.

A kialakulás, a formálódás folyamatát két, egymással nagyrészt párhuzamosan futó fejlődési folyamat jellemzi. *Helmholtz, Campbell* és *Hölder* elméletei határozzák meg az egyiket, és magam úgy jellemezném, hogy inkább a természettudományos mérések elméleti hátterének megvilágításáról szól. A másik tradíció elsősorban *Stevens* nevével és az általa konceptualizált skálatípusokkal jellemezhető (Stevens 1946), és talán joggal mondhatjuk, hogy a pszichológia tudományának „harcáról" van benne szó. Arról a harcról, amit ez a tudomány, illetve általában a társadalmi, emberi jelenségek mérésével foglalkozó diszciplínák vívnak az egzaktságra törekvés elismeréséért, mérési módszereik legitimálásáért. *Díez* szerint a 20. század közepén, ahogy mondtam, elsősorban *Suppes* tanulmányának megjelenésével, létrejön a szintézis, és megformálódik az egységes matematikai elmélet, amely elvileg arra hivatott, hogy mind a természet, mind a társadalom (az ember) vizsgálatában a szükségessé váló mérések elméleti hátterét adja.

Ha nem is mindegyik esetben megyek bele a technikai részletek ismertetésébe, és nem is tárgyalok minden egyes méréselméleti megfontolást, igyekszem áttekinteni ezt a folyamatot, követve lényegében *Díez* gondolatmenetét. Megjegyzem, hogy a méréselméleti szakirodalomban ez a gondolatmenet, vagyis a méréselméletek történetének ilyen felfogása széles körben elfogadottnak tekinthető, nagyon hasonlóan írja le a fejleményeket *Joel Michell* (1990, 1999) és *Duncan Luce* (1987, 1992, 1996).

A történet 1887-ben kezdődött, amikor megjelent *Helmholtz* alapvető tanulmánya. Már ekkor világosak az alapok: a mérés hozzárendelés, egy függvénykapcsolat, amely a mért dolgokat és a hozzájuk rendelt matematikai objektumokat (mint mondtam, a legtöbbször számokat) köti össze. Mérni természetesen már az ókori kultúrákban is mértek, illetve a természettudományok felgyorsult fejlődésében már 200-250 évvel

[1] Több forrás nem ezt a korábbi publikációt tekinti az egységes megközelítés kialakulása időpontjának, hanem egy későbbi, *Patrick Suppes* és *Dana Scott* által írt tanulmány megjelenésének évét (1958). Elképzelhető, hogy Suppes eredeti tanulmánya 1951-ből – mivel egy angol nyelvű, de portugál matematikai folyóiratban jelent meg – kevésbé volt ismert a témával foglalkozók előtt, mint ez a második, amely egy valószínűleg sokkal olvasottabb folyóiratban, a *Journal of Simbolic Logic* hasábjain jelent meg.

ezelőtt is fontos, sőt, alapvető szerepet játszott a mérés, azonban a 19. század végén e tevékenység *elméleti alapjai* kerültek középpontba. Ha úgy tetszik, a feladat a már sokféle formában gyakorolt, bizonyos értelemben nagyon jól ismert emberi tevékenység, a mérés mélyebb, rejtettebb alapjainak, belső összefüggéseinek feltárása és egyfajta formalizálása volt. A kiindulópont ebben a hozzárendelés, de nyilván adódik a következő kérdés, vajon mi szerint, milyen szabályoknak megfelelően történik.

Helmholtz mutatott rá először, hogy a mérés során a „dolgoknak" adott tulajdonság alapján történő összehasonlítására kerül sor. Általában azt tudjuk, hogy bizonyos tulajdonságát tekintve az egyik „dolog" *nagyobb értékkel* rendelkezik, mint a másik „dolog", vagy esetleg éppen *egyformák* ebből a szempontból. Rudak hosszát, vagy a térben elhelyezkedő pontpárok távolságát (az általuk meghatározott szakaszok hosszát) lehet például ilyen módon összehasonlítani. *Helmholtz* az ugyanakkora értékből indul ki, tehát az „egyformaságot" tekinti egyik alapfogalomnak. A másik alapfogalom a „dolgok" *összetétele*, egyesítése, mint ahogy a rudakat egymás folytatásaként (egy egyeneshez illeszkedő módon) helyezhetjük el. Vagy a tömegmérésnél a testeket együtt tesszük a mérleg egyik serpenyőjére, vagy ahogy egy időintervallum befejeződésekor, ugyanabban az időpontban kezdődik egy következő, és a két időintervallum egyesítését vizsgáljuk. Vagyis itt az „összeadhatóság", más, precízebb kifejezéssel a *konkatenálhatóság*, a *konkatenáció* lesz a fontos. *Helmholtz* célja az volt, hogy bemutassa, hogyan válik a mért értékek, vagyis a mértékek összeadhatósága meghatározó tényezővé a mérést jelentő hozzárendelés megalkotásakor.

Nem követjük egészen pontosan *Helmholtz* gondolatmenetét, hiszen abban vannak olyan elemek, amelyek ma már másképpen szerepelnek a korszerűbb leírásokban. Mindenesetre a tulajdonságok értékeinek *azonossága*, valamint *rendezettsége* (ugyanolyan hosszú, hosszabb, rövidebb, ugyanolyan tömegű, nagyobb-, kisebb tömegű, ugyanolyan-, nagyobb-, kisebb időtartamú, stb.), továbbá a *mérendő értékeknek az összeadódása* játszik alapvető szerepet ebben a legelsőnek tekinthető méréselméletben. Az összeadhatóság (additivitás) lehetővé teszi, hogy önkényesen kijelölve egy „dolgot", ami esetében az adott tulajdonság mértékét egységnyinek tekintjük, addig egyesítsünk egymással ilyen egységnyi mértékű „dolgokat", amíg nem kapunk egy a mért „dologgal" megegyező tulajdonságút, ahogyan az empirikusan definiált. Ez a bonyolult fogalmazás, mondjuk, a hosszúság mérése esetén azt jelenti, hogy eleve rendelkezünk egy eszközzel, a mércével, amelyen már sok-sok, milliméteres hosszúságú „dolog" (mondjuk így: rudacska, vagy kis darabja a szalagnak) össze van rakva, folytatólagosan össze van illesztve, segítve ezzel a munkánkat. Hogy a leírt művelet valóban elvégezhető, azt a rudak egymás folytatásaként (egy egyenesben) való elhelyezhetősége, az „egyenlő hosszú" reláció fennállásának empirikus meghatározhatósága biztosítja. És ez jelöli ki a hozzárendelés módját: úgy rendelünk hozzá a hossz kezdetben kvalitatív értékéhez számokat, hogy ez a bizonyos összerakás, egyesítés, annak eredménye pontosan tükröződjék a hozzárendelt számok közötti viszonyokban. Pontosabban: ha az a rúd és a b rúd egymáshoz (egy egyenesben) illesztésével a c rúddal egyező hosszúságú rudat kapunk, akkor az a rúdhoz rendelt hosszúságmérték, a $h(a)$, valamint a b-hez rendelt hosszúságmérték, a $h(b)$ összege egyezzék meg a c-hez rendelt hosszúságmértékkel, a $h(c)$-vel. Ez a művelet az egységnyi hosszúságúnak kijelölt rudakkal is elvégezhető, így közvetlenül meghatározhatjuk (egy általam is később ismertetendő eljárás segítségével tetszőleges pontossággal) a rudak hosszát, vagy két pont távolságát (ami persze fogalmilag ugyanaz).

Szinte hallom az alaposabb matematikai tudással rendelkezők felszisszenését. Természetesen ez a leírás igencsak pontatlan, hogy ne mondjam: pongyola. De az eljárásnak van precíz, axiomatikus megfogalmazása is, ami *Höldertől* származik (1901).

Filozófiai szempontból fontos, hogy a *klasszikus méréselmélet* képviselői számára a *számok* azonosulnak a mértékek közti *arányokkal*. Például a mindenki által jól ismert π, amelynek konkrét értéke egy végtelen, nem szakaszos tizedes tört (irracionális szám), a 3,14…, nem más, mint egy körvonal hosszának és ugyanazon kör átmérője hosszának a hányadosa, egymáshoz viszonyított aránya. Amikor tudományos kutatást folytatunk, akkor a „dolgok" közti relációkat azért fordíthatjuk le számok közötti összefüggésekre, mert maguk a számok és a köztük lévő viszonyok – a klasszikus méréselméleti felfogás szerint – „ott vannak bent a dolgokban", s a kutatás során megmutatják magukat. Így a mérhetőség feltétele az, hogy a mért mennyiség additív legyen, egységet tudjunk kijelölni, és meg tudjuk állapítani valamilyen alkalmas empirikus eljárással, hogy a mért „dolog" vizsgált tulajdonsága hányszorosa az egységnyi mértékkel rendelkező „dolog" tulajdonságának. *A mérés során a „dolgokban", egymás közötti viszonyaikban megmutatkozó számokat olvassuk ki*.

A klasszikus méréselmélet elsősorban a természettudományok számára koncipiálta a mérést, de a pszichológia fejlődésének kezdetén a pszichofizikai mérések számára is útmutatást jelentett (Michell 1999, Fechner 1860). Ám nem bizonyult kellően általánosnak, hiszen már a természettudományok egyes fogalmainak megfelelő tulajdonságok mérésére sem nyújt elméleti hátteret. Kétségtelen, hogy a természettudományok legfontosabb mennyiségei additívak, illetve egész mértékegységrendszereket lehet felépíteni ezekre a fogalmakra, mennyiségekre. A hossz, az idő, a tömeg, az elektromos töltés fogalmaihoz köthető mennyiségek additívak, vagy ahogy a fizika nevezi őket: extenzívek. Ha anyagi rendszereket megfelelő módon egyesítünk, akkor az e fogalmakkal leírt tulajdonságaik értékei összeadódnak, és ez lehetővé teszi, hogy a mértékeket is összeadjuk, ha ki akarjuk számítani az egyesített rendszer tulajdonságához tartozó mértéket. Nem véletlen, hogy a későbbiekben, a *Helmholtz*, *Hölder* és *Campbell* által megalapozott, tehát az extenzív fizikai mennyiségek mérésére vonatkozó eljárásokat általánosító méréselméletet *extenzív mérésnek* nevezik el (Luce és Narens 1981).

Ugyanakkor már a fizikában sem minden mennyiség összeadódó, vagyis extenzív. És ezt mindenki tudja, aki egy minimális szinten képes a körülötte zajló eseményeket tudományos fogalmakkal is értelmezni. Mindenki tudja, bár nem biztos, hogy tudatosította is, hogy például a hőmérséklet nem additív, nem összeadódó mennyiség. Ha két anyagi rendszert egyesítünk, akkor a keletkező rendszer hőmérséklete nem a két eredeti rendszer hőmérsékleteinek összegével lesz azonos, hanem egy kiegyenlítődési folyamat játszódik le, s valahol a két érték között kapjuk a hőmérsékletet. Szemben a klasszikus viccel, ha egy szobában a levegő hőmérséklete csak 15 fok, és odakint 5 fok van, akkor nem érhetjük el a 20 fokot úgy, hogy kinyitva az ablakot, „beengedjük" az 5 fokot. Ha egy pohárban 30 fokos víz van, egy másikban 50 fokos, és összeöntjük őket, akkor ugyan nagyon sok kisgyerek gondolja úgy, hogy 80 fokos vizet kapunk (a részleteket ld. pl. itt: Nahalka 2002), de ezt később „kinőjük", és a felnőttek közül valószínűleg kevesen válaszolnának így egy tesztkérdésre.

Az úgynevezett intenzív, tehát a kiegyenlítődő mennyiségek „kibújnak" a klasszikus méréselmélet hatálya alól, nem érvényes rájuk az összeadódás, az additivitás. Így a klasszikus méréselmélet kiegészítésre szorult, és kialakult a természettudományokban a *származtatott mennyiségek mérésének elve* (Campbell 1920). E szerint a fizikai mennyiségek a természet törvényei által meghatározott összefüggésekben állnak egymással, egy mennyiség lehet más mennyiségeknek a függvénye. Kijelölhetünk úgynevezett alap mennyiségeket, és minden más fizikai mennyiség meghatározható a konkrét szituációkban mért értékeik függvényeként. Az alap mennyiségek pedig éppen az additív mennyiségek (a legfontosabbak: hosszúság, idő, tömeg), tanuljuk őket az iskolában, a hozzájuk tartozó, nemzetközileg egységesített alap mértékegységek rendszerét (SI) ismerjük, és hétköznapjainkban alkalmazzuk is.

A formálódás szakasza: a másik vonulat, a skálák osztályozása

A 20. század harmincas éveiben kapott megbízást a *British Association for the Advancement of Science*[2] egyik bizottsága, hogy vizsgálja meg *a pszichológiai mérések validitásának* kérdését. Pontosabban azt a problémát vetették fel, hogy az ebben a korban rendelkezésre álló méréselméleti elvárásoknak mennyire felel meg mindaz, ami a pszichológiában tudományos mérés néven zajlik. A bizottságnak *Campbell* is tagja volt, és elsősorban ő szorgalmazta, hogy a testület lesújtó véleményt fogalmazzon meg. A verdikt lényege, hogy a pszichológiai méréseknek mondott eljárások nem felelnek meg a tudományos igényeknek, mert a pszichológiai jelenségek világában nem léteznek additív mennyiségek (British Association for the Advancement of Science 1933). A bizottság elsősorban a hallásintenzitás mérésével kapcsolatos kifogásait tárta a tudományos közvélemény elé. Az egyik tag megfogalmazta, hogy a hallási érzéklet és az inger erőssége közti kvantitatív összefüggés nem csak, hogy hibás, hanem értelmetlen mindaddig beszélni róla, amíg az érzékletre vonatkozóan nem értelmeztük az összeadódás fogalmát. *Stevens* az Egyesült Államokban éppen a hallás kutatásával foglalkozott pszichológusként, és nyilván mélyen érintette a tekintélyes bizottság álláspontja. A ma is ismert, és kutatásmódszertani alap kurzusokon részleteiben is tanított mérési skálák elméletét ő fogalmazta meg 1946-ban kiadott művében. Ahogy *Díez* leírja (Díez 1997, 179. o.): az alapvető összefüggések már a 30-as években *Stevens* rendelkezésére álltak. Elmélete egyfajta válasz kívánt lenni a bizottság véleményére, és a kérdésfelvetés, ami a középpontjában áll, számunkra is rendkívül fontos.

Stevens elsősorban azt kérdezi, hogy lehet-e a tudományosnak tekinthető mérések körét az additív-, illetve a velük függvénykapcsolatban álló származtatott mennyiségek mérésére szűkíteni. A válasza természetesen az, hogy nem, *a klasszikus méréselmélet túl szűkre szabja a mérhető mennyiségek körét*.

Számunkra azért kiemelkedően fontos ez a kérdésfelvetés, mert az emberi kognitív képességek fejlettségének mérése pontosan az itt jelzett problémával küzd. Tudjuk, hogy a 20. század 30-as éveiben a pszichometria, itt még inkább csak az intelligencia mérése, előrehaladott gyakorlatnak volt tekinthető, miközben a tudósok egy körében alapvetően megkérdőjeleződött, hogy az ilyen méréseknek egyáltalán van-e létjogosultságuk. Ha a bennünket még közelebbről érdeklő, a képességek mérését szolgáló gyakorlatot nézzük, akkor világos, hogy a bizottság véleményében megfogalmazott kritika alapvetően vonatkozik erre a területre is. *A képességek tekintetében a közvetlen szemléleten alapuló ismereteink szerint nincs semmiféle összeadódás*. Szó sincs arról, hogy egy ember egy adott képessége fejlettségének mértéke viszonylag egyszerűen belátható módon valamilyen egység többszöröse lenne. És olyan pszichológiai mennyiségeket sem ismerünk, amelyek additívak lennének, és a képességek fejlettségének mértékei ezekkel valamilyen függvénykapcsolatban állnának, ezért mai ismereteink szerint ezek a mennyiségek

[2] Az 1831-ben alakult szervezet a brit tudósok egyfajta elit társaságaként működött, évente tartott találkozókkal, azokon szekciókban aktuális tudományos kérdések megvitatásával.

leszármaztatott mennyiségekként sem értelmezhetők[3]. Vagyis a *Stevens* által kialakított megközelítésre a pszichológiának, az emberi képességek mérésének igencsak nagy szüksége volt.

Stevens számára – ahogy ez korábban már szerepelt – minden olyan eljárás mérés, amelyben valamilyen empirikus entitásokhoz (objektumokhoz, folyamatokhoz, jelenségekhez, viszonyokhoz) rendelünk hozzá számokat valamilyen szabályok szerint. Bár *Stevens* munkáját a későbbiekben sok kritika érte, különösen *Joel Michell* ír kemény szavakkal a mérések ilyetén való definiálásával szemben (Michell 1999), illetve *Paul Velleman* és *Leland Wilkinson* (1993) kritizálják élesen a konstrukciót, ennek ellenére azt kell mondanunk, hogy *Stevens* skálái fontos előrelépést jelentettek a méréselmélet fejlődésében. *Stevens* kimondja ugyanis, hogy amikor az empirikus entitásokhoz hozzárendelünk számokat, akkor egy szoros kapcsolatnak kell lenni az empirikus entitásokon elvégzett műveletek és a nekik megfelelő, már a számokon végrehajtott műveletek között. Ezek között csak az egyik lehet az, hogy két empirikus rendszer alkalmas egyesítése (összefűzése, konkatenációja) megfelel az egyesítésre szánt rendszerekhez rendelt számok összeadásának, úgy, hogy az eredeti rendszerekhez rendelt számok összege megegyezik az egyesített rendszerhez rendelt számmal. *Stevens* igyekszik számba venni a további lehetőségeket, és így jut el öt skálatípushoz (mi négyet szoktunk emlegetni). Más megfogalmazásban még több skálatípust említ, és említenek mások is (ld. pl. Díez 1997, 182. o.). A későbbi kutatások egy általános keretbe helyezve megmutatták, hogy bár némi módosítások szükségesek, de *Stevens* lényegében helyesen írta le a skálákat (Narens 1981a, b, Luce és Narens 1985, Luce 1996).

A mai kutatásmódszertan kurzusokon elsősorban szereplő *nominális*-, *ordinális*-, *intervallum*- és *arányskálák* definiálása sokféleképpen történhet. Most bemutatok egy a hazai szakirodalomban ritkábban szereplőt. Forrásként sok művet említhetnék, de itt elég az eddig is több vonatkozásban „zsinórmértékként" használt *Díez* (1997) írásra hagyatkoznunk.

Skálák

A skálák definiálása során abból indulunk ki, hogy egy konkrét mérés meghatározása során szemügyre vesszük, hogy a mért empirikus entitások halmazán milyen *relációk* érvényesülnek. E relációk alapvető szerephez jutnak az elméletben, ezért a fogalmat érdemes „jól megrágni".

A reláció fogalma, bár a hétköznapi értelmezésre épül, itt meglehetősen elvont jelentést kap – mint oly sok más fogalommal ez megesett a matematikában. De induljunk ki a hétköznapi jelentésből! A reláció hétköznapi értelemben – és ez a matematikai jelentés lényege is – valamifajta viszonyt jelent két-, vagy több „valami" között. Ha a hosszúság példájánál maradunk, akkor az „egyenlő hosszú" viszony egy reláció, amely mondjuk két rúd közt áll fenn (vagy nem áll fenn). Megtehetjük, hogy a rudakból párokat alkotunk, és megvizsgáljuk, hogy melyek azok a párok, amelyek esetén a pár tagjai egyenlő hosszúak. E párok esetén a reláció fennáll, az összes többi esetében viszont nem (az összes többi pár esetében valamelyik rúd hosszabb). Vagyis a rudakból álló párokat most egy új halmaznak tekintve, e halmaz egy részhalmazát alkotják azok a párok, amelyek esetén fennáll az „egyenlő hosszú" reláció. Ez azt jelenti, hogy bátran definiálhatjuk magát a

[3] E kijelentések esetén nagyon óvatosan kellett fogalmaznom. Majd látjuk a modern tesztelmélet részleteinek ismertetésénél, hogy konstruálható a képességfejlettségek halmazán összeadási (konkatenációs) művelet. Az azonban igaz, hogy ez a művelet szemléletességben közelébe sem érhet a hossz, a tömeg, az idő mérésénél szerepet játszó konkatenációhoz, nem közvetlenül, hanem csak összetett matematikai lépések segítségével értelmezhető.

relációt úgy, hogy az nem más, mint *a rendezett párokból alkotott halmaz egy részhalmaza*. Az a részhalmaza – konkrét példánkban – amelynek az elemei esetén a párokat alkotó rudak egyforma hosszúak. Párok alkotásával vizsgálhatók más relációk is, a rudak esetén a „hosszabb", „rövidebb", de bármilyen összehasonlítás a vizsgált entitások körében ilyen relációt eredményez, pl. „melegebb", „azonos tömegű", „fejlettebb", „szebb", „ugyanolyan fontos", hogy – az utóbbiak esetében – mondjak egy-két extrém példát is.

A relációk fogalma általánosítható, és megfelelő matematikai fogalmak használatával kellően precízzé, kellően absztrakttá tehető. A példáink kétváltozós relációk voltak, de vannak több változós, vagyis kettőnél több elemet összekapcsoló relációk is. Az a, a b, és a c rudak ebben a sorrendben a „toldás" relációban állnak egymással, ha igaz, hogy a b rudat az a rúd folytatásaként, egy egyenes mentén elhelyezve az így keletkező egyesített rúd ugyanolyan hosszú, mint a c. A lehetséges „rúdhármasok" közt vannak olyanok, amelyekre ez az összefüggés áll, és vannak olyanok, amelyekre nem. A megfelelő rúdhármasok (mint az előbb a megfelelő párok) halmazával lehet meghatározni a „toldás" relációt. És csak a fantáziánktól függ, hogy milyen relációkat találunk ki, hogy „hányasokból" állítjuk össze, s a kialakított n-esek halmazának mely részhalmaza alkotja a relációt. Sőt, az egészen általános relációfogalom esetén az n-esek nem is kell, hogy ugyanazon halmazok elemei legyenek[4].

Természetesen egy adott gyakorlati terület, vagy konkrét tudományos diszciplína embereit nagyon is célszerűségi szempontok vezérlik, amikor ilyen relációkat hoznak létre, és a legtöbbször nem is kell foglalkozniuk az elvontabb, matematikai definícióval.

Nagyon fontos, hogy eddig nem volt szó számokról, mérésről. Egyelőre egy X halmaz elemei közti relációkról beszéltem. Ez a halmaz a mért „dolgok" halmaza, a jelölést e könyvben még számtalan esetben fogom használni. Nem úgy állapítom meg (egyelőre), hogy két rúd egyforma hosszú-e, hogy megmérem, hány milliméter hosszúak, s ha ugyanaz a szám adódik, akkor konstatálom, hogy igen, egyforma hosszúak, legalábbis ha elhatároztam, hogy milliméter pontossággal fogok mérni. Nem, erről egyelőre szó sincs. Ahogy mondani szoktuk: „összemérem" a két rudat. Egymás mellé állítom őket, és szabad szemmel, esetleg valamilyen alkalmas eszközzel ellenőrzöm, hogy egyenlő hosszúak-e. Miközben nehéz megfogalmazni, hogy ilyenkor mi is történik pontosan, a gyakorlatban semmi problémát nem jelent számunkra az ellenőrzés. Két lécdarab hosszának azonosságát szinte bárki nagyon gyorsan, gondok nélkül ellenőrizni tudja a két lécdarab egymás mellé helyezésével, megvizsgálva, hogy a végeik együtt vannak-e, esetleg még tapintással is ellenőrizzük, hogy az egyik nem nyúlik-e túl a másikon. Persze az, hogy milyen pontosan tekinthetjük egyforma hosszúaknak a lécdarabokat, már érdekes, ha tetszik: izgalmas episztemológiai kérdés. Ismeretelméleti probléma, hogy van-e egyáltalán objektív értelemben vett hossza egy rúdnak (egyáltalán mit jelent az, hogy „objektív értelemben vett"), valójában már az is kérdés, hogy van-e rúd, de ne szaladjunk előre, többek közt éppen e kérdésekkel szeretnék később foglalkozni a tüzetesebb vizsgálat során. Most még csak a méréssel kapcsolatos értelmezések egy „biztonságosabb", nagyon is realista, objektivista kifejtése zajlik, a kritikára később szeretnék kitérni. Most elég annyi, hogy a lécek hosszának előbb leírt összehasonlításakor a gyakorlati igények fogják megszabni, milyen kis különbségeket hanyagolunk el, s tekintjük még egyenlő hosszúaknak a rudakat.

---

[4] Egészen pontosan: ha adottak az A_1, A_2, …, A_n tetszőleges, nem üres halmazok, akkor ezek ilyen sorrendben vett direkt szorzatának egy (bármelyik) nem üres részhalmazát az adott halmazokon értelmezett, n-változós relációnak nevezzük. A halmazok direkt szorzatát így jelöljük: $A_1 \times A_2 \times \ldots \times A_n$, és az olyan $(a_1, a_2, \ldots, a_n)$ rendezett n-esek halmazát jelenti, ahol $a_i \in A_i$ $(i = 1, 2, \ldots, n)$.

Az *X* halmaz elemei közti relációk fennállásának vagy nem fennállásának ellenőrzésére tehát vannak eljárásaink egy-egy mérés körében. Legalábbis azokban az esetekben, amelyek ilyen egyszerűen kezelhetők. Két oldat pH-jának (savasságának) összehasonlítására már nincsenek ilyen egyszerű eljárásaink, legalábbis olyanok, amelyek kémiában járatlanok számára is hétköznapiak, jól ismertek lennének, mint amilyen a lécdarabok hosszának összehasonlítása. De vannak ott is kvalitatív eljárások, amelyek lehetővé teszik az ellenőrzést. A tudomány nyilván sokkal szélesebb körben ismer kvalitatív eljárásokat a „dolgok" tulajdonságainak értékei közti, vagy mondhatjuk azt is, hogy a „dolgok" közti relációk vizsgálatára. Az a fontos, hogy ezek egyelőre kvalitatív, gyakorlati jellegű, valamilyen cselekvést is igénylő eljárások. Az *X* halmaz, és a vizsgált tulajdonsággal összefüggésben rajta értelmezett relációk együttesen alkotják az *empirikus rendszer* fogalmát. Sokszor *empirikus relációs rendszert* mondanak, én is így használom majd a fogalmat. Még jelölést, rövidítést is alkalmazok: ERR. Minden mérés esetén dolgunk van egy ilyen empirikus relációs rendszerrel, ERR-rel, valamilyen a szándékainkat, a mérési tevékenységünket a „valóságba" beágyazó rendszerrel[5].

Ezek után (remélem) bátran írhatok relációkat a mért entitások tulajdonságai közti viszonyok matematikai leírásával kapcsolatban. Formalizáltan: legyenek $x_1, x_2, \ldots, x_n$ az X halmaz elemei, vagyis kiválasztunk X elemei közül n darabot. Ezek az R-rel jelölt, n-változós relációban állnak egymással, ha érvényes, hogy viszonyukat ez az R kvalitatív összefüggés jellemzi, amelynek érvényesülését empirikusan meg is tudjuk vizsgálni. E tényt, tehát hogy az $x_1, x_2, \ldots, x_n$ elemek az R relációban állnak egymással, így jelöljük: $R(x_1, x_2, \ldots, x_n)$. A mérés során az X halmaz minden egyes eleméhez hozzárendelünk egy számot, ez a hozzárendelés nem más, mint egy függvény. Jelöljük ezt a függvényt f-fel, és azt a számot, amit az x elemhez rendelünk (az x entitás adott tulajdonságának mértékét) pedig jelölhetjük $f(x)$-szel. *Stevens* értelmezésében a mérhetőség azt jelenti, hogy van a valós számok halmazán egy S reláció, amely az R-hez hasonlóan n-változós (n szám közötti viszonyt fejez ki), és azt, hogy a fenti n db. entitáshoz rendelt számok ebben a relációban állnak egymással, így fejezhetjük ki: $S(f(x_1), f(x_2), \ldots, f(x_n))$. Az R relációnak is megfeleltetjük a számokon értelmezett S relációt, nem csak az entitások képeződnek le a számokra. Rendkívül fontos, hogy az empirikus rendszeren értelmezett, R-rel jelölt relációknak vannak tulajdonságaik, és ezek a tulajdonságok ugyanolyanok, mint az adott relációnak megfelelő, a számok világában érvényesülő S relációk tulajdonságai. Arról van szó, hogy az empirikus rendszerben megkeressük azokat a sajátosságokat, amelyek a számokra is jellemzők, nyilván azért, hogy mintegy „helyettesítsük" a mért dolgokat a számokkal, a helyettesítőkkel elvégezve a számításokat levonjunk bizonyos következtetéseket (a számokra vonatkozóan), majd az eredményeket „visszavetítsük" magára az empirikus rendszerre.

Vagyis az egész folyamat „arról szól", hogy lefordítjuk a számok nyelvére azt, amit kvalitatív módon megállapítottunk az ERR-ben. Így például ha az R reláció kétváltozós, s azt fejezi ki, hogy az egyik rúd hosszabb vagy egyenlő hosszú a másikhoz viszonyítva (ez az empirikus reláció), akkor ennek a relációnak megfelel a valós számok halmazán a „nagyobb vagy egyenlő" reláció (ez a fenti jelölésben az S, és szokásosan a „≥" jelet

---

[5] Ez a leírás pontatlanná válik majd a fogalmak később sorra kerülő, precízebb tisztázása során. Meg fogom mutatni, hogy az ERR (empirikus relációs rendszer) valójában matematikai objektum, axiómarendszer állítandó fel rá. Ez az értelmezés azonban egy megváltoztatott ismeretelméleti alapokon zajló elemzés keretében kap majd szerepet. Az lesz a célom, hogy világosan elkülönítsem egymástól azokat az „objektumokat", amelyek a számunkra külső valóságban léteznek, amelyek a tapasztalati világunkban léteznek, s amelyek éppen e tapasztalati világ összefüggéseinek formalizálására születnek, s ezen utóbbi „objektumok" „világa" lesz majd az *X* halmaz és a vizsgált ERR. Most még azonban ez utóbbiakról mint „objektív", „valóságos", „hús-vér" dolgokról beszélünk, bevallottan egy objektivista „szemüveget" használva az elemzéshez.

használjuk a leírására). A mérés, vagyis a hozzárendelés során, amikor létrehozzuk az *f* függvényt, hosszabb rúdhoz nagyobb számot kell rendelnünk. Ha két rúd az *R* relációban áll egymással, akkor a nekik megfelelő valós számok az *S* relációban kell, hogy álljanak egymással.

Ha adott az ERR (empirikus relációs rendszer), megtaláltuk azokat a matematikai, vagyis a számok halmazán érvényes relációkat, amelyek „tükörképei”, vagy inkább leképezései az ERR relációinak (a dolgok közti kvalitatív relációknak), akkor még általában számtalan konkrét megfeleltetés (tehát számok hozzárendelése, *f* függvény) létezhet, vagyis a lehetséges mérési kimenetel még nem egyértelmű. Elég csak arra hivatkoznunk, hogy a lécek hosszát centiméter, inch, könyök, méter, láb, tengeri mérföld és számtalan más egységben megadva, ugyanazon hosszúság jellemzésére sokféle szám alkalmas lehet. Vagy itt van a hőmérsékleti skála, amelynél nem csak a mértékegységek lehetnek különbözők, vagyis nem csak az egységnyinek tekintett hőmérsékletváltozás lehet nagyon más, hanem még a 0 érték is egészen más állapothoz tartozhat. Vagyis sok hétköznapi példát ismerünk, amelyek azt mutatják, hogy a számoknak az ERR elemeihez való hozzárendelése nem teljesen egyértelmű, hanem bizonyos önkényes döntéseken múlik.

Stevens azt a kérdést teszi fel, hogy *egy konkrét mérés esetén vajon mi fűzi össze az összes lehetséges hozzárendelést*. Jóllehet végtelen sok hozzárendelés lehetséges (egy esetet kivéve, ld. később), de *nem lehetséges bármilyen hozzárendelés*. Ha valaki kitalálná a hosszúság mérésével kapcsolatban a „matár” mértékegységet, megadná a lehető legpontosabban, bárhol, bármikor reprodukálhatóan, hogy mely két pont közti távolság az 1 matár, akkor nem jelentene gondot lécek hosszának, utak hosszúságának, bármilyen távolságnak a megadása ezzel a mértékegységgel, vagyis lenne hozzárendelésünk. De ennek a hozzárendelésnek is „tudnia kell” azt, hogy ha két léc ugyanolyan hosszú, akkor a hozzájuk rendelt számok meg kell, hogy egyezzenek; hogy ha az egyik hosszabb, akkor a hozzá rendelt szám nagyobb, mint a másikhoz rendelt szám; és hogy két lécet egy egyenesben egymáshoz illesztve a hosszak összeadódnak. Ha bármelyik szabályt sérti a hozzárendelés, akkor „baj van”, méréselméleti szempontból diszkreditálódik az eljárás.

Mi fűzheti tehát össze az ugyanazon tulajdonság mérésekor kialakított megfeleltetéseket, *f* függvényeket? Mi írja le azt a végtelen sok hozzárendelést, ami lehetséges, és hogyan zárhatunk ki más leképezéseket? Itt is a matematikai függvényfogalom segít. Ha két leképezés egyaránt ugyanannak a tulajdonságnak (pontosabban az értékeinek) a mérését szolgálja, akkor az egyik is, és a másik is szolgáltat mértékeket, konkrét számokat. Például ugyanabban a szobában a levegő hőmérsékletét lehet mérni úgy, hogy a mértéket Celsius fokokban fejezem ki, és megtehetem, hogy Fahrenheit fokokat adok meg. Minden méréskor két különböző szám lesz az eredmény (pontosabban van egyetlen olyan hőmérséklet, amely esetén a Celsius és a Fahrenheit fokok megegyeznek, ez kb. a -40 °C = -40 °F). De mindig ugyanazt a függvényt kell alkalmaznom a Celsius fokban kifejezett mértékekre, hogy megkapjam a Fahrenheit fokban kifejezett mértéket. Képlettel:

$$F = \frac{9}{5}C + 32\,,$$

ahol *F* jelöli a hőmérséklet Fahrenheit skálán mért értékét, és *C* jelöli a hőmérséklet Celsius skálán mért értékét. A *C* értékeket kaphatjuk az *f* függvény értékeként (a mért érték), az *F* értékeket egy *g* függvény értékeként, vagyis a képletet így is írhattuk volna:

$$g(x)=\frac{9}{5}f(x)+32,$$

ahol *x* az adott test adott állapotát jelenti, az ERR egyik eleme. A lényeg, hogy van egy függvény, amelyet az egyik mértékre alkalmaznom kell, hogy megkapjam a másikat, és ez a függvény, ez a képlet mindig ugyanaz, bármely két összetartozó mértékről van szó az éppen vizsgált tulajdonság esetén. A fenti képletet így is írhatom:

$$\varphi(f(x))=\frac{9}{5}f(x)+32.$$

Ez a forma azt fejezi ki, hogy előbb az *f* függvény „elvégzi" az ERR leképezését a számok halmazára, majd a φ függvény már a számok halmazán belül „működik", számokhoz (a konkrét esetben Celsius fokokban megadott hőmérséklet mértékekhez) rendel hozzá számokat (Fahrenheit mértékeket).

Általánosan: mindig azt kell megvizsgálni, hogy az ERR valós számok halmazára történő lehetséges leképezéseit mi köti össze egymással. Milyen további, most már a valós számokat a valós számokra leképező függvények szükségesek ahhoz, hogy megkapjuk az összes mérési skálát? A skálatípusok úgy jönnek létre, hogy azt találjuk, számtalan hasonló függvény adódik a vizsgálatok során. Kiderül majd, hogy egy-egy jól elkülöníthető csoportot alkotnak azok a mérések, amelyekben a különböző skálákat egymásba a kölcsönösen egyértelmű leképezések viszik át, vagy azok, amelyek esetén ezek a leképezések rendezéstartók, azok, amelyek lineárisak, s azok, amelyek „csak" egy számmal való szorzást jelentenek (hasonlóság). De nézzük meg ezt konkrétan, az elvont fogalmazás itt jórészt érthetetlen a területet nem ismerők számára, ugyanakkor matematikailag nem is elég precíz.

Nominális skála

Vannak olyan mérések, amelyekben az *R* reláció kétváltozós, az „azonos tulajdonsággal rendelkeznek" viszonyt írja le, vagyis $R(x_1, x_2)$ (más jelöléssel: x_1Rx_2) akkor és csak akkor, ha $x_1 = x_2$, az egyenlőséget a megegyező tulajdonság értelmében véve (nem tévesztendő össze az azonossággal, amely relációban bármi csak önmagával állhat). Ilyen esetekben az *f* függvény bármi lehet, ami különböző számokat rendel hozzá a különböző tulajdonságokkal leírható elemekhez, és azonosakat az adott tulajdonság szerint megegyező elemekhez. Itt ez jelenti a reláció „átvitelét" a matematikai struktúrára. Ha a két entitás azonos tulajdonságú, akkor feleljen meg nekik ugyanazon szám, és legyen különböző e két szám, ha a mért entitások is különbözők. Az ilyen függvények majdnem bárhogyan transzformálhatók, arra kell csak vigyázni, hogy nehogy a transzformált függvény ugyanolyan tulajdonságú elemekhez nem egyező mértékeket rendeljen, és nehogy különböző tulajdonságú elemekhez azonos számokat rendeljen. Ezért megfelel minden olyan transzformáció, amely az $f(x)$ értékekre egy *kölcsönösen egyértelmű függvény* alkalmazását jelenti. Világos, hogy ebben az esetben nincs is jelentősége annak, hogy számokat rendeljünk hozzá az entitásokhoz, használhatjuk bármilyen halmaz elemeit. Egy példa: ha az „entitások" emberek, és a tulajdonság, amit vizsgálunk, a nemük, akkor az *R* reláció az „azonos nemű"-t jelenti, az *f* függvény lehet például az, amely a férfiakhoz az 1 számot rendeli (vagy mondjuk a 'férfi' jelsorozatot), a nőkhöz pedig a 2 számot (alternatíva a 'nő' jelsorozat). Bármilyen kölcsönösen egyértelmű függvény segítségével végrehajthatunk transzformációt az *f* függvényen, pl. az *f* függvény minden értékéhez, ha azok számok, hozzáadhatunk 3-at, megszorozhatjuk 2-vel, vagy ha a 'férfi' és ’nő’ jelsorozatokat használjuk, akkor bármilyen, e két értékhez két különböző „valamit” hozzárendelő függvény megfelel.

Szokás mondani, hogy a nominális skálán való elhelyezés nem is mérés. Magam ezzel nem értek egyet, hiszen a mérés definíciójának megfelel. Néha szerepel, hogy azért nem mérés ez, mert csak átnevezi az elemeket, ami azonban nem igaz, nem az elemeket nevezzük át, hanem esetleg a tulajdonságok értékeinek adunk új nevet. A nominális skálán való mérés valójában osztályba sorolás, azt adjuk meg, mely elemek tartoznak azonos osztályokba (pl. a ’férfi’ és a ’nő’ névvel illetett osztályokba).

Ettől függetlenül számtalan olyan matematikai leírás létezik a mérési skálákról, amely a nominális skálát nem sorolja a mérési skálák közé. Ennek van egy matematikailag jól alátámasztható indoka is. Az X halmazon értelmezett *ekvivalencia reláció*, vagyis az *egyenlőség* segítségével osztályokat alakíthatunk ki a mért dolgok halmazán, egy osztályba sorolva azokat, amelyek ekvivalensek, vagyis a mért tulajdonságuk szempontjából egyforma értékkel rendelkeznek. Matematikailag természetesen itt ekvivalencia reláció által generált ekvivalencia osztályokról van szó. A mérések matematikai elemzése nagyon gyakran nem is az eredeti objektumoknak megfeleltetett absztrakt entitásokon történik, hanem azok ekvivalencia osztályain. Vagyis az eredeti elemek közti egyenlőség mintegy „kikerül” a vizsgálódás köréből, elveszti jelentőségét. Az ekvivalencia osztályok egyenlősége már azonos az azonossággal, minden elem csak önmagával egyenlő (azonos), így már a nominális skálának valóban nem sok értelme marad. De mindez azon múlik, hogy a matematikai módon értelmezett ERR-t az eredeti mérendő entitásoknak megfelelő absztrakt elemek alkotják-e, vagy ezek ekvivalencia osztályai, s ez döntés kérdése. Kétségtelen, hogy az esetek döntő részében kényelmesebb, indokoltabb az ekvivalencia osztályokat az ERR elemeinek tekinteni.

Ordinális skála

Vannak olyan mérések, amelyekben az R reláció a *megelőzés* relációja. Itt tehát a mért entitásokon a vizsgált tulajdonságuk tekintetében valamilyen sorrend, rendezés értelmezett. Bármely két elemről meg tudjuk állapítani, hogy melyik előzi meg a másikat, vagy esetleg, hogy egyenlők a vizsgált tulajdonság tekintetében. Matematikai okokból a „nem követi” relációt használjuk sokszor, ez a számok körében – talán mindenki emlékszik iskolai matematikai tanulmányaiból – a „nagyobb vagy egyenlő” (≥) fogalmával rokon. Vagyis $R(x_1, x_2)$, vagy másképpen x_1Rx_2 akkor és csak akkor, ha $x_1 \gtrsim x_2$, ahol a két elem közti furcsa jel a „nagyobb vagy egyenlő” jelre hasonlít, de a vonalak azért olyan „görbék”, hogy meg is különböztessük a számok közti relációt leíró ≥ jeltől[6]. Az f függvénnyel szembeni elvárásunk most az, hogy ahhoz az elemhez, amely nem követi a másikat, nagyobb vagy egyenlő számot rendeljen, mint a másikhoz, egyenlőt csakis akkor, ha a két elem vizsgált tulajdonsága azonos[7]. Formálisan: ha $R(x_1, x_2)$, vagyis $x_1 \gtrsim x_2$, akkor

[6] Itt a „nagyobb vagy egyenlő” típusú relációt használjuk, de megtehetnénk, hogy a „kisebb vagy egyenlő” típusú relációt használjuk. Az egyik segítségével definiálható a másik. A megnevezések is megváltoznak persze, a „nem követi” kifejezés helyett a „nem előzi meg” kifejezés használható például.

[7] Egy ordinális skála értelmezésénél az elemek közt ekvivalencia reláció is létezik. Amennyiben a rendezési reláció egy úgynevezett gyenge rendezési reláció, vagyis nem a határozottan „megelőzi” ($\succ$) relációról, hanem a „nem követi” ($\gtrsim$) relációról van szó, akkor ennek létezése már lehetővé teszi egy ekvivalencia reláció értelmezését. Ugyanis akkor mondjuk egy a és egy b elemről, hogy ekvivalensek ($a \sim b$), ha egyszerre igaz, hogy $a \gtrsim b$, és $b \gtrsim a$. Bebizonyítható, hogy az így definiált reláció valóban ekvivalencia reláció, és például a valós számok esetén, ha a ≥ relációt a szokásos módon definiáljuk, megegyezik a számok egyenlőségét kifejező relációval.

$f(x_1) \geq f(x_2)$, de úgy, hogy $f(x_1) = f(x_2)$ akkor és csak akkor, ha $x_1 = x_2$. Vagyis ismét arra vigyázunk, hogy a számok közti relációk közül azokat feleltessük meg az entitások közti relációknak, amelyek megfelelnek a fenti definíciónak. Ha e függvény értékeire, vagyis a mérés eredményét jelentő számokra egy olyan függvényt alkalmazunk, amely egyrészt az f függvény azonos értékeihez azonos számokat rendel, egyébként pedig monoton növekvő függvény, akkor a két függvény egymás utáni alkalmazásával olyan függvényt kapunk, amely szintén megfelel a definíciónak. Így egy skálából kiindulva végtelen sok másik skálát kapunk, ha az eredeti függvényünket „megtoldjuk" egy a leírt tulajdonsággal rendelkező függvénnyel. Az ordinális skálát tehát a rendezéstartó transzformációk jellemzik.

Példaként szerepeljen itt az iskolai végzettségek négy értéke, a legfeljebb általános iskolai végzettség, a szakmunkás végzettség, az érettségi és a diploma. Az X halmaz nyilván emberek halmaza, a rajtuk értelmezett R reláció szerint az egyik ember nem követi a másikat, ha azonos az iskolai végzettségük, vagy az elsőé magasabb az előbbi felsorolás szerint, mint a másodiké. Az f függvény hozzárendelhet bármilyen négy, az előbbi felsorolásnak megfelelően növekvő számot a végzettségekhez, csak a hasamra ütve, mondjuk az 5, 10, 15, 20 számokat, de természetesen a kutatómunka során általában az 1, 2, 3, 4 számokat használjuk a legtöbbször (és még a növekedés sem fontos, egy csökkenő számsorozat is szerepelhetne). Nincs akadálya azonban annak, hogy ismét „megtoldjuk" ezt a függvényt, és az értékeire egy újabb függvényt alkalmazzunk, amelynek, mint mondtuk, monoton függvénynek kell lennie. Csak a példa kedvéért legyen ez a függvény az $x + 7$ képlettel leírt összefüggés. Az 1, 2, 3, 4 számok helyett ilyenkor a 8, 9, 10, 11 számok rendelődnek hozzá az iskolai végzettségekhez. De korántsem szükséges, hogy a függvény ilyen „szabályos" (képlettel egyszerűen kifejezhető) legyen. Az 1, 2, 3, 4 számoknak megfeleltethetnénk mondjuk az 5, 400, 401, 1328 számokat is. Talán ezekből a példákból is jól látható, hogy nagy gyakorlati jelentősége annak, amit a hozzárendelés transzformációjáról mondunk nincs, ilyen műveleteket nem, vagy csak nagyon ritkán szoktunk elvégezni a kutatások során. A transzformációnak, a lehetséges transzformációk jellegének a matematikai leírás, a skáláknak részhalmazokba történő sorolása szempontjából van értelme. Láttuk, hogy a nominális skálákat az jellemzi, hogy az f függvény értékeire (a mérési eredményekre) alkalmazható transzformációk köre az egy-egyértelmű függvények halmaza, az ordinális skálákat az jellemzi, hogy szigorúan monoton függvények segítségével transzformálva a hozzárendelést megadó f függvényeket, a definíciónak megfelelő hozzárendeléshez jutunk.

Intervallumskála

Ez a skála számunkra a legfontosabb. Ennek az az oka, hogy a képességek fejlettségének mérésekor kapott eredményekkel olyan a statisztikában lépten, nyomon használt műveleteket végzünk, amelyekhez szükséges, hogy adataink legalább intervallumskálán helyezkedjenek el. Az átlag, a szórás, a korrelációs együtthatók kiszámítása, a szóráselemzés, a regresszióelemzés, a klaszter- és faktoranalízis mind olyan módszerek, amelyekhez szükséges, hogy intervallumskálával legyen dolgunk.

Vannak olyan mérések, amelyek során az R reláció nem az X halmaz elemein, hanem az X halmaz elemeiből alkotott rendezett párokon definiált. Olyan esetekben van erről szó, amikor egyfajta „távolság", „különbség" van a vizsgált, mérendő elemek között az adott tulajdonságuk tekintetében. Annyit követelünk meg (de ez a fogalmazás még pontatlan), hogy a párok elemei közti „távolságok" között fennálljon valamilyen rendezési reláció (nagyobb-kisebb, előrébb-hátrébb, stb.). Természetesen (ritka esetektől eltekintve) nem a geometriai távolságról van szó, hanem valami más olyan viszonyról, amivel kapcsolatban el tudjuk dönteni empirikusan, hogy a párok között fennáll-e a rendezés vagy sem. Egzakt módon meg tudjuk adni, milyen tulajdonságai vannak annak a relációnak, amely egy halmaz elemeinek párjaiból képezett új elemek között kell, hogy érvényesüljön, ha azt akarjuk, hogy ez a reláció távolság jellegű tulajdonsággal legyen kapcsolatos. Ahogyan a nominális és az ordinális skálák esetén is a kiindulásként értelmezett ekvivalencia- és rendezési relációk tulajdonságai is nagyon pontosan definiáltak. (Az intervallumskála konstrukciójához szükséges reláció tulajdonságait sokkal pontosabban specifikálom a könyvnek abban a részében, ahol a modern tesztelmélet egy részterületét, a legegyszerűbb Rasch-modellt mutatom be.)

Ha egy vasrudat valamilyen állapotában felmelegítünk, akkor megnő a hossza. Most képzeljük el, hogy egy másik állapotban, amikor mondjuk jóval melegebb ez a vasrúd, megint melegítjük, méghozzá nagyobb mértékben (pl. ugyanazzal a melegítési eljárással hosszabb ideig, mint az előbb), hogy a hossza nagyobb mértékben változzék, mint az előbbi melegítés során. Ilyen módon egy rendezési relációt hozunk létre egy vasrúd hőállapotainak párjai között (vagyis nem az állapotok között közvetlenül, hanem ezen állapotok párjai között). Ez a reláció segíthet a hőmérsékleti skála definiálásában.

Igazság szerint azt is mondhatnánk, hogy állapot-négyesek között van reláció, az x_1, x_2, x_3, x_4, elemek a D relációban állnak egymással, vagyis $D(x_1, x_2, x_3, x_4)$, vagy egy más jelöléssel $x_1x_2Dx_3x_4$. Lényegében ugyanezt mondjuk, ha sikerült konstruálnunk egy olyan R relációt az X halmaz elemeiből alkotott párokon, amely rendezési reláció, és $R((x_1, x_2,), (x_3, x_4))$, vagy más jelöléssel: $(x_1, x_2)R(x_3, x_4)$. Vagyis *Stevens* általános definíciója itt is alkalmazható, csak kicsit bonyolultabb a relációval kapcsolatos kívánalom megfogalmazása. A D relációnak – s ahogy említettem, ezt meg is mutatom később – pontosan megadott ulajdonságai vannak.

Milyen az f függvény? A nominális és az ordinális skálák értelmezése során meglehetősen „liberálisak” lehettünk a lehetséges függvények kijelölésében. Itt most tovább szűkül a kör. Az f függvénynek „tudnia kell”, hogy az R reláció szerinti kapcsolatban lévő párok elemeihez olyan számokat rendeljen, amelyek különbsége az egyik párban nagyobb, mint a másik pár elemeihez rendelt számok különbsége. Formálisan: ha $R((x_1, x_2), (x_3, x_4))$, vagy másképpen, ha $D(x_1, x_2, x_3, x_4)$, akkor $f(x_1) - f(x_2) > f(x_3) - f(x_4)$. Lehet természetesen, hogy nem az itt leírt, hanem a „fordított” $R((x_3, x_4), (x_1, x_2))$ reláció áll fenn, és akkor a megfeleltetés során az $f(x_3) - f(x_4) > f(x_1) - f(x_2)$ összefüggésnek kell érvényesülni, vagy esetleg egyik sem érvényes a kettő közül, s akkor szükségképpen $f(x_1) - f(x_2) = f(x_3) - f(x_4)$.

Milyen transzformációk engedhetők meg az f függvénnyel kapcsolatban? Megmutatható, hogy az f függvényt megszorozva egy pozitív valós számmal, és hozzáadva egy valós számot, olyan függvényt kapunk, amely szintén megfelel *Stevens* elvárásainak. (Emlékeztetek, ez azt jelenti, hogy ha bármely x_1, x_2, x_3, x_4 elemekre igaz volt, hogy $x_1x_2Dx_3x_4$, és ezért $f(x_1) - f(x_2) > f(x_3) - f(x_4)$ is igaz volt, akkor a valós számokat a valós számokra leképező, lineáris φ fügtgvény, mely esetén $\varphi(a) = \alpha a + \beta$ olyan, hogy $\varphi(f(x_1)) - \varphi(f(x_2)) > \varphi(f(x_3)) - \varphi(f(x_4))$ is igaz lesz.) Ezt egy példával már illusztráltam korábban: a hőmérséklet mérésének Celsius és Fahrenheit skálája közti összefüggéssel. Láttuk, hogy ha sikerült az ERR elemeinek megfeleltetni a Celsius skálán elhelyezkedő számokat, akkor e számokra alkalmazva az ott megadott összefüggést, függvényt, megkaphatjuk a Fahrenheit skálán a hőmérsékletértékeket. Nem nehéz ellenőrizni, hogy ezzel az átszámítással a hőmérsékletek közötti különbségek rendezési relációja nem változik meg. Mindez azt jelenti tehát, hogy az intervallumskálát az specifikálja, hogy *a lehetséges skálák közt egy pozitív lineáris leképezés teremt kapcsolatot* (bármelyiket megkapjuk egy másikból, ha ez utóbbi értékeit ugyanazzal a pozitív valós számmal szorozzuk, és ugyanazon számot adjuk hozzá).

Arányskála

Nem megyek végig pontosan a definíción. Nagyon hasonló az intervallumskála definíciójához, csak itt az *X* halmaz elemeiből alkotott párok között olyan rendezési relációnak kell érvényesülni, hogy a párok elemeinek megfeleltetett számok hányadosai között a „nagyobb" összefüggés legyen érvényes (az intervallumskálánál a különbségek között érvényesült ez a rendezési reláció). Az *f* függvény tehát ennek kell, hogy megfeleljen, és a rajta végrehajtott transzformáció egy nullától különböző számmal való szorzás lehet. Egy adott *X* halmaz esetén a különböző arányskáláknak ugyanott van a zéruspontja (ha 0 az *f* függvény értéke, akkor azt akármilyen valós számmal szorozzuk is, mindig 0-t kapunk). A hőmérséklet példájánál maradva, az a skála, amelynek 0 pontja az abszolút 0 foknál van (ez a leghidegebb állapot, ez alá nem csökkenhet a hőmérséklete semmilyen testnek, de ez a 0 pont nem is érhető el), arányskála, és bármely a hőmérséklet mérésére létrehozott arányskálának az abszolút 0 hőmérsékleten kell legyen a zéruspontja. Az egységeik lehetnek különbözők, a fizikában használt Kelvin skála esetében például az egység (az egy foknyi hőmérsékletváltozás nagysága) megegyezik azzal, ami a Celsius skálára jellemző, de ez nem lenne kötelező, önkényes definíció.

Stevens munkájának értéke

Mint mondtam, ezzel valójában nem írtuk le az összes lehetséges skálatípust. Pedagógiai, pszichológiai mérések esetén megelégszünk ezekkel a típusokkal, sőt, valójában (a fizikai mennyiségek mérésével összefüggő, például időméréssel kapcsolatos vizsgálatok kivételével) az arányskála nem is fordul elő a mi területünkön. Ez teljes összhangban van azzal a ténnyel, hogy a pszichológiában (és a pedagógiában) nincsenek additív természetű mennyiségek. Ugyanis kimutatható, hogy a *Stevens* által bevezetett skálák közt az arányskála megegyezik az additív mennyiségek skálájával. *Stevens* éppen azzal lépte túl a hagyományos megközelítéseket, hogy újabb skálákat épített fel (intervallum-, ordinális-, nominális skálák), mert újabb empirikus és matematikai relációkat vont be a skálák értelmezésébe.

Vagyis *Stevens* munkája kellett volna, hogy megmentse az emberi képességek fejlettségének mérését, a pszichometriát. Látjuk majd azonban, hogy a képességmérésekkel kapcsolatban nem csak skálaproblémák merülnek fel. Bár *Stevens* javaslatai legitimálták az olyan méréseket, amelyekben nem additív, viszont a különbségek, a távolságok értelmezésére lehetőséget adó mennyiségek szerepelnek, ám annak bizonyítása, hogy a képességek mérésekor intervallum-skálán elhelyezkedő mennyiségekkel van dolgunk, általában nem része a vizsgálódásoknak. A modern tesztelmélet alkalmazásának keretei között, ha nem is egyszerű matematikai „trükkökkel", de sikerült intervallumskálákat fabrikálni. Erre azonban később térek ki, a történeti áttekintés után. Ott majd látni fogjuk, hogy a pedagógia számára rendkívül fontos területeken, az összetettebb képességek fejlettségének mérésében a problémák így is változatlanul fennállnak.

Reprezentácionizmus, a szintézis

A reprezentációs méréselmélet az a szintézis, amelyhez a *Helmholtz* által elkezdett folyamat a 20. században elvezetett. Ez az elmélet matematikailag formalizált, absztrakt leírása a mérési folyamatnak. Nem új bizonyos értelemben, hiszen magába foglal minden korábbi eredményt, inkább talán általánosításnak tekinthető. Itt nem fogom bemutatni a legelvontabb formájában, hiszen annak jó megértéséhez magasabb szintű matematikai tudásra van szükség, csak a lényeget jelzem, inkább leíró, magyarázó módon, és nem a formalizált kifejtést követve. Érdeklődők a matematikai szempontból „teljes" leírásokkal megismerkedhetnek elsősorban a következő művekből: Luce és Suppes 2002; Narens 1985, 2002a, b; Michell 1990, 1999; Luce 1996; Luce és Narens 1981, 1983; Roberts 1979; Krantz és mtsai. 1971.

A reprezentációs méréselmélet lényegét, igaz még nem a kellő matematikai precizitással, és nem is a lehetséges maximális általánossággal, már *Stevens* megfogalmazta. Az elmélet egzakt matematikai alapokra való építése és kiterjesztése, általánosítása *Suppes* (1951) művének megjelenésével kezdődött, és lényegében még napjainkban is tart. E folyamat jellemzője, hogy a kutatók mindannyian ugyanabban a paradigmában gondolkodnak, a használt nyelv, a jelölések, a fogalmak, a definíciók, a tételek egységesen elfogadottak. Ilyenek az érett elméletek, uralják az adott kutatási területet, ahogyan azt *Thomas Kuhn* (1962/1984) a normál tudomány fogalmával leírta. Ez nem jelenti azt, hogy a későbbiekben nem jelentkezhet egy olyan matematikai elmélet, amely új alapokra helyezi a mérési folyamatok elemzését. Sőt, a szakirodalomban már meg is jelent olyan méréselmélet, amely alternatívának tekinthető. *Zoltan Domotor* (a *Pennsylvania Egyetemen* oktató filozófus) és *Vadim Batitsky* (a *New Yorkban* működő *St. John's Egyetem* filozófia tanára) írtak le legutóbb egy olyan új méréselméletet (Domotor és Batitsky 2008), amely azonban struktúrájában, követelményeiben sokkal inkább igazodik a természettudományok igényeihez, mint a pszichológiához. Saját meglátásom szerint ez az új, analitikusnak „elkeresztelt" méréselmélet a reprezentációs méréselméletnél is nehezebben lesz alkalmazható a pszichológiában és a pedagógiában, pedig, mint majd a részletes kifejtésben látni fogjuk, a reprezentációs méréselmélet igényeit is nagyon nehezen tudják kielégíteni a mi szeretve tisztelt diszciplínáink.

Stevens elgondolásainak elemzése során a reprezentációs méréselmélet kiindulópontjai tulajdonképpen már szerepeltek. A vázlatos ismertetéshez szükségünk van a *relációs struktúra* és a *homomorfizmus* fogalmaira. Relációs struktúrának nevezünk a matematikában egy olyan rendszert, amely alaphalmazokból, azok elemei közötti relációkból, és axiómákból áll. Az axiómák olyan állítások, amelyeket nem bizonyítunk. Egy-egy axióma tartalma lehet az, hogy valamelyik reláció valamilyen tulajdonságú, egy axióma összekapcsolhat több relációt, az alaphalmaz kitüntetett (külön megnevezett) elemeiről állíthat valamit[8]. Ilyen rendszer például a természetes számok struktúrája, amely egy halmazból (a természetes számok), az elemek közt a 0 kitüntetett elemből, a rákövetkezés relációjából, valamint a természetes számok struktúráját ténylegesen megformáló axiómarendszerből (mondjuk a *Peano*-féle axiómákból) áll. Egy relációs rendszert egy másikra leképezhetünk *homomorf* módon, ha az alaphalmazok, illetve természetesen elemeik között van nem feltétlenül egy-egyértelmű, de függvénykapcsolatot jelentő leképezés (tehát a leképezés egyértelmű), illetve az egyik struktúrában szereplő relációknak is megfelel a másik struktúrában (megint nem feltétlenül egy-egyértelmű módon, de egyértelműen, tehát függvénykapcsolatot teremtve) egy-egy reláció, úgy, hogy az alaphalmazok közti megfeleltetés relációtartó. Ez utóbbi azt jelenti, hogy ha az eredeti halmaz bizonyos elemei egy adott relációban álltak egymással, akkor a másik halmazból nekik megfelelő elemek a relációnak megfelelő relációban állnak majd egymással.

[8] Az alaphalmaz egy-egy kitüntetett eleme nem újabb entitása az axiómarendszernek, mert az alaphalmaz elemei valójában egyváltozós relációk.

A méréselméletre koncentrálva kissé érthetőbbé tehetjük a fogalmat. Már eddig is erről a homomorfizmusról volt szó, csak a fogalmat nem említettem. A méréselméletben is két relációs struktúra szerepel, egy empirikus relációs rendszer (ERR) és egy másik, a matematikában egyébként is használatos relációs rendszer, rendszerint valamilyen számhalmaz a szokásos értelmezéssel. Az egyszerűsítés kedvéért itt ezt a másik halmazt, amelyre az ERR elemeit leképezzük, *képhalmaznak* fogom hívni. Egy eljárás akkor nevezhető *mérésnek*, ha meg tudjuk adni ezeket a rendszereket, tehát a következő összetevőket:

- alaphalmazok (az ERR alaphalmaza és a képhalmaz),
- elemeikként a kitüntetett elemek,
- relációk,
- axiómarendszerek (itt előfordul, hogy a képhalmaz elemei nem alapfogalmai a rendszernek, de van egy háttérelmélet, amelyben a mérendő objektumoknak megfelelő matematikai objektumok definíciója létezik);

továbbá létezik a két struktúra között egy homomorfizmus, vagyis a relációkat megtartó leképezés[9]; és létezik mindezeknek empirikus megvalósítása, azonosítása is:

- azonosítani tudjuk az elemeket (tudjuk, hogy az ERR absztrakt, matematikai halmazai elemeinek mik felelnek meg a tapasztalati világunkban, tudjuk, hogy milyen „valóságos” viszonyok, műveletek felelnek meg az ERR absztrakt relációinak, ismerjük, és reprodukálható módon azonosítani tudjuk az ERR kitüntetett elemeit),
- meg tudjuk állapítani, hogy az így a tapasztalati világban azonosított halmaz adott, megfelelő számosságú, véges részhalmazának elemei egy adott relációnak megfelelő viszonyban, kapcsolatban állnak-e egymással,
- empirikusan azonosítani tudjuk azokat az elemeket, amelyekre valamelyik alap- vagy definiált fogalomhoz kapcsolódó tulajdonság illik,
- empirikusan ellenőrizni tudjuk az axiómák tapasztalati világbeli viszonyokra, történésekre való lefordításának hibahatárokon belüli érvényesülését,
- és létezik egy empirikus eljárás, amely lehetővé teszi, hogy ténylegesen megmondjuk, hogy az ERR egy elemének a képhalmaz mely eleme (rendszerint szám) felel meg.

Vagyis összefoglalóan: a mérés korrekt kidolgozásához minden esetben egyrészt komoly *háttérelméleteknek* kell rendelkezésre állniuk (ERR, matematikai relációs rendszer, homomorfizmus), másrészt a mérést ténylegesen kivitelezhetővé tevő *empirikus eljárások egy megfelelő készletével* kell rendelkeznünk.

A méréselmélet *konkretizációinak* nevezhetjük a „relációs rendszer típusoknak” a leírását, matematikai formalizálását, amelyek az egyes méréstípusokat formálják meg. Pontosabban: megalkotható egy-egy axiómarendszer, amely

- a nominális struktúrákat,
- a rendezési struktúrákat,
- a különbségi struktúrákat,
- az extenzív struktúrákat

[9] Korábban már megjegyeztem, hogy az ERR struktúrájának leírása során hagyatkozhatunk csak a mért entitásoknak megfelelő absztrakt matematikai halmazelemek közti ekvivalencia reláció ekvivalencia osztályainak elemzésére. Ha a matematikai axiómarendszereket ezekre az (agregált) elemekre építjük, akkor viszont az ERR és az adott számhalmaz közti leképezés izomorfizmus, vagyis relációtartó egy-egyértelmű leképezés.

írja le. Ezek azok a struktúrák, amelyek szorosan kötődnek a Stevens által is elsősorban leírt skálákhoz (de vigyázat, nem mindegyik lehetséges skála szerepel itt, és nem is az összes mérési típus). Rendre a nominális-, az ordinális-, az intervallum- és az arányskálákon való mérés alapjait teremtik meg ezek az axiómarendszerek.

Minden ilyen axiómarendszer felállítása után kimondatik két a reprezentációs méréselméletben alapvető szerepet játszó tétel. Az egyik, a *reprezentációs tétel*, amely azt állítja, hogy ha egy relációs rendszer rendelkezik az axiómarendszerben leírt tulajdonságokkal (tehát ha nominális struktúra, rendezési struktúra, stb.), akkor létezik homomorfizmus az e struktúrával jellemezhető alaphalmaz (ERR) és a valós számok halmaza (képhalmaz) között (lehetséges a mérés). A másik alapvető állítás az *unicitási tétel*, amely viszont – egyszerűen fogalmazva – azt mondja meg, hogy az adott struktúrát alkotó ERR és a valós számok közötti homomorfizmusok egymással milyen kapcsolatban vannak, vagyis milyen skáláról van szó.

E matematikai „bűvészkedéseknek” még egy fontos pontja van. Az összes művelet, amelyről eddig beszéltünk, azt a célt szolgálja, hogy számunkra alkalmas módon (relációtartás, homomorfizmus) leképezzük az ERR alaphalmazát a valós számok halmazára. Ezt részben azért tesszük, mert a számok világában érvényes szabályokat szeretnénk alkalmazni, nem az eredeti alaphalmazzal és struktúrával (ERR) akarunk műveleteket végezni, hanem az elemeinek a reprezentánsaival, a számokkal. Csakhogy a számok világa rendkívül gazdag, csak e halmazra figyelve olyan műveleteket is végezhetünk, amelyeknek, illetve amelyek eredményeinek nincsenek megfelelőik az ERR-ben. Ezek a műveletek értelmetlenek. Így például ha az ERR-re csak rendezési struktúra jellemző, akkor értelmetlen dolog kiszámolni az ERR elemeinek megfeleltetett számok különbségét, összegét, átlagát, adatok szórását, stb. Megtehetjük persze, a matematika szenvtelen ebből a szempontból, nem akadályozza meg, hogy mi ilyesmiket számoljunk, csak a műveletek semminek nem felelnek meg az ERR struktúrájában, és a számítások eredményei sem mondanak semmit az ERR-ről. Ezért a méréselméletben megszületett az „értelmesség” (meaningfulness) fogalma. Nagyon egyszerűen a lényeg: egy az adatokon (az ERR elemeihez rendelt számokon) érvényes relációt akkor nevezünk értelmesnek, ha annak megfelel egy az ERR-re felépített axiómarendszerben definiálható reláció, s annak van is valamilyen értelmes, kvalitatív mondanivalója a számunkra.

Mivel azonban ez nyilván sokak számára „kínai”, kissé másképpen, talán érthetőbben is leírom. Egy mérés során – ahogy láttuk – általában sokféleképpen lehet leképezni a vizsgált halmaz elemeit homomorf módon a számok halmazára. Egy a vizsgálat tárgyáról a számok segítségével megfogalmazott kijelentés akkor értelmes, ha igaz vagy hamis volta megmarad, ha áttérünk egy akármilyen másik (persze ugyanúgy legitim) skálára. Az, hogy két rúd hossza ugyanannyi, értelmes kijelentés, mert bármilyen skálát használnánk is, ez lenne a helyzet, ha egy skála esetén ez volt. Ha van két edényben víz, és az egyikről kijelentjük, hogy melegebb, mint a másik, szintén független a hőmérsékleti skála választásától, vagyis értelmes. Mindegy, hogy a hőmérsékletet Celsius-, Fahrenheit-, Réaumure- vagy abszolút fokokban adjuk meg, mindegyik skálán nagyobb számot kell hozzárendelni a melegebb vízhez. Ám ha azt mondjuk, hogy az ebben a lábasban lévő víz kétszer olyan meleg, mint a másikban lévő, az már nem értelmes kijelentés. Mert lehet, hogy Celsius fokokban megadva a hőmérsékleteket valóban ez a helyzet, mondjuk az egyik 20 °C-os, a másik 40 °C-os, azonban ugyanezek a hőmérsékletek °F-okban: 68 és 104, szó sincs róla, hogy az utóbbi az előbbi kétszerese lenne.

Még inkább átérezhető talán az értelmesség fogalmának jelentősége számunkra, ha a következő problémára gondolunk. Tegyük fel, hogy egy teszt alkalmazásával két tanuló arányszámítási képességének fejlettsége ugyanannyi, mondjuk 68% (ha százalékosan adjuk meg az adatot). Van biztosíték arra, hogy egy másik, elvileg szintén az arányszámítási készség (képesség?) mérésére szolgáló teszt segítségével is egyenlőknek bizonyulnak a fejlettségek? Sőt, vajon minden lehetséges teszt (legalább elvileg) ugyanazt az értéket mérné (becsülné) a két tanuló esetén? Később, amikor a képességmérések problémáját tárgyalom, erre a kérdésre még részletesen visszatérek, de talán az értelmesség fogalma jelentőségének érzékeltetésére a példa megfelelő volt.

A kutatásokban, és sok más gyakorlati tevékenységben szükségessé váló mérések esetén ez a fogalom azért is nagyon fontos, mert erősen behatárolja az adatokkal végezhető műveletek körét. Jelentős problémák származnak abból, ha egy mérés során mondjuk csak rendezési struktúrával van dolgunk, ám mi a számokkal olyan műveleteket végzünk, amelyek nem értelmesek. Ez fordul elő minden olyan esetben, amikor csak ordinális skálán elhelyezkedő adatokkal (pl. iskolai osztályzatokkal) számolunk átlagokat, szórásokat, sőt, ilyen adatokat használunk fel regresszióanalízisben, faktoranalízisben, klaszterelemzésben. (És azt a problémát nem is említettük, hogy az osztályzatok ordinális skálán való elhelyezkedése is alapvetően kérdéses.)

Általános értelemben akkor nevezhetünk egy mennyiséget *kvantifikálhatónak*, illetve *kvantifikáltnak*, ha az itt leírtak érvényesek rá. A kvantifikálhatóság tehát jelentős feltételek kielégítését követeli meg. Ez érthető is, gondoljunk csak arra, hogy számos tudományos fogalmunk esetén a kvantifikálhatósággal, s az erre épülő mérésekkel alakíthatjuk ki a kapcsolatot a tudományos elképzelések és az általuk leírt, a tapasztalati világunkban megjelenő entitások között. De az is kiemeli a kvantifikálhatóság jelentőségét, hogy ez teszi a tudományos terminusokat alkalmassá arra, hogy az összefüggéseket (törvényeket, elméleteket) matematikai formában, mint mennyiségek közti összefüggések kereteit adjuk meg.

Nem szükségszerű, hogy ennek minden esetben így kell lennie. Nem biztos, hogy a tudomány minden fogalma kvantifikálható, sőt, nagy valószínűséggel vannak olyan tudományos fogalmak, amelyek soha nem lesznek kvantifikálhatók. Éppen ezért jelentős kérdés, hogy egy-egy tudományos fogalom kvantifikálható-e egyáltalán. E könyv részben arról szól, hogy a komplex képességek esetén a képességfejlettség egy olyan fogalomnak tűnik, ami nem kvantifikálható. Persze nagyon sok esetben rendkívül nehéz vállalkozás bebizonyítani valamiről, hogy az nem tartozik bele egy adott halmazba, s így nehéz bebizonyítani egy tudományos fogalommal összefüggésben, hogy az *nem* kvantifikálható. Legfeljebb azt teheti a szakember, hogy a meglévő kísérletekről bebizonyítja, hogy azok sikertelenek, illetve maga is „gyárthat" reményteljes értelmezéseket, amelyekről szintén azt látja be, hogy nem adják a probléma megoldását. Én is csak ennyit vállalhatok itt. És ez azt jelenti, hogy a kérdés nyitott, nincs kizárva, hogy valamikor születnek jó javaslatok az összetettebb képességekre vonatkozóan az emberek képességfejlettségének kvantifikálására (miközben nekem ezzel kapcsolatban erős kétségeim vannak).

Példa

El kell ismernem, hogy matematikai szövegekhez nem szokott olvasók számára az előző oldalakon szereplő információk nehezen emészthetőek voltak (ráadásul a kvalitatív, nem formalizált leírás miatt számos pontatlanságot kellett elkövetnem, ami viszont nyilván a precízebb kifejtést igénylők, és azokat értők számára jelent problémát). Ezért most egy példával igyekszem érthetőbbé tenni a reprezentációs méréselmélet mondanivalóját. A hosszmérés már sokszor szolgált példaként, használjuk most is ezt az egyszerű mérést!

Először is, a hosszúság mérése számára az extenzív mérés jelent elméleti hátteret. Az extenzív mérések esetén az ERR-en két reláció „működik": egyrészt az ERR alaphalmazának rendezettnek kell lennie, tehát van egy (a matematikusok úgy mondják: gyenge) rendezési reláció, továbbá egy olyan háromváltozós reláció, amely a konkatenációt írja le (kétváltozós művelet). A formalizált leírásban mindkét reláció tulajdonságait pontosan megfogalmazott axiómák formálják meg. Itt tehát a méréselmélet felépít egy általános struktúrát, amelyhez aztán csak „hozzá kell igazítani” a hossz, az idő, a tömeg, a térfogat, stb., mint extenzív mennyiségek mérését.

A hosszmérésnél az ERR alaphalmaza a szakaszok halmaza. A szakasz elvont matematikai (geometriai) objektum, s valójában nem is alapfogalma a geometriai struktúráknak, hanem definiált fogalom. A háttérelmélet tehát a geometria valamely axiómarendszere, amely definiálja a szakaszokat, amelyek a mérendő entitásokkal hozhatók majd kapcsolatba. Két relációt definiálunk: (1) egy rendezési relációt, neve lehet például az, hogy „legalább olyan hosszú”, és (2) a konkatenációt, a szakaszok összefűzését, amely egy háromváltozós reláció. A geometriai háttérelmélet (mondjuk a *Hilbert*-féle axiómarendszer) lehetővé teszi számunkra, hogy az így definiált relációkról bebizonyítsuk, hogy teljesítik az extenzív struktúrák általános leírásában szereplő két relációval szemben elvártakat. És ez valóban belátható, a szakaszokon definiált relációk megfelelnek az igényeinknek. A részleteket itt nem írom le, a korábban már említett szakirodalomban megtalálhatók, de leírok egy példát annak érzékeltetésére, hogy milyen tételek bizonyítására van szükség annak belátása során, hogy a geometriai szakaszok extenzív struktúrát alkotnak. A példa a rendezési relációval kapcsolatos. Legyenek a, b, és c szakaszok, és jelölje a rendezési relációt R. Annak, hogy a szakaszok halmaza extenzív struktúrát alkosson, az egyik teljesítendő feltétele, hogy igaz legyen a következő: ha aRb és bRc, akkor igaz aRc is. Kicsit szemléletesebben: hívjuk az R relációt így: „nem övidebb”. Ha a szakasz nem rövidebb, mint b, és b nem rövidebb, mint a c, akkor az a sem rövidebb, mint a c szakasz. Ez az állítás bebizonyítható az euklideszi geometria axiómáinak felhasználásával. Van vagy másfél tucat ehhez hasonló követelmény (axióma), amelyek érvényesülését be kell látni ahhoz, hogy bátran állíthassuk, a geometriai szakaszok extenzív struktúrát alkotnak. És ez valóban megtehető, remélem, az olvasók ezt most elhiszik nekem[10].

Van egy kitüntetett szakaszunk. Az elméleti felépítésben természetesen ez bármilyen szakasz lehet, a gyakorlati megvalósítás során majd meg kell adnunk konkrétan is egy bárhol reprodukálható hosszúságú szakaszt (bárhol tudni kell előállítani a gyakorlarban egységnyinek választott hosszúsággal megegyező hosszúságot, pontosabban olyan tárgyat, amelynek egy kitüntetett mérete megegyezik a kiválasztott tárgy adott méretével, de ez most még nem feladat, egyelőre a matematikai háttér kiépítésénél tartunk).

A matematikai struktúra, a képhalmaz a pozitív valós számok struktúrája (a 0-t és a negatív számokat egyáltalán nem használjuk az ilyen jellegű méréseknél). Ennek is van természetesen korrekt axiómarendszere. Az ERR-beli rendezési relációnak a „nagyobb vagy egyenlő” relációt feleltetjük meg, míg a konkatenációnak a megfelelő számok összeadását. Nagyon könnyű belátni, hogy a pozitív valós számok is extenzív struktúrát alkotnak. A kitüntetett szakasznak az 1 számot feleltetjük meg.

Felépítettük a két konkrét matematikai struktúrát. Eddig maximálisan megfeleltünk az extenzív struktúrákkal szembeni elvárásoknak. Akkor viszont érvényes a reprezentációs tétel is, az unicitási tétel is, és átfogó képünk van arról, hogy mi számít a szakaszok világában értelmes relációnak, összefüggésnek. A reprezentációs tétel

[10] Az euklideszi geometria axiomatikus leírásának van egy olyan változata, amely a *Hilbert* által megalkotott alapokra épül, de annak olyan, *Birkhoff*, amerikai matematzikus által javasolt módosítása, amelyben éppen két pont távolsága lesz az egyik alapfogalom.

biztosítja számunkra, hogy a szakaszokhoz úgy tudunk hozzárendelni számokat, hogy a hosszabb szakasznak (ERR-beli reláció) nagyobb szám (képhalmazbeli reláció) feleljen meg, és azt is, hogy ha két szakasz konkatenáltja egy harmadik (ERR-beli reláció), akkor a két szakaszhoz rendelt szám összege a harmadik szakaszhoz rendelt szám lesz (képhalmazbeli reláció). A reprezentációs tétel nem ad meg konkrétan hozzárendelést, csak annak megfelelő tulajdonságokkal való létezését mondja ki. Az unicitási tétel szerint pedig itt arányskáláról van szó, vagyis bármely két, a feltételeknek megfelelő leképezést az köt össze egymással, hogy az egyik szerinti mértékeket (a szakaszok hosszmértékét) ugyanazzal a számmal kell megszorozni, hogy megkapjuk a másik szerinti mértékeket. Az elméleti háttér tehát azt biztosítja, hogy végtelen sokféle módon rendelhessünk hozzá számokat a szakaszokhoz, de azt is megmondja, hogy ez a végtelen sok hozzárendelés milyen lehet, egymással ezek milyen kapcsolatban vannak. Ha a végtelen sok hozzárendelés közül egyet szeretnénk kiemelni (a konkrét mérésnél már erre lesz szükségünk), akkor annyit kell csak tenni, hogy meg kell adni egyetlen konkrét szakaszra vonatkozóan a hozzárendelést, tehát, hogy mennyi legyen a hosszának a mértéke, s ezzel már rögzítettük a többit is. A gyakorlatban ez annak a szakasznak a megadását jelenti, amelyhez az 1 számot rendeljük hozzá, és a mértékegység megadásával jelezzük, hogy melyik hozzárendelésről van szó a végtelen sok közül. Ha például azt mondjuk majd egy szakaszról (mondjuk egy gerendáról), hogy az 3,5 méter hosszú, akkor jelezzük, hogy azzal a hozzárendeléssel dolgozunk, amely esetén az etalonnak választott 1 méternek van egy pontos meghatározása. De ez már a mérés konkrét, gyakorlati kivitelezéséhez tartozik.

Fontos, hogy mivel extenzív mérésről van szó, ezért bármilyen a statisztikában használatos módszert legitim módon alkalmazhatunk a hosszadatokkal kapcsolatban (ezt az extenzív mérés általános tulajdonságai biztosítják, külön a hosszmérésre nem kell bebizonyítani).

A leírt összefüggések azonban egyelőre csak az elméletet jelentik. Az ERR is egy matematikai struktúra, a pozitív valós számok halmaza is az, a közöttük megalkotható leképezés is egy elméletben létező „valami". Szükség van még a mérés konkrét kivitelezését lehetővé tevő algoritmusok, cselekvések megadására. Vigyázat, ezek már nem matematikai precizitással bíró műveletek, minden részlet esetén *hibalehetőségek merülnek fel*, s a méréselméleteknek ezért egyik nagyon fontos fejezete a hibáknak, és azok hatásainak elemzése (ebben a könyvben ezzel a kérdéssel nem foglalkozom). Először is, szükségünk van az ERR elemeinek megfelelő, tapasztalati világbeli „dolgok" azonosítására. Ez általában nem nehéz feladat, a képességfejlettség mérés esetén kifejezetten könnyű, hiszen egészen egyszerűen: az alaphalmazban emberek vannak. A hosszúságmérésnél is csak egy kicsivel komplikáltabb a kérdés megválaszolása: a szakaszokat a testeken elhelyezett, igen kis méretű jelek párjaival azonosítjuk. Hogy mennyire kell kicsiknek lenniük a jeleknek, az a mérés pontosságával szembeni elvárásoktól függ. Nem is biztos, hogy emberi kéz által létrehozott jelről (vonalról, karcolatról, pontozóval készített jelről, stb.) van szó, lehet, hogy jelként valamilyen természetes, tőlünk függetlenül, az adott felületen létező „pontot" (kis méretű foltot), „vonalat", egy test egy élét, stb. használunk. A „pontok" kijelölése nyilván azért szükséges, mert az ERR matematikai háttérelméletében az alaphalmazt a pontok (a geometria absztrakt entitásai) alkotják[11].

[11] Megjegyzem, hogy a geometria elméletében a pontokhoz nem szükséges semmilyen reprezentációt alkotnunk, nem kell azt mondanunk, hogy a geometria pontjai nagyon-nagyon kicsiny térdarabkák, ideálisan dimenziótlan, nulla méretű valamik, ennek egyébként sincs az égvilágon semmi fizikai értelme. A pontok egyszerűen pontok a geometriában, azt tudjuk róluk, amit a geometria axiómarendszere elárul, s ha sikerülne olyan reprezentációt

Meg kell mondanunk, hogy mit értsünk két szakasz egyenlő hosszán és azon, hogy az egyik hosszabb, mint a másik. Már többször írtam, s még a későbbiekben is szerepel majd, hogy erre nagyon jó gyakorlati eljárásaink vannak, a közvetlen összeméréstől kezdve a segítő eszközök igénybevételéig (mondjuk egy zsineget használunk). Vagyis elég könnyen tudunk algoritmust mondani arra, hogy megmondjuk, milyen reláció érvényesül két szakasz (rúd, út, stb.) között. A konkatenáció kivitelezése is egyszerű. Ha rudakról van szó, akkor az egyik rúd egyik végéhez helyezzük a másiknak az egyik végét, a második rudat az első folytatásaként helyezzük el, vigyázva, hogy egy egyenesben legyenek, s a konkatenáció eredménye az így kapott egyesített rúd. A mérendő „szakasz” az első rúd szabad vége és a második rúd szabad vége, mint két „pont” által meghatározott „szakasz”.

Kijelölhetjük az egységet (csak emlékeztetek rá, ennek az a jelentősége, hogy a végtelen sok hozzárendelés közül egyet kiválasszunk). Biztos sokan emlékeznek: Párizs mellett, a Sèvre levéltárban őrzött, platina-iridium ötvözetből készült ősméter két karcolata közti távolság az egység. Modernebb módon is definiálható, atomfizikai jelenségeket kihasználva. A lényeg azonban az, hogy bármelyik eszközön a megfelelő egység ugyanolyan hosszú. Az 1 m egy bolíviai iskola méterrúdján ugyanakkora, mint Oroszországban egy barkácsoló által használt mércén, stb. Hogy ez egyáltalán nem biztos? Igen, ez így van, lehetnek pontatlanságok (a szabás-varrás közben használt ún. „centik” összehasonlítása például érdekes eredményekre vezet általában), ez azonban a lényegen nem változtat, a pontosság növelésének lényegében nincs soha elvi akadálya.

Hogy lehet megvalósítani magát a mérést? Hogy mérjem le tetszőleges két „pont” távolságát (rúd, vagy út hosszát, stb.)? Csak az alapismereteket használhatom, vagyis azokat az összefüggéseket, amelyek az axiómáknak felelnek meg a gyakorlatban. Ha van egy méterrúdunk, amely, tegyük fel, kellő pontossággal ugyanolyan hosszúnak tekinthető, mint az 1 m-nek választott hosszúságú „szakasz” (Párizs mellett…), akkor megnézhetjük, hogy a mérendő szakasz (ezután nem teszem már idézőjelek közé, ami nem jelenti azt, hogy a geometria absztrakt szakaszaira kell gondolni) hosszúsága nem egyezik-e meg éppen az 1 m hosszú szakasz hosszúságával. Ha igen, vagyis az összemérés azt mutatja, hogy éppen egyenlő hosszúak (a mérés hibahatárán belül), akkor vége a mérésnek, a mérendő szakasz 1 m hosszú. Ha a mérendő szakasz 1 m-nél hosszabb (ezt is meg lehet állapítani, azt mondtuk, hogy a hosszabb-rövidebb összehasonlításra vannak eljárásaink), akkor még megvizsgálhatjuk, hogy több 1 m-es szakasz konkatenációjával nem jutunk-e olyan szakaszhoz, ami egyenlő hosszú a mérendő szakasszal. Ha véletlenül szerencsénk van, akkor ez előfordulhat, és használva a konkatenációval kapcsolatos szabályokat, a mérendő szakasz hossza annyi méter, ahány 1 m-es szakaszt kellett konkatenálnunk. A mércéken a konkatenációk már eleve adottak (pl. milliméteres szakaszok vannak konkatenálva), így előfordulhat, hogy befejezhető a mérés.

De tegyük fel, hogy még ez sem igaz, nem fejezhető ki az egység egész számú többszöröseként a mérendő mérték! Az eljárás ekkor a következő. Készítünk ugyanolyan hosszú rudakat (ha éppen rudakat mérünk), mint amilyen a mérendő rúd hossza. Mindig fel kell tételeznünk, hogy ennek nincs akadálya, a megfelelő konkrét méréselméleti leírásokban ennek axiómák felelnek meg. A gyakorlatban megtehetjük, hogy egy rúddal egyenlő hosszú zsinegdarabokat készítünk, sokat, mondjuk *k* db-ot. A gyakorlatban ezt kivitelezhetjük úgy, hogy egy nagyon hosszú zsinegdarabon az egyik végpontjától kezdve

(precízebben modellt) alkotni a geometria axiómarendszerére, amelyben a pontok szerepét a tapasztalati világunkban szekrényeknek nevezett valamik játsszák, akkor ez a reprezentáció működne. A geometria axiómarendszere azonban rendkívül hasznos, amikor a geometria pontjait a tapasztalati világunkban létező nagyon kicsi dolgoknak feleltetjük meg, s működtetve az axiómarendszert ezeknek a nagyon-nagyon kicsi „dolgoknak” a viszonyairól tudunk meg valamit, amikor kihasználjuk a megfeleltetést.

kijelölünk „pontokat", amelyek esetén két szomszédos „pont" távolsága megegyezik a mérendő szakasz hosszával. A zsineg persze ki van feszítve. Ha a mérendő hosszt h-val jelöljük, akkor a konkatenációra vonatkozó szabály miatt a k db h hosszúságú zsinegdarab konkatenációja egy kh hosszúságú zsineget jelent. Most kezdjünk el egyre több 1 m-es szakaszt konkatenálni (1 m-es távolságban megjelölni „pontokat" egy másik zsinegen), és mindig nézzük meg, hogy a következő 1 m-es darab hozzá vételével még rövidebb szakaszunk van-e, vagy már éppen átléptük a kh hosszúságú szakasz hosszát. Tegyük fel, hogy ez a K-adik lépésben még nem következik be, de a $K + 1$-edikben már igen. Ez azt jelenti, hogy a K méter hosszú szakasz még rövidebb, vagy esetleg egyenlő hosszú a mérendő szakasz hosszának k-szorosával, a $K + 1$ méter hosszú szakasz azonban már hosszabb. Vagyis

$$K \text{ méter} \leq kh \text{ méter} < K + 1 \text{ méter}.$$

Hagyjuk el a méter mértékegység kiírását, és osszunk végig k-val. Ezt kapjuk:

$$\frac{K}{k} \leq h < \frac{K+1}{k}.$$

A k értékét, vagyis hogy hány a mérendő hosszúsággal megegyező hosszúságú zsinegdarabból rakjuk össze a hosszú zsineget, mi választhatjuk meg, a K-t a leírt folyamatban határozzuk meg, vagyis mind a K/k, mind a $(K + 1)/k$ hányados kiszámolható (tetszőleges pontossággal), s mint látjuk, e két szám között kell lennie a mérendő szakasz hosszának, h-nak. De a két szám közötti különbség $1/k$, vagyis k növelésével nagyon kicsivé tehető, tehát a K/k hányados akár nagyon kis eltéréssel becsülheti a h hosszt. Amint látható, a folyamat nem feltétlenül szolgál pontos mérési eredménnyel, azonban nincs akadálya annak, hogy – legalábbis a technikai problémák megoldása után – a pontosságot tovább növeljük, ehhez csak az kell, hogy még több szakaszból rakjuk össze a mérőzsineget (még nagyobb legyen a k).

A leírt példa egy konkrét megvalósulása a reprezentációs méréselméletnek. Konkréták az alaphalmazok, a relációk. Csak két reláció van, ezek közül az egyik egy rendezési reláció, a másik a konkatenáció. Ez a típus több esetben is előfordul a fizikában, és gyakorlatilag megfeleltethető a klasszikus méréselméleti hagyományoknak. A klasszikus méréselmélet tehát a rendezett és additív struktúrák elmélete. De mint láttuk, már a fizikában sem igaz, hogy minden mennyiség additív. Más mennyiségek esetén mások lesznek a struktúrákban szerepet játszó relációk, de a méréselmélet általános megállapításai azokra is érvényesek, ha kielégítik az alapkövetelményeket.

Minden mérésnél az itt leírtakhoz hasonlóan kellene eljárni. Látható, hogy a méréselméleti szempontból korrekt mérés kialakításához jelentős elméleti bázis szükséges, és eredményes empirikus eljárásokat kell létrehoznunk, gyakran ugyanazon tulajdonság esetén meglehetősen különbözőket (mennyire mást jelent „sublerrel" megmérni egy munkadarab egy hosszadatát, és mondjuk valamelyik bolygónak a Földtől való távolságát, miközben mindkettő hosszmérés). Látjuk majd, hogy ezeket a követelményeket a képességfejlettség mérése nem teljesíti, illetve, alapos munkával is csak rendkívül egyszerű képességekre vonatkozóan tudunk kialakítani megfelelő rendszereket, homomorfizmusokat és empirikus eljárásokat.

Itt most extenzív struktúrát használó mérésre adtam példát. Kiválasztásának indoka a könnyebb érthetőség volt, szemléltetni szerettem volna az alapelvek érvényesülését. Egyrészt ezek a struktúrák a pedagógiában, pszichológiában nem játszanak szerepet, mert nincsenek additív mennyiségek (a konkatenáció nem értelmezhető egyik relációként). Másrészt azt is meg kell jegyeznem, hogy természetesen a mérések leírt módja (a *kh* hosszúságú zsineg, a méteres darabok konkatenálása, stb.) a valóságban nem fordul elő, senki sem mér így hosszúságot. Az adott feladattal kapcsolatban van egy pontossági követelmény, mondjuk, egy ácsnak milliméter pontossággal kell dolgoznia. Az ő igényeit kielégíti egy milliméter beosztású mérce, amelyen – ahogy volt róla szó – 1 milliméter hosszúságú „rudak” vannak már eleve konkatenálva, és megelégszik azzal, hogy megnézze, a mérendő hossz hány darab egy milliméteres „rúd” konkatenálásával keletkező „rúd” hosszával lesz egyenlő, az 1 milliméteres pontossággal. De a módszert, amit leírtam, mégsem kell elvetnünk, mint gyakorlati eljárást. Például iskolában érdekes feladat lehet a szokásos A4-es lapok vastagságának meghatározása. Van ugyan olyan mérőeszköz, amely erre alkalmas (mikrométer), de nincs szükség feltétlenül az alkalmazására, lehet, hogy sok iskola nem is rendelkezik ilyennel. Ám ha egy csomagban lévő, 500 darab A4-es lap együttes vastagságát mérjük meg milliméter pontossággal (a matematika órákon használt vonalzóval), és a kapott értéket elosztjuk 500-zal, akkor kb. századmilliméteres pontossággal tudjuk megadni egy lap vastagságát. És ebben a megoldásban éppen a fentebb leírt módszer alkalmazása történik.

A pedagógiának és a pszichológiának elsősorban különbségi struktúrákra lenne szüksége (annak érdekében, hogy az adatok intervallumskálán helyezkedjenek el, s alkalmazhassuk a hatékony statisztikai eszközöket). Hogyan nézne ez ki a *képességfejlettség mérése* esetén? Itt használhatjuk a hosszmérés analógiáját.

Abból indulunk ki, hogy a képességfejlettségek különbségi struktúrát kell, hogy alkossanak. A különbségi struktúrákat egy reláció jellemzi, ez egy általában *D*-vel jelölt, négyváltozós reláció. A különbségi struktúra elméleti leírásában a struktúra axiómarendszere pontosan megfogalmazza, hogy e *D* reláció milyen tulajdonságokkal kell, hogy rendelkezzék. (Később még én is megmutatom ezt az axiómarendszert, amelynek leírása megtalálható például itt: Roberts 1979.)

Vagyis szükségünk lenne egy elméletre (a hosszmérés esetén ez a geometria volt, mondjuk a *Hilbert*-féle axiómarendszer, vagy annak *Birkhoff* által megalkotott változata), amely definálhatóvá tesz egy négyváltozós relációt, amiről az axiómák felhasználásával be tudjuk bizonyítani, hogy rendelkezik a különbségi struktúra leírásában megadott *D* relációtól megkövetelt tulajdonságokkal. Majd látjuk, hogy a klasszikus tesztelmélet nem teszi lehetővé, hogy ezt megtegyük. A modern tesztelmélet és az együttes mérés elmélete igen, de majd azt is látjuk, hogy rendkívül erős feltételekkel, amelyek meggátolják, hogy a pedagógiát jobban érdeklő komplex képességek esetén működjék az értelmezés. Most tegyük fel, hogy olyan képesség fejlettségének méréséről van szó, amely képesség esetében igaz, hogy a modern tesztelmélet, és/vagy az együttes mérés elmélete megfelelő hátteret jelent, tehát hogy a négyváltozós reláció definícióját lehetővé tevő elmélet létezik.

Sikerült tehát belátnunk – most még csak elvben –, hogy az adott esetben a képességfejlettségek struktúrája különbségi struktúra. E struktúra esetén is van természetesen reprezentációs tétel, unicitási tétel, és meg tudjuk vizsgálni, hogy milyen műveletek lehetnek értelmesek ezen a struktúrán. Később a részleteket is megmutatom, de a lényeg az, hogy ennyi már elég ahhoz, hogy a képességfejlettség mérhető legyen.

A konkrét mérés feladatok megoldásával történik, ennek az az oka, hogy maga a *D* reláció is feladatok megoldásához kapcsolódik (ld. később a modern tesztelméletről szóló fejezetben). Kialakíthatjuk tehát a mérés gyakorlati módját, és ezzel minden követelményt teljesítünk. Még egyszer: ez a program tényleg végrehajtható, ha nem is a klasszikus tesztelmélet keretei között, de a modern tesztelméletet és az együttes mérés elméletét alkalmazva. Végrehajtható a program, azonban csak viszonylag egyszerűbb esetekben. És ez lesz a nagy probléma: ha nem is állíthatom, hogy a pedagógia számára érdektelenek a megoldások, de az alkalmazhatóság igencsak szűkkörű lesz.

Általánosítás

Milyen *általánosításokat* hajtanak végre a reprezentációs méréselmélet kidolgozói? Az elvont matematikai axiómarendszer jó néhány technikainak nevezhető, bár a pontos kifejtést mindenképpen segítő változáson ment keresztül. Az általános axiómarendszer nem csak egyetlen reláció fennállásáról szól, általánosan n számú, a kvalitatív, empirikus rendszeren értelmezett-, és k számú ($k \leq n$) matematikai relációról beszél. Talán a legfontosabb fejlemény, hogy a reprezentációs méréselmélet axiómarendszere semmilyen előfeltételezést nem tartalmaz a *skálatípusokkal* kapcsolatban, a különböző skálák mintegy eredményként „kijönnek" az elemzésből, és egyben egy általánosabb és elvontabb definíciót kapnak (Luce és Narens 1984). Az általánosítást szolgálta, hogy a méréssel kapcsolatos matematikai struktúra a reprezentációs méréselméletben valójában bármilyen matematikai rendszer lehet.

Néhány matematikai természetű, azonban a gyakorlati mérési problémákat is érintő kérdést vetettek fel a matematikusok, és igyekeztek azokat megoldani. Így például az axiómarendszer konkrét modelljeiben fontos szerepet játszó relációkra vonatkozó kikötések enyhítésével foglalkoztak (pl. Menestrel és Lemaire 2006). A feladat illusztrálásaként álljon itt egy egyszerű példa: miképpen kezelhető a hosszméréssel kapcsolatos empirikus tapasztalat, hogy ha mondjuk két léc hosszúságát egy hibahatáron belül azonosnak tekintettük, és a másodiknak egy harmadikkal is azonosnak találtuk a hosszúságát, akkor még nem biztos, hogy az elsőt a harmadikkal összehasonlítva is azt mondanánk, hogy azok egyenlő hosszúak. Pl. ha a 0,5 mm-nél kisebb különbségek esetén azonosnak tekintjük a lécek hosszúságát, és a második az elsőnél 0,3 mm-rel hosszabb, s a harmadik ugyanígy 0,3 mm-rel hosszabb a másodiknál, akkor mivel a harmadik az elsőnél 0,6 mm-rel hosszabb, nem tekintjük egyforma hosszúnak őket. Az 'egyenlő hosszú' reláció ilyenkor nem tranzitív, vagyis nem ekvivalencia-reláció, ami alapvetően megváltoztatja az elemzést. Ugyanilyen matematikai kérdésfelvetés, de szintén gyakorlati következményekkel, hogy az axiómarendszer alapján bizonyítható, hogy az entitások X halmazának végtelennek kell lenni. A gyakorlatban azonban számtalanszor találkozunk véges halmazokkal, a pedagógiában szinte kizárólagosan.

Van egy kiterjesztés, általánosítás, amely a mi munkánkat is közvetlenül érinti, ez az „*együttes mérés*" („conjoint measurement") elmélete (Luce és Tukey 1964, Bouyssou és Pirlot 2005). Abból a kérdésfelvetésből indul, hogy vajon lehetséges-e kvantifikálni olyan kvalitatív jellemzőket, amelyek vagy ugyanazon empirikus entitásokhoz kapcsolódnak (azoknak különböző tulajdonságait írják le), vagy külön-külön mindegyik a mért entitások egy-egy halmazának elemeihez, és együtt határoznak meg valamely rendezhető, sorrendbe állítható értékekkel rendelkező tulajdonságot. *Michell* példájával élve (Michell 1999): tegyük fel, hogy egy kognitív teljesítményt vizsgálunk (ez az, amire más jellegzetességek hatnak). Tegyük fel, hogy ennek a teljesítménynek az értékei sorba rendezhetők, és tegyük fel, hogy a teljesítmény kialakításáért a szintén sorba rendezett értékekkel rendelkező képességet, valamint a motivációt tesszük felelőssé. Tudnunk kell még, hogy ugyanazt a teljesítményt a képesség és a motiváció milyen „párjai" alakítják ki, illetve további, matematikailag jól definiált ismeretekkel kell még rendelkeznünk. Az együttes mérés eljárása ilyen feltételek mellett lehetővé teszi, hogy mind a teljesítmények, mind a képességek, mind a motivációk halmazán intervallumskálát hozzunk létre.

Talán világos lehet, hogy ez a kérdés miért oly fontos a képességmérésekkel összefüggésben. A pszichológusok rendszeresen azt a kritikát kell, hogy elviseljék és kivédjék, hogy mentális jelenségeket leíró változóik legfeljebb ordinális skálán mérhetők (pl. Meehl 1991). Ugyanakkor a kutatók rendszeresen használnak olyan módszereket, amelyekhez – ezt már többször is említettem – a változóknak intervallum-szinten értelmezetteknek kellene lenniük. Az együttes mérés elmélete egy lehetséges megoldásnak látszott a problémára, és sokak számára ma is az. Különösen akkor, ha meggondoljuk, hogy a modern tesztelméletben egyrészt a felmért vizsgált személyek képességfejlettségi szintje, másrészt a megoldandó feladat nehézségparamétere határozza meg, hogy milyen valószínűséggel oldják meg majd a felmérésben résztvevők a feladatot. Ez éppen az együttes mérés szituációja (Perline és Wright 1979). Úgy tűnhet tehát, és a hivatkozott irodalom pontosan ezt állítja, hogy minden probléma megoldódik. Ennyire azért sajnos nem egyszerű a helyzet, de ezt később, a saját elemzésem kifejtése során szeretném megmutatni.

A reprezentációs méréselmélet kidolgozásaként, finomításaként felfogható fejlemények azonban alig érintették a pszichológiát, a pedagógiát meg szinte egyáltalán nem. Az együttes mérés elmélete sokkal inkább a közgazdaságtudományban játszik szerepet. Egyébként is igaz, hogy számos esetben a matematika alkalmazása iránti igények közgazdasági elméleti kérdések kapcsán vetődnek fel. Van olyan vélemény, miszerint (legalábbis az Egyesült Államokban) a közgazdászok egyetemi képzésében jóval nagyobb szerepet kap a matematikai ismeretek elsajátítása, mint a pszichológiát tanulók esetében, ezért a kétségtelenül magasabb matematikai felkészültséget igénylő elemzések, módszerek befogadása a közgazdászok körében könnyebben megvalósulhat.

A klasszikus- és a reprezentációs méréselmélet összevetése

A klasszikus méréselmélet a *matematikai platonizmushoz* áll közel, míg a reprezentációs méréselmélet a *strukturalista iskolához*. A klasszikus méréselméletben a valós számok azért rendelkeznek azokkal a tulajdonságokkal, amikkel rendelkeznek, mert ezek a tulajdonságok a világ objektumainak és azok folyamatainak a természetében vannak benne. Természetesen magunk nem közvetlenül ezekkel az ideákkal kerülünk kapcsolatba, csak a „barlang falán vetített képet” van módunkban vizsgálgatni, s így a mérés sem egy – legalábbis ebben az értelemben vett – objektív ismeretszerzés, de itt valóságosan létező matematikai objektumok, valós számok a kiindulópontjai a mérési folyamatnak. A matematikai strukturalizmus ezzel szemben nem hisz abban, hogy a matematika tárgyai, pontosabban azok a „valamik”, amiknek a struktúráját írja le a matematika, valóságosan léteznének. Valóságosnak csak a struktúrát képzeli el, a viszonyoknak azt a rendszerét, amelyben a tényleg valóságosan létező dolgok, és folyamataik egymással állhatnak. Az empirikus relációs rendszerre jellemző egy struktúra, és hozzá tudjuk rendelni azt a matematikai struktúrát, amely pontosan megfelel ennek, ugyanolyan relációkat tartalmaz, ugyanolyan tulajdonságokkal. A strukturalista matematika ismeri azt a struktúrát, amit a valós számok struktúrájának hív, de nem azért, mert elfogadná, hogy létezik a világban egy olyan valami, ami a valós számok halmaza, és az rendelkezne ezzel a struktúrával, hanem egyszerűen csak „megszokásból” használja ezt a kifejezést. Indokoltabb lenne valós számstruktúráról beszélni, de kicsit „nyögvenyelős” ez a szóhasználat. A reprezentációs méréselmélet azt mondja, hogy vizsgálva az objektív valóság tárgyait és folyamatait, ezen belül is vizsgálva valamilyen kitüntetett tulajdonságot, azt vesszük észre, hogy e tulajdonság egyes tárgyakra jellemző értékei olyan jellegzetességekkel rendelkeznek, amelyek tükrözik a valós számstruktúra (maradjunk egyelőre ennél a megnevezésnél) sajátosságait. A reprezentációs méréselmélet azt mondja, hogy a dolgok közti viszonyokban megtalálja ezt a struktúrát. És ezt használja ki a mérés során. Még jobban egyszerűsítve: a klasszikus és a reprezentációs méréselmélet között az az egyik legfontosabb elvi különbség, hogy a klasszikus elmélet szerint *a számok ott vannak az objektumokban és folyamataikban*, mi mintegy kihozzuk őket, amikor mérünk, míg a reprezentációs méréselmélet szerint *az ismert struktúrát vesszük észre a dolgok és folyamataik viszonyrendszerében*, s ezt a struktúrát hozzárendelve a dolgokhoz, folyamataikhoz, végezzük el a mérést, és minden tevékenységet, ami hozzá kapcsolódik.

Az itt csak nagyon vázlatosan, és minden formalizálás nélkül ismertetett reprezentációs méréselmélet is rendkívül erősen kötődik a megismerés, a tudományos tevékenység *objektivista* megközelítéséhez. Az objektivista megismerési szemlélet hívei hisznek abban, hogy amikor megismerjük az objektív valóságban létező dolgok tulajdonságait, azok mértékeit, és megismerjük a tulajdonságok, illetve a dolgok közötti szükségszerű viszonyokat, vagyis az összefüggéseket, akkor olyasminek a tükörképe jön létre bennünk, illetve a tudomány objektivált ismeretrendszerében, ami immanens módon ott van magában az objektív valóságban, nekünk „csak" az a dolgunk, hogy onnan kiolvassuk. A rúd maga, a rúd hossza, és hogy annak mekkora a mértéke, tőlünk függetlenül is léteznek, objektívek. A róluk alkotott tudás tükörkép, vagyis megfeleltetésben van az eredetivel, és ez a megfeleltetés annál jobb, minél alaposabb a megismerés, minél részletesebben, minél inkább sokoldalúan tárjuk fel a dolgok természetét. Ismereteink valóságos, magában a valóságban is létező tulajdonságokat, mennyiségeket, összefüggéseket, viszonyokat, relációkat tükröznek. Nem véletlen, hogy a mérés, vagyis az objektív valóság viszonyainak kvantitatív feltárása, az összefüggések „leolvasását" megkönnyítő rendkívül fontos művelet mindenfajta modern objektivista meggondolás egyik középpontban álló, alapvető eleme. Éppen a mérés teremti meg azt a kapcsolatot, amely a valóság és a róla alkotott tudás között jön létre, s ez a kapcsolat a valóság által alapvetően meghatározott. A valóság mintegy bevetül a tudatunkba, s ennek legfontosabb eszköze a mérés. E szemléletmódban osztozik egymással a klasszikus- és a reprezentációs méréselmélet.

Operacionalizmus

Bár eljutottunk a modern, a szintézist jelentő méréselméletek nagyon vázlatos bemutatásáig, de a történeti folyamatok elemzése során még egy „vonulat" áttekintésére van szükségünk, ha a későbbiekben a képességméréssel szeretnénk foglalkozni. Ez a méréselméleti hagyomány az *operacionalizmus*, amely inkább egyfajta gondolkodásmód, a felmerülő problémák megoldásához való viszony, és nem önálló méréselmélet. Magam úgy látom, hogy alapvetően jellemzi ma is mind a pszichometriai látásmódot, mind azt a gyakorlati és empirikus tudományos tevékenységet, amely a pedagógia keretei között a képességmérésekkel kapcsolatban zajlik.

A 20. század első felében, elsősorban *Bridgman* (1927), valamint *Dingle* (1950) munkássága nyomán formálódott meg az a méréselméleti gondolkodásmód, amely a tudományos terminusok jelentését erősen a tudományos terminussal összefüggő megfigyelésekhez, és az azok eredményeire alapozott, gyakorlatiasan elvégezhető műveletekhez kívánta kötni. *Bridgman* a fizika 20. századi fejlődésének, az új, nagy elméleteknek a newtoni fizikával való szakítását szerette volna értelmezni, a newtoni fizikában született koncepciók hiányosságait kimutatni és megszűntetni. Kiindulópontja az volt, hogy a newtoni fizikában azért merülnek fel hiányosságok, mert e tudomány terminusai nem megfelelően értelmezettek. A megfelelő értelmezés pedig operacionális abban az értelemben, hogy nem abban keresendő, hogy a terminusról az ember mit vél, mit gondol, hanem abban, hogy mit kezd vele, mit tesz vele. Vagyis a teoretikus terminusok értelmezését a megfigyelésekre kell alapozni (verifikácionizmus), és csakis a tudományos gyakorlathoz köthető megfigyelések lehetnek relevánsak (pragmatizmus). Így fogalmaz: „Egy fogalmon semmi mást nem értünk, mint műveletek egy halmazát; egy fogalom a hozzá rendelt műveletek halmazának szinonimája" (Bridgman 1927, 5. o.).

Ugye felismerjük ebben az intelligenciahányados híres-hírhedt definíciójának sémáját: IQ az, amit az IQ tesztek mérnek (Boring 1923, 35. o., 1945, 244. o.)? Világos, hogy a pszichológia számára ennek a gondolatnak az elfogadása erős elköteleződést jelentett a mentalizmussal szemben, és természetesen alapvetően erősítette a behaviorista paradigmát. Van egy eljárásunk, amely produkál valamilyen számokat, ezzel már meg is valósítottunk egy mérést. Megtaláltuk azokat a műveleteket, amelyek valamilyen mennyiséget, változót operacionalizálnak. Sőt, valójában az operacionalizmus magát a mennyiséget azonosítja is mindezekkel a műveletekkel (Borsboom 2006a). Fontos megjegyeznünk *Michellt* (1990) követve, hogy az operacionalizmus egy általános kutatásmetodológiai doktrína, és valójában az a félelem motiválja, amit a kutatók azzal kapcsolatban éreznek, hogy nem tudnak megfelelő kvantitatív változókat létrehozni a jelenségek magyarázatára. Ezért aztán módosul a cél: a tudományos kijelentések hatókörét kell korlátozni a megfigyelés praxisának leírására, szemben a spekulációval. *Michell* keményen fogalmaz: „Természetesen ez a gondolkodás nem tudományos. Bárki, aki fél a hibázástól, és ezért keresi a biztonságos tudományos módszert, soha nem léphet be a tudományos kutatás világába" (i.m. 23. o.).

Az operacionalizmust azért tárgyalom, mert a pszichológiai mérések, közelebbről a képességfejlettség mérések esetén nagyon komolyan felmerül az operacionalizmus „vádja". *Mitchell, Borsboom* és más kritikusok gyakran fogalmaznak így, azt érzékeltetvén, hogy valójában a klasszikus tesztelméletben (és fontos, már többször említettem, hogy a modern tesztelméletben a kritikusok által is elfogadottan nem ez a helyzet) valójában a mérés specifikálja a képességet magát. Egy képesség nem más, mint amit a fejlettségének mérésére kidolgozott teszttel mérünk. Ez színtiszta operacionalizmus, valóban. A helyzet azonban nem ilyen egyszerű. E könyv legfontosabb fejezetében, a képességfejlettség méréséről szólóban részletesen is leírom, hogy ez a kritika nem állja meg a helyét. A képességfejlettségnek még a klasszikus tesztelmélet esetében is van „tisztességes", nem operacionális definíciója (bár ez most sokak számára „kínai": a képességfejlettség a képességhez tartozó feladatokból alkotott tesztben szerezhető pontszám várható értéke). Vagyis a klasszikus tesztelmélet sem vádolható operacionalizmussal, a probléma egészen máshol lesz (a kiépített matematikai rendszerek alapján nem formálható olyan szisztéma, amely megfelelne a reprezentációs méréselméletnek).

A klasszikus tesztelméletnek megfelelő eljárások alkalmazása során mérni szeretnénk a vizsgált személyek valamilyen képességének fejlettségét. A reprezentációs méréselmélet szerint ehhez elengedhetetlen lenne, hogy a vizsgált személyek halmaza az adott képességgel kapcsolatban valamelyik, a reprezentációs méréselméletben konkrétan is felépített kvalitatív struktúrával rendelkezzék. Mivel az a célunk, hogy magas szintű statisztikai vizsgálatokat végezhessünk az adatokkal, azt szeretnénk, ha a vizsgált személyek halmaza az adott képességükkel összefüggésben különbségi struktúrát mutatna. Szükségünk lenne a megfelelő relációra, amelynek rendelkeznie kellene az elméletben leírt tulajdonságokkal ahhoz, hogy itt a számok hozzárendelése után intervallumskáláról beszélhessünk (amennyiben a hozzárendelés valóban homomorf módon jön létre, és bizonyítható, hogy e hozzárendelésekkel kialakuló skálák intervallum jellegűek). E könyv hamarosan következő nagy fejezetében éppen azt szeretném megmutatni, hogy ez ma nem teljesül, és rendkívül kétséges, hogy teljesíthető-e.

A tesztelés „világában” e probléma jelenléte azonban nem evidens tény. Elhomályosíthatja a látásunkat az a körülmény, hogy a klasszikus tesztelméletnek precíz matematikai, axiómákra épülő elmélete van. Ám ez az elmélet nem a reprezenrtációs méréselméletnek megfelelő építmény, empirikus relációs rendszerről a leírásokban szó sincs. Van viszont definíciója (egy axiomatikus háttérrel) annak, hogy miképpen rendeljünk hozzá számokat az emberek képességfejlettségéhez. Akkor működne „tisztán” az operacionalizmus a képességfejlettség mérésében, ha egyáltalán nem rendelkeznénk semmilyen meghatározással a mért mennyiségekkel kapcsolatban, ha valóban azt tennék pszichológusok, neveléstudományi kutatók, hogy lényegében ad hoc módszerek segítségével rendelnének hozzá számokat a vizsgált személyekhez, anélkül, hogy a képességfejlettségnek bármiféle definíciója létezne. Ez azonban igaztalan vád lenne. A képességfejlettségnek igenis van definíciója, még a klasszikus tesztelméletben is, még akkor is, ha nem mindig világos, hogy amit éppen olvasunk egy tesztelméleti könyvben például, az a képességfejlettség meghatározása. A tesztelméletekkel foglalkozó fejezetben részletesen is kifejtem, hogy a klasszikus tesztelmélet valójában megadja, hogy mit mér. Az a probléma, hogy a klasszikus tesztelmélet a korrektül definiált képességfejlettségre nem épít fel olyan matematikai struktúrát, amely a használt mérési skálát legalább intervallumskálává tenné, vagyis nem épít fel különbségi struktúrát. Éppen ezért az így kialakított szám hozzárendelés üres marad, a mérés során kapott számok struktúrája, amelyet bőven kihasználunk a statisztikai számítások elvégzésekor, nem tükröz semmilyen kvalitatív sajátosságokat.

A klasszikus tesztelmélet azonban fokozatosan a perifériára szorul, kiszorítja a modern tesztelméleti logikán alapuló, úgynevezett „itemválasz-elmélet” (Item Response Theory = IRT), amely a hazai, ma még nagyon szűkös szakirodalomban (nem nagyon szerencsés módon) „valószínűségi tesztelmélet” néven vált ismertté[12] (Molnár 2013, 2006). Ez egy kuhni tudományos forradalom, amely nálunk egyelőre nagyon kevéssé érezteti a hatását. Az azonban igaz, hogy a nagy nemzetközi vizsgálatok (PISA, TIMSS, IALS, stb.), a hazai kompetenciamérések, a professzionális keretek között kivitelezett intelligenciamérések, a korrekt metodológiára épülő, tesztek segítségével folytatott kutatások már mind valamilyen IRT modellre épülnek, tehát operacionalizmussal semmiképpen nem vádolhatók (ezt e könyvben később részletesen is bemutatom). Említettem már, és most is hozzá kell tennem mindehhez: a modern tesztelmélet alkalmazásakor olyan követelményeket kell teljesíteni, amelyeket a pedagógia számára fontos, összetett képességek nem teljesítenek, ezért az én meglátásom szerint a modern tesztelmélet egy szigorú metodológia alkalmazása mellett a pedagógiai mérések során nagymértékben használhatatlan. Erre is természetesen csak a részletek tárgyalása során, később térhetek ki.

És az is igaz, hogy számtalan kutatás, vizsgálódás, iskolai értékelő tevékenység zajlik még ma is klasszikus tesztelméleti alapokon. Hazánkban ennek a metodológiának az alkalmazása még mindig túlsúlyban van. Ezért is fontos elemezni az operacionalizmus logikáját követő klasszikus tesztelméleten alapuló méréseket.

Nagyon karcos, és inkább vagdalkozó egyelőre az, amit itt mondtam, ezért, ahogy már többször is ígértem, a későbbiekben természetesen nagyon alaposan alá kell támasztanom az állításaimat. De maga a tétel hadd álljon itt tisztán megfogalmazva: *ami ma a képességmérések területén az esetek egy nagy részében (kivéve a modern tesztelméleten alapuló vizsgálódásokat) történik, az, ha nem is operacionalizmus, nem felel meg a reprezentációs méréselmélet által támasztott igényeknek*.

[12] Azért nem szerencsés a „valószínűségi tesztelmélet” megnevezés, mert a klasszikus tesztelmélet is valószínűségi jellegű, alapvetően a valószínűség fogalmára épít.

Az intelligencia és annak mérése (vagy általában a pszichometria) állatorvosi ló. Ahogy kifejlődött, és működött a legelején, abban „dühöngött az operacionalizmus". Fogalmunk sem volt róla, hogy mi is az az intelligencia (nem, mintha ma sokkal jobban tudnánk). Ahogy megformálódott a klasszikus tesztelmélet, úgy „bújt ki" az intelligencia mérése az operacionalizmus alól. Már mondhattuk azt, hogy nem a mért IQ az intelligencia, hanem „rendes" matematikai definíció állt a rendelkezésünkre. Ám az intelligenciának ezután sem volt még matematikai struktúrája, nem volt olyan ERR, amely e „tulajdonságot" megfelelően specifikálta volna. A 20. század '60-as éveitől kezdődően aztán fokozatosan tért hódított az intelligencia mérésében is a modern tesztelmélet (IRT modellek) alkalmazása, vagyis hihettük, sokan hiszik is, hogy minden rendben van, van már ERR, sikerült „beterelni" az IQ-mérést a modern, reprezentációs méréselmélet alá. Ezt akkor lehet megtenni, ha az adott képesség valóban rendelkezik azzal a struktúrával, amelyet az ERR-ben feltételezünk. Meg fogom mutatni, hogy az ilyen komplex emberi tulajdonságok esetén, mint amilyen az intelligencia is, ez teljességgel elképzelhetetlen. Lényegében azt fogom megmutatni: nemhogy különbségi-, de még rendezési, sőt, nominális struktúra sem építhető.

Az intelligencia mérése körül bontakozott ki a pszichológia történetének talán legösszetettebb vitája, amit akár harcnak is nevezhetünk. Az IQ mérése nagyon sok ember életét befolyásolta, és befolyásolja ma is, az IQ mérés jelentős társadalmi érdekeket érint. Sokan állították ebben a száz éve tartó vitában, hogy önmagában a mérés bizonyos társadalmi érdekek kielégítését szolgálja, s körülötte az értelmezések tisztázatlansága nem is jött rosszul embereknek, akár nagy társadalmi csoportoknak, szervezeteknek sem. Ha az IQ mérés az iskolában – és ez valóság még ma is egyes oktatási rendszerekben – az úgymond „képesség" szerinti szelekciót („tracking", „ability grouping") szolgálja, akkor maga a mérés bekapcsolódik egy olyan folyamatba, amelyben nem a tudományos kérdéseké a főszerep, hanem a szegregációé, a diszkriminációé, az elnyomásé, a meglévő társadalmi előnyök továbbélésének biztosításáé. Ha az IQ mérések az Egyesült Államokba való bevándorlás mértékének csökkentését szolgálják (Gould 1981), ahogyan erre volt példa az USA történelmében, akkor a tudományos eredményeket ezen a módon és e célból felhasználni akarók nem fognak rákérdezni a fogalmak definiálásának mikéntjére, és annak logikájára. Nem lesznek skrupulusaik azzal kapcsolatban, hogy most éppen egy operacionalista gondolkodásmód gyakorlati megvalósulásáról van szó, hanem céljaik elérése érdekében alkalmazzák az IQ mérést úgy, ahogy rendelkezésükre áll, ha kell, még ma is a hagyományos, klasszikus tesztelméleten alapuló metodológiával. Társadalmi értelemben az iskolaalkalmasság felmérését szolgáló többféle teszt adaptív, ha sokak megelégedésére szűrni lehet a halmozottan hátrányos helyzetű gyerekeket, fogyatékosnak nyilvánítva őket. Hogy amúgy semmilyen korrekt, megbízható elméleti háttere nincs e méréseknek, már kevésbé lesz érdekes annak számára, aki a szűrésben érdekelt. Örül, hogy van egy tudományos köntösbe bújtatott módszer, amely az igényeit kielégíti.

Vagyis a legkülönbözőbb pszichológiai és pedagógiai mérések úgy lesznek társadalmilag adaptívak, hogy valamilyen jól érzékelhető, *létező társadalmi igényt elégítenek ki*, és ez legitimálja, adaptívvá teszi a mérések elméleti megalapozását jelentő eljárásokat is, amennyiben azok a klasszikus tesztelméleten alapulnak.

Valamit javult a helyzet azzal, hogy számos ilyen mérésben, és a professzionális IQ tesztek alkalmazása során a legtöbb esetben már a modern tesztelméletet alkalmazzák. Két okból sem lehet azonban felhőtlen az örömünk. Egyrészt a modern tesztelméletben – ezt magam is bemutatom később – a reprezentációs méréselméleti háttérnek egy a pszichológiai tényekkel rendkívül nehezen kapcsolatba hozható megalapozása történik, és ezért kell azt mondanom, hogy ma is bizonytalanok vagyunk abban, hogy mi is az IQ. Definiálható különbségi struktúra a képességfejlettségek (a felmért személyek) halmazán, de annak nem adható szemléletes pszichológiai jelentés. Ez egy súlyos validitási probléma (azt mérjük-e, amit mérni szeretnénk?). A másik gond, amiben biztos nagyon sok kutató nem ért velem egyet, az, hogy az IQ mérésben, de nagy valószínűséggel minden komplexebb emberi képesség fejlettségének mérése során nem teljesülnek az IRT modellek alkalmazásának elemi feltételei sem. Ezt mutatom meg a képességfejlettség méréséről szóló fejezetben a későbbiekben.

A mérés episztemológiai státusa

Miért és mennyiben ismeretelméleti kérdés a mérés? Egy példa

Először is kíséreljük meg világosabbá tenni, hogy a mérés egy elvont értelemben miért *ismeretelméleti kérdés*! Induljunk ki abból, hogy a vizsgált „valamik" halmazán értelmezett tulajdonság értékeihez tartozó mértékek meghatározása, vagyis a mérés számos önkényes elemet tartalmaz! Vagy legalábbis olyat, ami más területeken kialakított meggyőződéseinktől, elkötelezettségeinktől, vagy egészen egyszerűen bizonyos elhatározásainktól függenek. Vizsgáljuk meg közelebbről, hogy amikor az M (vagyis a matematikai, a legtöbb esetben a valós számokat tartalmazó) halmaz elemeit rendeljük hozzá az X halmaz elemeihez (az empirikus relációs rendszer, az ERR elemeihez), illetve amikor ennek a hozzárendelésnek az empirikusan adott objektumokon, jelenségeken, folyamatokon végrehajtható megfelelőjét kivitelezzük, vagyis mérünk, akkor azt vajon miért éppen úgy tesszük, ahogy tesszük! Elemezzük ezt a kérdést egy már sokszor használt példánkon, a hosszúság mérésén! Az itt következő leírás egyes részletei ismétlésként hatnak majd, mert valóban szerepeltek már, most azonban új összefüggésekre szeretném felhívni a figyelmet a már jól ismert példa kapcsán.

A hosszúság mérésekor a rudakhoz rendelünk hozzá számokat. A mérendő entitásokat tartalmazó halmaz a rudak halmaza. Az M halmaz a valós számok halmaza. Mint már volt róla szó, a rudak (vagy valamilyen más módon kijelölt, meghatározott anyagi objektumok) „mögé képzeljük" a geometria által egzakt módon leírt, axiomatizált geometriai teret, annak pontpárjait, vagy másképpen a szakaszokat. A gyakorlatban, vagyis az empirikus objektumok esetén is gyakran nem rudakról, hanem „szakaszokról", pontpárokról van szó, valamely testen (vagy testeken) kijelölt két „pont" közti távolság meghatározását kell elvégeznünk. Ha egy ház egyik oldalának hosszát akarjuk lemérni (mondjuk egy járda megtervezéséhez, az anyagszükséglet kiszámításához van szükségünk erre a méretre), akkor sem egy rúd hosszát mérjük, hanem a házon alkalmasan kiválasztott két „pont", a „sarkok" közti távolságot.

Végtelen sokféleképpen lehet számokat hozzárendelni a tér szakaszaihoz. Kezdetben ebben nem korlátoz bennünket semmi. Az egyik természetes módon felmerülő korlát, hogy egyenlő szakaszokhoz ugyanazt a számot kell hozzárendelni. De mi az, hogy két szakasz a hossz szempontjából egyenlő? Volt már erről szó, azt mondtuk, hogy ez az „egyenlőség" az X halmazon egy elméletileg definiált reláció, miközben a gyakorlatban, a valóságos rudak körében a reláció fennállásának vagy nem fennállásának ellenőrzése empirikus módszerekkel történik. Ha rudakról van szó, akkor a rudakat egymás mellé állítjuk, és szemmel, tapintással, vagy más módszerekkel ellenőrizzük, hogy az egymásnak megfelelő végpontok nagyon közel vannak-e egymáshoz. Ha a tér pontpárjairól beszélünk, akkor azokat nem lehet egymáshoz közelíteni, nem hozhatjuk fedésbe a szakaszokat. Mondhatná valaki, hogy – maradva a gyakorlati megoldásoknál – húzzunk ki a két pont között, megfeszítve egy fonalat, amely nem nyúlik. Jelöljük meg a fonáldarab két azon pontját, amelyek egybeesnek a vizsgált szakaszunk végpontjaival. Ezután könnyű ellenőrizni, hogy vajon bármely más szakasz egyenlő hosszú-e ezzel az eredeti szakasszal, egyszerűen a fonálon megjelölt egyik pontot az új szakasz egyik végpontjába kell vinni, és meg kell vizsgálni, hogy a kifeszített fonál másik megjelölt pontja fedésbe hozható-e a szakasz másik végpontjával. Ez természetesen a gyakorlatban járható út, mi azonban a hozzárendelés módját vizsgáljuk, vagyis a gondolkodásunk logikáját, alapjait szeretnénk megérteni e kérdésben, ezért nagyon kritikusnak kell lennünk a tekintetben, hogy vajon egy-egy ilyen művelet esetén nem használunk-e fel már valami olyasmit, amit csak később fogunk bebizonyítani, vagyis gondolkodásunk nem forog-e körbe.

Természetesen nagyon sok mindent felhasználunk hallgatólagosan. Elsősorban azt, hogy miközben a zsineget elmozdítjuk, hogy a két megjelölt pontja közti szakaszát összemérjük a mérendő szakasszal, eközben nem változik a hossza. Rendkívül adaptív az a gondolat, hogy a gyakorlati alkalmazások során használt mérőeszközök, összehasonlításra ad hoc módon használt tárgyak, fonalak, rudak, mércék hossza a mozgatás közben nem változik, a mérnöki gyakorlat egésze nyugszik ezen és sok más hasonló feltételezésen. De hogy ez valóban így van, az csak egy hipotézis. És nem is igaz olyan helyzetekben, amikor a két pontpár (vagy a két rúd) egymáshoz képest mozog (nem mindkét szakaszra merőleges irányban). A relativitáselmélet állítása szerint az összemérésre használt eszköz hossza más lesz az egyik pontpárhoz (rúdhoz) rögzített koordinátarendszerben, mint a másikban, ha azok egymáshoz képest mozognak. Jelentős különbségek azonban csak óriási sebességek esetén vannak, ezért a nem szélsőséges helyzetekben releváns mérnöki gyakorlatban ez az effektus teljességgel elhanyagolható.

Azt is felhasználjuk a leírt folyamatban (egy zsineg segítségével pontpárok közti távolságok egyenlőségének ellenőrzésekor, illetve két rúd összemérésekor), hogy a pontpárok távolsága egy értelmes kifejezés, mint ahogy a rudak hossza is az. Pedig ez nem ilyen egyszerű. A gyakorlatban egy „pontot” valamilyen felületen tudunk kijelölni, vagy a felületnek van valamilyen „kicsúcsosodása”, amelynek a hegyét tekintjük egy „pontnak”. De mindenki tudja, és korábban egy részelemzésben már én is használtam ezt a tudásunkat, hogy itt természetesen szó sem lehet a matematika által tárgyalt, dimenziótlan pontokról. Mi egy rúd hossza? Hová helyezzem a mérce egyik végét? A rúd végéhez? Mi az? A rúdnak egy mikrofizikai értelemben nincs vége. Ha használunk egy rendkívül jó nagyítót, akkor a rúd végét egyenetlennek látjuk. Ennek az egyenetlenségnek melyik részéhez kell pontosan illeszteni a mérce végét? És a másik vég egyenetlenségeit figyelve annak mely részénél kell leolvasni a hosszat, ott is használva a nagyítót? Vegyük a legszélső pontokat? Jó (bár megkérdezhetjük: miért éppen azokat). De ha a nagyítást tovább fokozzuk, akkor rá kell jönnünk, hogy valójában a rúd végénél atomok vannak, amelyeknek bizony nincs határozott kontúrjuk, elektronfelhőiknek nincs határozott határa. Akkor most hová kell helyeznünk a mérce végét, azét a mércéét, amely szintén határozott „peremmel” nem rendelkező atomokból áll a végénél?

Ennek az egész gondolatmenetnek nincs természetesen az égvilágon semmilyen gyakorlati relevanciája. Egy rúd hosszát, egy munkadarab vastagságát, két karcolat közti távolságot a konkrét gyakorlati igényeknek megfelelő pontossággal kell meghatározni. Ez a pontossági követelmény lehet szigorúbb és kevésbé szigorú. Az atomok elektronfelhőiből adódó bizonytalanság ma biztosan semmilyen gyakorlati szempontból nem játszik szerepet. A felület egyenetlenségei már inkább, de minden ilyen esetben érvényes, hogy a felületek megmunkálásával szembeni követelmények összhangban vannak a mérési pontossággal kapcsolatos követelményekkel. Vagyis ha mondjuk, egy munkadarab valamely méretének meghatározása azért lenne problematikus, mert a mérésnél is szerepet játszó felületek túlságosan egyenetlenek lennének, akkor az a munkadarab nagy valószínűséggel nem lenne megfelelő éppen a praktikus célok szempontjából problematikus felületi egyenetlenség miatt. Azt látjuk tehát, hogy a gyakorlati mérési eljárások alkalmazása során jól elboldogulunk azzal az elképzeléssel, hogy a mérőeszközeink a megfelelő méreteket határozzák meg, mintha itt valóban pontokról, ideális felületekről, egyenesekről, síkokról beszélhetnénk.

Ez a kis eszmefuttatás nagyon fontos dolgokra figyelmeztet. A rudaknak nincs tényleges, valóságos hosszuk, két karcolatnak, vagy hegyes eszközzel ejtett két pontszerű jelnek nincs távolsága. Valójában távolságuk csak az elvont, a geometriában tárgyalt pontoknak van. A geometria axiómarendszereiben megalapozható a fogalom, de nem valóságos testek adataira érvényes módon, hanem teljes mértékben absztrakt formában. Amikor valóságos rudak hosszáról, akármilyen módon kijelölt „pontok" (karcolatok, jelek) távolságáról beszélünk, akkor a matematikai távolság fogalmat mintegy modellként használjuk. Mintha hozzárendelnénk a hegyes eszközzel ejtett két pontszerű jelhez egy-egy „matematikai pontot", és mindazt, amit a „matematikai pontokról" tudunk, azt átvinnénk a fizikai világban manipulált tárgyakra, jelekre, jelenségekre, folyamatokra. A fizikai világban nincsenek pontok, nincsenek egyenesek, síkok, ezért nincsenek távolságok sem, tehát a méréseink sem ezeket határozzák meg (ha egyszer nincsenek, nehéz is lenne).

Mit illusztrál a példa?

Az ember akár több tízezer éven keresztül is gyakorolhatta a mérés „tudományát", ennek segítségével lélegzetelállító kunsztokat produkáló technikai világot volt képes létrehozni. Felmerül a kérdés: hogyan lehet azt állítani, hogy amiket mérünk, azok nem is léteznek? A kérdésre adható válasz kulcsa az, hogy a mérés egy *adaptív emberi tevékenység*. A mérés segítségével a fizikai világról tudást *konstruálunk*; azt működtetve olyan következtetésekhez jutunk, amelyek értelmes, eredményes cselekvéshez vezetnek. A mérési eredmények csak *tapasztalati világunk* objektumaihoz rendelnek hozzá számokat, a valóság entitásaihoz soha nem tudunk hozzárendelni semmit. A tapasztalati világunkban létező tárgyakat, jelenségeket, folyamatokat, embereket, stb. magunk konstruáltuk, sőt, folyamatosan konstruáljuk, ezeket el tudjuk látni pontokkal, egyenesekkel, síkokkal, tömeggel, hőmérséklettel, problémamegoldási képességgel, intelligenciával, előítéletességgel. Belegondoljuk, belekonstruáljuk ezeket a mennyiségeket a tapasztalati világunk objektumaiba. És jól tesszük, mert – legalábbis számos esetben – rendkívül hasznos, rendkívül jól működő konstrukciókról van szó.

Az is nagyon figyelmeztető ebben az elemzésben, hogy a konstrukciónak *elméleti háttér* kell. A hosszmérés nem azért lehetséges, mert a reprezentációs méréselméletben alapvető szerepet játszó empirikus rendszer feltárul előttünk, és mintegy produkálja a szabályokat, az összefüggéseket. A hossz mérése nem azért lehetséges, mert e tevékenység során olyan jól tudjuk összekapcsolni a matematikai struktúráink elemeit és a valóságos objektumokat. Nem ez történik ugyanis. A számokat egy matematikai struktúra, az empirikus relációs rendszer (ERR) elemeihez rendeljük hozzá, az ERR pedig mintegy modellje a tapasztalati világunkban létező objektumoknak. Ahhoz azonban, hogy felépíthessük tapasztalati világunk objektumait, legalábbis az empirista, induktivista, pozitivista látásmóddal szakító ismeretelméletek, elsősorban a konstruktivizmus szerint, *konstruált tudásra* van szükség, s annak pedig megfelelő szerkezetre, vagyis *elméletekre*. Az lehet, hogy egy pontos, az euklideszi teret leíró geometriai axiómarendszer teljes részletességgel és működő formában csak az ezzel foglalkozó matematikusok agyában van meg, de mindenki, aki képes hosszúságot mérni jól, hordoz magában egy naiv tudásrendszert azokkal az absztrakt entitásokkal (pontok, egyenesek, síkok, távolságok, stb.) kapcsolatban, amelyekből felépíti tapasztalati világát. S úgy tűnik, a hétköznapi mérési feladatokhoz ez bőven elég.

A tudománynak természetesen kötelessége, hogy az ezek szerint a mérés folyamatában alapvető szerepet játszó háttértudást explicitté tegye, formalizált rendszert biztosítson számára. Ez történik meg az euklideszi geometria rendszerének kiépítésével. És ennek kellene történni a képességek fejlettségi szintjének mérésével kapcsolatban is. Rendelkeznünk kellene tehát egy olyan elmélettel, amelynek a képességfejlettségek, mint elvont entitások (mint a szakaszok a geometriában) az elemeit alkotják. Azt kell majd megvizsgálnunk, hogy ilyen elmélettel (vagy akár elméletekkel, hiszen geometria sem csak egy van) rendelkezünk-e.

Mérés és tudomány

A mérés ismeretelméleti státusa megvizsgálható a tudományos tudás keletkezése szempontjából is. A hagyományos, pozitivista tudományképben a mérés alapvető szerepet játszik, a valóságra vonatkozó ismeretek közt a konkrét adatok megszerzésével kapcsolatos tevékenység. Ebben a képben a tényekkel kapcsolatos ismereteink megalapozását szolgálja, míg a tényekről így kialakított ismeretek a tudományos ismeretek, köztük az absztrakt elméletek felállításához szükségesek. A pozitivista tudományképben a mérés objektív ismeretekhez juttat bennünket, legalábbis az így szerzett ismeretek fokozatosan egyre jobban megközelítik, egyre inkább hű módon tükrözik a valóságot.

Egy másfajta, elsősorban *konstruktivista* alapokon nyugvó tudományképben a mérésnek nyilván nem lehet ez a szerepe. A konstruktivista megismerés-felfogásban az ismeretek nem objektívek, nem a valóság valamifajta tükörképének elemei, hanem konstruktumok. Ebben a szemléletben a méréssel megalapozott tényismeret nem forrása az elméletek létrehozásának, mert a megismerés logikája itt nem írható le a pozitivista felfogás indukcióra alapozott módján (von Glasersfeld 1995, Nahalka 2002). Mérés azonban itt is szükséges, ám míg a pozitivista megközelítés esetén a mérés és a valóság szoros kapcsolatban vannak egymással, maga a mérés, annak eredménye mintegy „üzenet" az objektív valóságban, addig egy konstruktivista alapokat használó méréselméletben erről szó sincs, magát az egész procedúrát, a mérés lényegét kell új elméleti keretek közé helyezni.

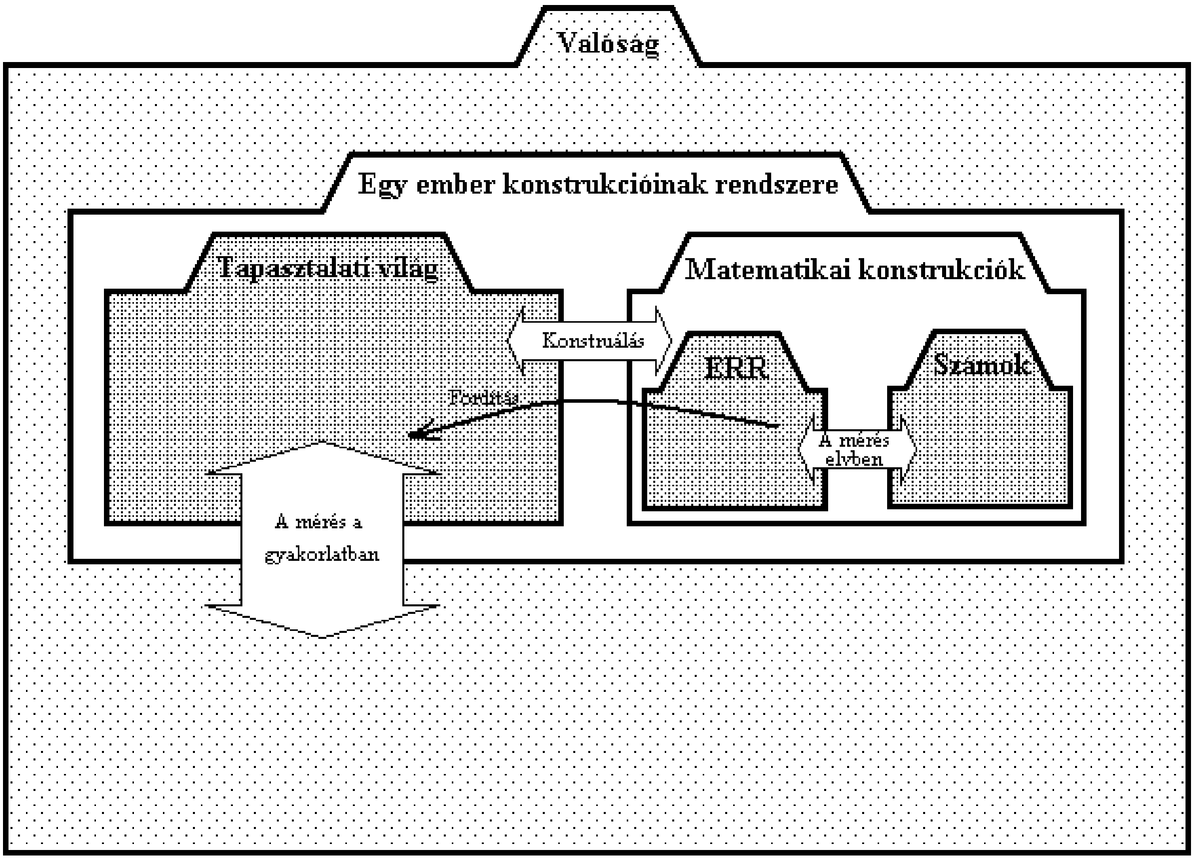
A konstruktivista megismerés-felfogás szerint az ember a világról konstrukciókat alkot, e konstrukciók és az „eredetijük" megfelelése olyan kérdés, amit ez a megközelítés nem az objektivista szemlélet szerint válaszol meg. A konstrukciók *adaptivitásáról* beszélünk, vagyis arról, hogy a konstruktumok illeszthetők-e meglévő világlátásunk (a „nagy konstrukciók") keretei közé, segítségükkel kielégítő, megfelelő gyakorlati tevékenység formálható-e. Egészen elvont értelemben egy tudományos elmélet felállítása akkor jogosult, ha konzisztens, pontosabban: ha van *ellentmondásmentes formális rendszere* a szimbolikus logika értelmezése szerint. Ennek a feltételnek csak nagyon kevés tudományos elmélet felel meg (pontosabban nagyon kevésről tudjuk, hogy megfelel). Ennek egyik oka, hogy a legtöbb tudományos elmélet jelenleg nem rendelkezik kifejtett formális rendszerrel (szimbolikus logikai alapokra épített axiómarendszerrel). Másik oka, hogy még az ehhez közelálló tudományos tudásrendszerek közt is van olyan, amelyről tudjuk, hogy ellentmondásos, ilyen például a klasszikus mechanikára, az elektromágneses jelenségek *Maxwell*-féle elméletére és a fenomenologikus termodinamikára épített fizikai elméletrendszer (miközben a hétköznapokban bátran használjuk). Az adaptivitásnak a gyakorlatban való sikeres alkalmazáshoz köthető követelménye sokszor nem szempont a tudományos vizsgálatok során – szerencsére –, hiszen gyakran fordul elő, hogy kutatók olyasmivel foglalkoznak, aminek közvetlenül semmilyen gyakorlati haszna nincs. A tudományos munka adaptivitása éppen ezért nem egy objektív módon megítélhető jellemző, ami vagy van, vagy nincs, az e kérdésben meghozott ítélet inkább társadalmi természetű, és számos tudományon kívüli tényező is befolyásolja.

Bármennyire is bizonytalan, és bármennyire is komplex társadalmi folyamatok eredményeként formálódik is meg, a tudományos elméletek adaptivitásának fogalmánál általánosabbat és „erősebbet" a konstruktivista felfogás szerint nem alkothatunk. Az „igaz tudományos elmélet" fogalma ebben a szemléletmódban nem létezik, az igaz és hamis megjelöléseket csakis gondolkodási rendszerek keretei közt használhatjuk, s csak abban az értelemben, hogy igaz az, ami nem bizonyított alaptételekből, más igaznak tartott, vagy igaznak bizonyult ismereteinkből egy elfogadott logikai rendszer szerint következik, és hamis az, aminek a tagadása bizonyítható be ilyen módon. A konstruktivista gondolkodásmód szerint az igaz és a hamis szemantikai értékek, mint a valósághoz való viszonyítás során a kijelentésnek adható értékek nem léteznek (pontosabban az elmélet ilyen fogalmakat nem értelmez).

A mérés tehát csakis a tudományos elméletek, paradigmák, kijelentések adaptivitásának eldöntése során játszhat szerepet. Vagyis itt alapvetően egy társadalmi jelenségről van szó, a társadalmi gyakorlat során alkalmazunk méréseket, amelyek a tudomány által szolgáltatott eredmények adaptivitásának lemérésére is szolgálnak. Azért szerepel itt „is", mert természetesen a mérések a tudományon kívül számos más társadalmi tevékenységrendszerben kapnak szerepet, egyébként minden ilyen alkalom egyben tudományos ismeretek adaptivitásának értékelését is jelenti, csak ez az oldala a tevékenységnek nem válik általában nyilvánvalóvá.

Miről beszélünk?

Mindezek után érdemes feltennünk a kérdést: ha egy nem objektivista szemüvegen keresztül vizsgáljuk a méréseket, milyen rendszerek (objektumok és tudásrendszerek) jönnek számításba (ld. 1. ábra).



1. ábra: A mérési folyamat során szerepet játszó rendszerek („világok", halmazok)

Minden rajzos séma nyilván tökéletlen, és általában nem nehéz hibákat sem felfedezni rajtuk, a mérés vizualizálásának itt használt módja is sok problémával küzd. De talán mutat valamit a lényegből. Azt akarja szemléltetni, hogy a mérés során milyenfajta „világok", rendszerek, emberi konstrukciók játszanak szerepet egy nem objektivista elemzés szerint, ha a szóba jövő részrendszereket minél teljesebben akarjuk számba venni.

A mérés az *objektív valóságban* lejátszódó folyamat (természetesen annak a számára, aki a belső rendszerében a világnak egy realista szemléletmódját alakította ki, amelyben tehát létezik objektív valóság). A legtágabb „világban", a valóságban léteznek a vizsgálatba vett valóságos dolgok. Ezek halmaza az, amiről egy nem objektivista ismeretelméleti felfogás alapján nem is beszélhetünk, közvetlenül számunkra nem elérhető. Ezzel egy konstruktivista kiindulópontot használva valójában nincs is dolgunk.

Az egész valóságon belül, annak részeként ábrázolható egyetlen ember (a viszonyokat, a rendszereket és világokat elképzelő, megkonstruáló ember) *konstrukcióinak* rendszere, kognitív apparátusa. Érdemes itt is megkülönböztetni kétféle konstrukciót, miközben e modellnek valószínűleg ez a leggyengébb pontja: vannak a *hétköznapi gondolkodásunkat megformáló részrendszerek*, amelyek a tapasztalati világunkat alkotják, amelyek nem teljes mértékben logikai alapokon építik ki viszonyaikat, s a hétköznapi életben való „mozgásunkat", a tevékenységeinket szolgálják; és van *a matematikai konstrukciók világa* bennünk. A két rendszer nagy valószínűséggel nem különíthető el olyan jól egymástól, mint azt az ábra mutatja, a matematikai gondolatok még egy matematikus fejében sem tisztán egy formális rendszer keretei közt jönnek létre. Csak egy példával illusztrálom ezt az összefüggést: ha egy matematikus egy geometriai pontot képzel el, akkor ugyanúgy, ahogy ezt bármelyik nem matematikus teszi, egy nagyon-nagyon pici, már ponttá zsugorodott „dolgot" képzel el, ami valahogyan ott lebeg a térben. Ezzel a képpel sok probléma van. Hiszen a geometria axiómarendszere egyáltalán nem kell, hogy még csak nagyjából is ilyen dolgokra és csak ilyenekre vonatkozzék. Bármilyen halmaz kielégíthetné a követelményeket, ha olyan relációk, fogalmak, axiómák érvényesülnének benne, mint amilyeneket a geometria valamelyik axiómarendszere előír. Vagyis a kognitív rendszerünk részét alkotó matematikai struktúrák olyan konstrukciók, amelyek a tapasztalati világban meglévő konstrukciókkal együttműködnek. Még talán az is igaz, hogy a matematikai világunk is teljes egészében része a tapasztalati világnak, e ponton tehát a felvázolt séma minimum bizonytalan. Modellként azonban jó szolgálatot tesz, s ha az elemzést a későbbiekben esetleg gátolja, akkor adaptívabb modellt is kereshetünk. Most mindenesetre egy olyan ember konstrukcióit képzeljük el, aki tisztában van a mérés lényegével, ahogy azt a reprezentációs méréselmélet leírja. Vagyis emberünk valóban birtokol fejlett matematikai struktúrákat, ezek közt a konkrét méréshez szükségeseket is. A matematikai világ konstrukció eredménye, ahogy ezt egy nyíl segítségével az ábra is igyekszik szemléltetni, a konstrukciós folyamatokban a tapasztalati világ fontos szerepet játszik.

A mérés bizonyos értelemben „két színtéren lejátszódó" folyamat. A mérés egyrészt „tisztán" matematikai kérdés, amennyiben létrejön konstrukcióval az ERR, létrejön, vagy már adott az a matematikai struktúra, amelyre leképezzük, és létrejön maga a leképezés, az ERR és (konkrétabban) a számok halmaza között, egy *homomorfizmus* formájában. Ez a mérés elvi, matematikai leírása, amelyben még szó sincs arról, hogy megmondjuk „mi mennyi", minek mi a mértéke, itt csak a mérés elvi alapjai tisztázódnak. A mérés másik „terepe" a konkrét mérő tevékenységek rendszere. A matematikai leírás elemeinek konkrét megfelelői ismeretében elvégzünk egy tevékenységsort, melynek eredményeként a tapasztalati világunkban létező dolgok mértékei jelennek meg. E tevékenység eredményeként mondhatunk olyanokat, hogy ez a léc 1 m 20 cm hosszú, vagy hogy ennek a tanulónak az arányszámítási képessége 78%-os fejlettségű. A mérés „két színtere" közt tehát egyfajta „fordítási művelet" létezik, ahogy az 1. ábra is egy nyíllal jelzi. „Lefordítjuk" a tapasztalati világ és a rajtunk kívüli világ interakcióinak „nyelvére" az ERR-ben lévő elemeket, a köztük lévő műveleteket, a hozzárendelést (homomorfizmust).

A *tapasztalati világunkban* van a valós világban létező, a mérés tárgyait képező dolgokról alkotott modelljeinknek a halmaza. Ezzel van „valóságosan" dolgunk, ez az, ami számunkra elérhető. Elemei még nem matematikai objektumok, nem axiómarendszerek határozzák meg őket, hanem a gondolkodásunk egésze, a fogalmaink, a képeink, stb., tehát az emberi kognitív rendszer maga. Ezekről kommunikálunk, ha mondjuk gyakorlati feladataink elvégzése érdekében kell, hogy egymással valamit közöljünk. Ezekről a dolgokról, vagyis tapasztalati világunk elemeiről gazdag tudással rendelkezünk, számtalan asszociáció kötődik hozzájuk. Az így kialakult tudásrendszer logikailag nem teljes mértékben rendezett. Ez alatt azt értem, hogy ez a tudásrendszer nem egy axiomatizált, szimbolikus logikai rendszer, hanem *tudáselemek szövevényes hálózata*, amelyben akár még ellentmondások is megbújhatnak. Ráadásul egy formálódó, alakuló rendszer is nagyon sok esetben. A dolgokat meg tudjuk figyelni, a dolgokkal műveleteket tudunk végezni, a hétköznapi életben mindezekről egy gazdag nyelv segítségével tudunk kommunikálni.

A „dolgoknak" vannak valamilyen *tulajdonságaik*, amelyekkel éppen foglalkozunk, s az éppen vizsgált tulajdonságukat tekintve a „dolgok" különbözők lehetnek. A tömegmérés példájában a „dolgok" tömege egy kvalitatív sajátosság, tehát még nem grammokban, fontokban, latokban megadott számérték. Ha emberek adott képességének fejlettségéről van szó, akkor a tulajdonság a képesség fejlettsége, egyelőre nem használva számokat a tulajdonság-értékek kifejezésére. Különböző „dolgoknak" lehet ugyanaz a tulajdonságuk, de egy „dologhoz" csak egy tulajdonságérték tartozhat. Még mindig a tapasztalati világunkban létező entitások körében mozgunk, azzal a sokszor akár „zavarosnak" is tekinthető „értelmezéssel", körülírással, amelyet az ilyen gondolati objektumokra használunk. A következő lépésben, amikor ezen asszociációk tengerében létező gondolati objektumok modellálása érdekében matematikai struktúrát hozunk létre, akkor válik igazán átláthatóvá, mi az, amit szikár módon, és nagy biztonsággal állíthatunk e halmazokról, az elemeikről.

Már a matematikai konstrukciók világába tartozik a dolgokkal és tulajdonságaikkal, valamint azok értékeivel (tehát a tapasztalati világunkban létező, nem formalizált struktúra keretében elgondolt elemekkel) kapcsolatos, matematikai, logikai szempontból korrekt, formalizált rendszernek az alaphalmaza. Ez alkotja az ERR „testét", jele (s ezt már használtuk korábban): X. Itt kell felhívnom a figyelmet a szempontváltásra. Eddig ugyanis az X halmaz jobbára (de volt már kivétel) egy objektivista elemzés keretei között maguknak a valóságos objektumoknak a halmazát jelölte. Most váltunk, és mivel nem óhajtok a valóság objektumairól értekezni, és még a tapasztalati világunk bizonytalan státusú objektumaira sem akarok hivatkozni a matematikailag szigorúnak tekinthető elemzések során, ezért az X a továbbiakban a matematikai modell része lesz, a mérendő objektumok absztrakt halmaza, amelyről csakis axiómákkal leírt tudással rendelkezünk.

Itt érdemes jelölnünk azt a matematikai struktúrát is, amelyet megalkotunk, ez az $\mathcal{X} = \langle X, R_1, R_2, \ldots, R_n \rangle$ struktúra, vagyis ez az ERR. Ez a jelölés azt akarja kifejezni, hogy az X halmaz elemeiből, valamint a rajta értelmezett $R_1, R_2, \ldots, R_n$ relációkból áll az $\mathcal{X}$ struktúra, természetesen valamilyen axiómarendszerrel meghatározottá téve a jelekben csak potenciálisan meglévő matematikai tartalmat[13]. Ez a halmaz nem úgy létezik, hogy

[13] Egy még alaposabb matematikai leírás esetén (ld. pl. Roberts 1979) e struktúrában külön jelölik az alaphalmaz elemein végezhető műveleteket. Így a struktúrát $\mathcal{X} = \langle X, R_1, R_2, \ldots, R_n, \mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_m \rangle$ sémával írhatjuk le, ahol az $\mathbf{o}_i$-k a műveletek. Egy művelet valójában egy speciális reláció, az alaphalmaz valahány eleméhez rendel hozzá

lennének valamilyen, csakis az elemeihez kapcsolódó reprezentációink, és e reprezentációk halmazáról beszélnénk (a tapasztalati világunkban létező, előbb leírt halmazok sokkal inkább ilyenek), hanem az adott axiómarendszer „által létezik": az axiómarendszer által „létrehozott" struktúráról van szó. Természetesen igyekszünk úgy elkészíteni ezt az axiómarendszert, hogy a tapasztalati világunkban meglévő összefüggések képeit hozzuk létre ezen a logikailag, matematikailag szervezett halmazon.

A klasszikus méréselmélet a konkatenálható objektumok additív tulajdonságaira vonatkozóan állít fel egy absztrakt ERR-t. *Hölder* (1901) axiómái rögzítik ezt a struktúrát, a relációk nyelvén pedig *Suppes* (1951) fogalmazta meg, hogy is kell „kinéznie" annak a matematikai struktúrának, amely logikailag, matematikailag rendezett „technológia" kialakítását teszi lehetővé a tömeg, a hossz, az idő vagy az elektromos töltés mérésére.

Ha testek tömegét mérjük, és megalkottuk az ehhez a méréshez szükséges matematikai struktúrát, akkor ez utóbbit nagyon gazdagnak képzeljük, vagyis bármilyen elem megtalálható benne, amelynek van, vagy lehet (később konstruálódik) megfelelője a tapasztalati világunkban. A matematikai struktúra megalkotása során az X halmaz elképzelésekor (pontosabban a rá vonatkozó axiómák megfogalmazásakor) lehetünk „bőkezűek", létrejöhet egy akár a tapasztalati világunkban létezőnél sokkal bővebb halmaz is. Az elméletnek olyannak kell lennie, hogy a korábban soha nem látott „dolgokkal" kapcsolatban is működjék. Ez néha furcsa következményekhez vezet. Így például kiderül majd, hogy az együttes mérés elméletében fel kell tételeznünk, hogy végtelen sok ember létezik, méghozzá nem is megszámlálhatóan végtelen sok. Ez nem probléma, mert az X halmaz definiálásában nem hús-vér emberekről van szó, hanem egy absztrakt halmazról. Akkor van csak gond, ha e bőkezű definiálás eredményeként, a mérés konkrét kivitelezése során olyankor is meg kell feleltetnünk X-beli halmazelemeknek valóságos embereket, amikor empirikusan nem lenne feltétlenül kivitelezhető a művelet. Majd látjuk (az együttes mérés részleteinek később sorra kerülő bemutatásánál), hogy fel kell tételeznünk, hogy akármilyen, valós számmal kifejezhető képességfejlettséghez találunk olyan X-beli elemet, tehát vizsgált személyt, akihez éppen ez a képességfejlettség-mérték tartozik. Ha a mérés során az lenne a feladat, hogy ilyen embert valóban találjunk is, komoly gondokkal kellene szembenéznünk. Szerencsénkre a mérés konkrét, empirikus kivitelezése során nem kell ilyen műveletet végrehajtanunk, csak a háttérelmélet megfogalmazásához szükséges. Maga a képességfejlettség mérés a modern tesztelmélet eljárásai alapján zajlik, ahhoz pedig nem szükséges, hogy ténylegesen találjunk olyan embert, aki esetében az adott képesség fejlettsége az a bizonyos megadott valós szám.

A másik itt szóba jövő matematikai struktúra alaphalmaza a *valós számok halmaza*, vagy annak valamely meghatározott részhalmaza. Megtehetjük, hogy a valós számokat a „dolgokhoz", pontosabban azok absztrakt modelljeihez, az X halmaz elemeihez rendeljük hozzá, ekkor függvénykapcsolatról beszélünk. Ha viszont a hozzárendelés a tulajdonságértékekhez történik, akkor kifejezetten egy-egyértelmű leképezésről van szó (ami persze szintén függvénykapcsolat)[14]. Az alkalmazott matematikai struktúra

szintén az alaphalmazból egyetlen elemet. Például a valós számok összeadása is egy ilyen jól ismert művelet, amely két számhoz egy harmadikat, az összegüket rendeli hozzá. Ám ez értelmezhető egy háromváltozós relációként is. A műveletek különválasztásának az ad értelmet, hogy speciális relációk, és az elmélet részletesebb, itt általában általam részletesen nem tárgyalt kifejtése során kényelmesebb a hosszabb struktúra leírás használata.

[14] Ha az X halmazon értelmezett egy ekvivalencia reláció, vagyis a matematikai struktúra (az ERR) pontosan megmondja, hogy az X halmaz elemei milyen ekvivalencia-osztályokba sorolhatók, akkor ezek az osztályok tekinthetők a mért tulajdonság értékeinek, és a számoknak ezekhez való hozzárendelése itt már izomorf leképezés, míg az X halmaz elemeihez hozzárendelve a számokat a legtöbb esetben csak homomorfizmusról beszélhetünk. A homomorfizmusok egyértelmű leképezések, amelyek esetében több halmazelemnek is

alaphalmaza lehet általánosan M, de nagyon gyakran a valós számok halmazáról van szó, azt pedig általában $\boldsymbol{Re}$ jelöli a matematikában. Az X halmaznak $\boldsymbol{Re}$-re való leképezését, vagyis a mért értékeknek az absztrakt, mért entitásokhoz való tényleges hozzárendelését, a mérés matematikai, elvi megvalósítását (eddig is használtuk ezt a jelölést) az f függvény jellemzi, vagyis $f: X \rightarrow \boldsymbol{Re}$. Érdemes tehát jól az eszünkbe vésni, hogy az f függvény írja le a számoknak az ERR elemeihez való hozzárendelését, s tudjuk, hogy ennek a valóságos cselekvések (a tényleges mérés) keretében megvalósuló megfelelője a mérés legfontosabb mozzanata.

Minden kolátozottsága ellenére az 1. ábra hozzávetőlegesen képes visszaadni valamit a mérés során szóba jövő halmazok, struktúrák, emberi konstrukciók kölcsönviszonyaiból. A lényeg az, és ezt fogom felhasználni a képességfejlettség mérésével kapcsolatos mondandóm kifejtésében, hogy viszonylag jól elkülöníthető matematikai rendszerek szükségesek ahhoz, hogy a mérés – legalábbis a reprezentációs elmélet szerint – helyes legyen. Egyrészt a konstrukciók világában mozgunk, használjuk a matematikai konstrukcióinkat, illetve a tapasztalati világunkat, és ezeket ügyesen „mozgatva” próbálunk úgy interakcióba lépni a rajtunk kívül elhelyezkedő objektív környezettel, hogy az eredmény a céljaink szempontjából adaptív legyen.

---

megfeleltethetjük ugyanazt a számot, mint például sok-sok testnek lehet 2,5 kg a tömege, vagy sok tanulónak lehet 62%-os fejlettségű a rendszerezési képessége. Az izomorfizmus esetén már egy-egyértelmű a leképezés. Ha egymástól valamilyen módon meg tudjuk különböztetni a mért értékeket, akkor már egy-egy ilyen értékhez egy és csak egy mérték, vagyis mérési eredmény rendelhető hozzá, s nincs két olyan érték, amelyhez ugyanaz.

Mérés tesztekkel

A mérés lehetősége

E könyv alapvető mondanivalója, hogy a képességek fejlettségének mérése a pedagógiában problémákkal küzd. Megmutatom azonban, hogy az e területen megfogalmazott sokféle kritika egy része méltánytalan. A szakmában kialakított mérési megoldások sok esetben nem marasztalhatók el amiatt, amit állítanak róluk, ám sokszor a szokásos kritikák egy része helyett másokat lehet megfogalmazni.

Az egész problémahalmaz mögött ott van a nagy kérdés, hogy vajon lehet-e alapvetően humán természetű, meglehetősen komplex jelenségeket számokkal jellemezni, *lehet-e mérni* ezen a területen. Minden kritikai beállítódásom ellenére ebben a kérdésben én pozitív álláspontot szeretnék kifejteni. Nem látom be, hogy a humán jelleg, a komplexitás miért lenne akadálya a kvantifikálásnak. Ha komolyan vesszük azt a tudományelméleti elvet, hogy a tudomány modelleket alkot, e modellek beválását empirikusan vizsgálja, és a vizsgálatok eredményeiből próbál következtetni arra, hogy a tapasztalati világunkban létező, tényleg komplex objektumok és jelenségek hogyan viselkednek, akkor teljességgel érthetetlen, miért kellene kizárni bizonyos diszciplínák esetén a kvantifikálás lehetőségét. Modelljeink igenis lehetnek matematikai struktúrákkal leírhatók.

Az persze lehetséges, hogy a kísérletek, ahogyan eddig próbáltuk a számokat hozzárendelni az objektumokhoz, jelenségekhez, rendszerekhez, emberekhez, stb., nem voltak kellően sikeresek, a megszületett leképezések használhatóságával szemben komoly kétségek merülhetnek fel. Az is lehet, hogy van olyan jelenségvilág, amelyben az elemekhez a számok hozzárendelése nem lesz soha adaptív.

A fizika is szembekerült ezzel a problémával. A kvantummechanika esetében a paradigma körében vizsgált rendszerek leírására használhatatlanok az egyszerű számok. Pontosabban nem lehet megalkotni úgy egy részecskerendszer kvantummechanikai leírását, hogy annak néhány fizikai mennyiség valós számokkal történő jellemzése legyen az alapja. Egy kvantumos rendszer állapotához egy komplexebb matematikai struktúra (az absztrakt Hilbert tér) elemeit kell hozzárendelni e leírás keretei között, márpedig ezek nem számok. De kapcsolatosak számokkal, a részletesebb jellemzésük is történhet (nem szükségszerűen) számokkal.

Elképzelhető, hogy a kvantummechanikához hasonlóan egyes humán diszciplínákban sem adhatunk kellően jó leírást a folyamatokról, az összefüggésekről pusztán bizonyos numerikus változók értékeinek, számoknak a megadásával. De ha ez így is van, akkor sincs szó a matematika száműzéséről, sőt, a matematika összetettebb, szofisztikáltabb struktúrái játszhatnak majd szerepet a jobb leírások megszületésekor.

Ám ezek csak spekulációk. Maradjunk a mai eszközeinkkel vizsgálható kérdéseknél! Ha a feladat a képességek fejlettségének mérésével kapcsolatos problémák tárgyalása, akkor azt érdemes maguknak az itt szóba kerülő legfőbb fogalmaknak a meghatározásával kezdeni.

Mi a képesség? Az első definíciós problémák

A tesztek alkalmazására épülő pszichológiai és pedagógiai mérési tevékenység esetén feladatokat adunk a vizsgált személyeknek, s e feladatokat ők megpróbálják megoldani. Ebben, mindegyik feladat esetén külön-külön, vagy sikeresek lesznek, vagy nem. Egy elméleti modell felállítása kezdődhet annak az állításnak a megfogalmazásával, hogy egy képességet jellemezhetünk azokkal a *feladatokkal*, amelyek a szóban forgó képesség „működése" által oldhatók meg.

Már ezen a ponton komoly definíciós problémák merülnek fel. Mit jelent vajon az, hogy egy feladat egy adott képesség „működése" által oldható meg? Tudunk vajon ilyen megfeleltetéseket létrehozni? Tudjuk, hogy egyes feladataink megoldásához milyen képességre, vagy képességekre van szükség? Egyáltalán, mik azok a *képességek*, amikor egy ilyet megnevezünk, akkor *miről beszélünk*?

Legyen szó mondjuk a *problémamegoldás* képességéről! A szakirodalomban is nagyon sokszor szerepel, számtalan elméleti elemzés és empirikus vizsgálat kapcsolódik hozzá (csak mutatóba: Bogard, Liu és Chiang 2013; Greiff és mts. 2013; Zhong, Wang és Chiew 2010; Langley és Rogers 2005; Molnár 2003; Mayer 1998; Taatgen 1997; Anderson 1993; Glaser 1983; Chi, Feltovich és Glaser 1981; Anderson 1976; Newell, Shaw és Simon 1958). Ha megkérdeznénk embereket, hogy milyen feladatokat tekintenek problémának, akkor valószínűleg nagyon különböző válaszokat kapnánk, de nem nehéz eljutni egy olyan körülíráshoz, amellyel nagyon sokan egyetértenek, talán a témával foglalkozó szakemberek döntő többsége is. Ez úgy hangzik, hogy problémának tekintjük azt a feladatot, amelynek megoldásához a vizsgált személy nem rendelkezik explicit, előhívható *algoritmussal*, azt *meg kell konstruálnia*. Természetesen ez sem egy egzakt tudományos definíció, hiszen a pszichológia képességekkel foglalkozó részterülete nem axiomatizált, nincs formális rendszere, így formális definíciók sem alkothatók. De gyakorlati célokra „elmegy" a mondott körülírás: ha van előttem egy feladat, akkor magam is el tudom dönteni, vagy az engem kikérdező pszichológus is dűlőre juthat abban a kérdésben, hogy a feladat számomra probléma vagy sem. És biztosan nagyon sok képesség esetén lenne hasonló a helyzet: nem lenne nehéz eldönteni, hogy egy feladat megoldása során szükség van-e szövegértésre, megfigyelésre, tanulásra, számolásra, matematikai szövegesfeladat-megoldásra, induktív gondolkodásra. Vagyis nem ördögtől való feltételezés, hogy egy képesség reprezentálható azoknak a feladatoknak a halmazával, amely feladatok megoldásához szükség van arra a képességre[15]. Ezt a meggondolást pusztán abban az egyszerűsítési műveletben szeretném kamatoztatni, miszerint a matematikai modell kidolgozása során csak arra építsünk, hogy dolgunk van *feladatok egy valamilyen halmazával*. Hogy ezek a feladatok éppen konkrétan melyek, s hogy miképpen választottuk ki őket, maradjon homályban, ne legyen a kiépítendő modell része, csak azt feltételezzük, hogy vizsgálataink kezdetekor már rendelkezésünkre állnak. Arra van csak szükségünk, hogy valamilyen ésszerű, a reprezentációs méréselméletnek megfelelő, lehetőleg kvantitatív módon jellemezzük az embereknek az adott feladatok körében nyújtott teljesítményét.

Néhány nyugtalanító megfontolás

Matematikai tudásom íratta le velem az előző részben a szavakat, az igény, hogy elhárítsak minden akadályt a „tisztán" matematikai vizsgálatok elől. „Pedagógiai lelkem" azonban figyelmeztet: ez az egész mégsem olyan egyszerű. Vegyünk egy példát, legyen ez a véges

[15] A problémamegoldás esete egyébként még az átlagosnál is „kacifántosabb". Ha a probléma olyan feladat, amelynek megoldásához a vizsgált személy nem rendelkezik algoritmussal, akkor a probléma jellegű feladatok halmazának ez a körülírása csakis személyenként lesz egyértelmű. Nekem lehet probléma olyasmi, ami nagyon sok más embernek nem az. A tesztelméletek alkalmazása során igencsak nagy jelentősége van annak, hogy egyszerre sok vizsgált személlyel végezzünk vizsgálatot, s velük ugyanazokat a feladatokat oldassuk meg. Így azonban könnyen előfordulhat, hogy ugyanabban a vizsgálati mintában lesznek olyanok, akik számára egyes kiválasztott feladatok nem jelentenek problémát, és lesznek olyanok, akik számára igen. Ha egy képességnek a definíciójaként használjuk a hozzá tartozó feladatok megadását, akkor a problémamegoldás esetén komoly problémák merülnek fel.

tizedes tört alakú számok írásbeli szorzásának elvégzése! Ezt egy képességnek tekintjük, amelyhez nagyszerűen tudunk feladatokat hozzárendelni, ezt lényegében vita nélkül megtehetjük. Ilyen feladatból végtelen sok van, azonban nyilván minden ember legalább egy hibát vét azokban az esetekben, amelyekben két nagyon-nagyon hosszú szám írásbeli összeszorzásáról van szó. Ahogy egyre hosszabb számokat kell összeszorozni, úgy egyre nő a hiba elkövetésének valószínűsége, nullává válik a jó megoldás valószínűsége (szinte lehetetlen nem hibázni). Éppen ezért van itt egy nyugtalanító kérdés: ha az írásbeli szorzás feladatait valaki valóban csak „elszámolások" miatt rontja el (rosszul ad össze egy ponton számokat, téveszt két egyjegyű szám összeszorzása során, eltéveszti az átviendő számot), akkor vajon ilyen feladatok megoldatásával tényleg azt vizsgáljuk, hogyan tud ez a valaki írásban szorozni? Hiszen magát az algoritmust jól ismeri, az algoritmus alkalmazásában egyáltalán nem vét, csakis figyelmetlenségből, fáradtságból, és véletlen hatásokból adódó tévesztései vannak. Ez azonban nem kell, hogy megzavarja a matematikai modell kiépítését, legföljebb a matematikusok arra figyelmeztethetik a pszichológusokat, hogy nézzenek utána, feladataik helyes vagy helytelen megoldása, és így a teljesítmény vajon egyértelműen az adott képességgel függ-e össze. A matematikus modellt alkot, az már nem az ő dolga, hogy megvizsgálja, konkrét esetben alkalmazható-e a modell.

Rendkívül nehéz azonban „tiszta képességeket" mondani. „Tiszta képesség" alatt azt értem, hogy egy képesség a hozzá rendelt feladatok megoldásában egyedül játszik szerepet, semmi más nem befolyásolja a megoldás minőségét. Természetesen a helyzet tisztázása egy konkrét esetben csakis definíció kérdése. Egy képességet úgy határozhatok meg, ha már kijelöltem azokat a feladatokat, amelyeket hozzá tartozónak vélek, hogy a képességbe belefoglalom mindazokat a műveleteket, amelyek a feladatok megoldásához szükségesek. Az írásbeli szorzás esetében például, nem pusztán az írásbeli szorzás algoritmusának ismerete, alkalmazni tudása értendő bele a képességbe, hanem a szorzótábla ismerete, az egyjegyű számok és más egész számok összeadása is.

Ha egyébként valaki lényegében tökéletesen ismeri a szorzótáblát, akkor is véteni fog a rendkívül nagy számok összeszorzása során, már csak azért is, mert egy idő után nem képes kellően koncentrálni, jelentős mértékben elfárad, illetve véletlen hatások befolyásolhatják. Vagyis ilyen esetben a tévesztés sem az algoritmusalkalmazásra, sem a számolással kapcsolatos képességekre nem jellemző, hanem azzal kapcsolatos, hogy bármilyen feladatunk végrehajtása során megnövekszik a hibázás lehetősége, ha fáradtak vagyunk, vagy a túl hosszú idő (ameddig a nagyméretű feladat végrehajtása tart) valószínűsíti, hogy valamilyen véletlen zavaró körülmény szerephez jut. Említsük meg ezzel összefüggésben a pszichológiai, pedagógiai mérésekkel foglalkozók egyik szörnyű rémét, azt a helyzetet, amikor egyes feladatok megoldása nem független más feladatok megoldásától (ez alapvetően kétségessé teszi a mérés elvégezhetőségét). Márpedig itt könnyen előfordulhat ez a helyzet. Bekerül egy vagy több, sok jegyet tartalmazó szám összeszorzása is egy feladatsorba, a teszt feladatainak megoldása rendkívül fárasztó, a végefelé előkerülő feladatok megoldásának minőségét már igencsak befolyásolja a fáradás, vagyis valójában az előző feladatok megoldása.

Ezek a megfontolások még élesebbé válnak, ha arra gondolunk, hogy pedagógiai szempontból az ilyen jellegű méréseknek mi a jelentőségük. Oktatási helyzetben feladatunk lehet annak megállapítása, hogy mi az a képesség (a feladatok által leírt „nagy" képességen belül jelentkezők között), amely esetleg nem elég fejlett, és a fejlesztésére lenne szükség. Egyáltalán nem mindegy, hogy az írásbeli szorzás algoritmusának jobb elsajátítására, a biztosabb kivitelezésre kell-e tanítani a tanulót, vagy a szorzótáblával, netán az egész számok összeadásával van baj, esetleg szó sincs ilyesmiről, hanem csak a fáradás okozott problémát a tesztfeladatok megoldása során. Ha az írásbeli szorzás feladatainak jó vagy rossz megoldása az egyetlen információm, akkor e kérdésre nem tudok válaszolni. Erre még lehet azt mondani, hogy külön meg kell nézni, hogy ismeri-e a tanuló a szorzótáblát, vagy tud-e egész számokhoz egyjegyű egész számokat hozzáadni. Ha e képességek esetén baj van, akkor nem érdemes egyelőre az írásbeli szorzás algoritmusával vesződni, elemibb műveletek biztos kivitelezésének megtanítására van szükség. Ha azok megfelelően kiépülnek, akkor már vizsgálhatjuk az írásbeli szorzást. Akkor sem túl nagy számokkal, hogy a fáradás és más hatások ne érvényesülhessenek, és az ott tapasztalható eredmények már megbízhatóan jelezzék az algoritmus elsajátításának színvonalát.

Ez az eszmefuttatás megfelelő lehet egy pedagógiai nézőpontból, azonban további megfontolások szükségesek, ha a képesség és a képességfejlettség általános definíciója a kérdés. Látni kell egyrészt, hogy vannak olyan képességek, amelyek egyáltalán nem ilyen viszonylag egyszerű módon írhatók le, mint az írásbeli szorzás. Mi van például a problémamegoldással? Ha feltételezzük is, hogy van valamilyen mechanizmus, amely magát a problémamegoldást jelenti (hogy ez reális-e, azt én erősen kétlem, de az elemzés érdekében most fel kell tételeznem), hasonlóan ahhoz, ahogy az írásbeli szorzás lényegét egy algoritmus adja, akkor is nyilvánvaló, hogy a problémamegoldást igénylő feladatokban megszámlálhatatlan más képesség, illetve ismeret játszik szerepet. Sőt, mivel problémamegoldás igénye felmerülhet az általunk birtokolt ismeretek bármelyikével kapcsolatban, ezért a problémamegoldások színvonalát valójában a világról alkotott teljes ismeretrendszer befolyásolja. Az is elég nyilvánvaló, hogy a problémamegoldások során szinte minden más képességünk is „munkába kerül" (hol a képességeknek egyik csoportja, hol egy másik). Mégsem tehetem azt, hogy előbb minden más képesség, ismeret fejlettségét megvizsgálom, és annak tudatában próbálok mondani valamit a problémamegoldás képességéről!

Érdemes még egy példát elemezni. Nézzük a szövegértést! E képesség vizsgálható olyan feladatokkal, amelyek konkrét szövegekhez kötődnek. Egy adott szöveg több feladat „hordozója" is lehet (sajnos itt is felmerül a feladatok nem független voltának problémája). Mindenesetre nem az a legfőbb gondunk, hogy a szövegértés képességének közreműködésével megoldható feladatokat azonosítsuk. A probléma itt is az, hogy a szövegértési feladatok megoldása szoros kapcsolatban van számtalan más képességgel, és ismét a világról alkotott teljes ismeretrendszerünkkel. Ha feltételezzük, hogy a szövegértésnek van az agyunkban egy elkülönülő mechanizmusa, akkor annak az ide tartozó feladatok segítségével való vizsgálata éppen ezért igencsak nehéz feladat lesz. A problémamegoldás és a szövegértés tehát e szempontból rendkívül hasonlók, és melléjük sorolhatjuk bátran a tanulás, a megfigyelés, a döntés, a kommunikáció képességeit (és nyilván még nagyon sokat). Mindezek esetén tehát végtelen sok feladat van, és azok megoldásában az eredményességet nem csak valamilyen, az adott képességhez tartozó központi mechanizmus határozza meg (ha van!), hanem nagyon sok további képesség is, illetve a világ egészéről alkotott tudás (ismeretrendszer) is. Ez azért probléma, mert amikor a képességfejlettség definícióját igyekszünk megalkotni, problematikus kiindulni abból, hogy az adott képességhez nagyon jól hozzárendelhetők a feladatok. Ha persze azt mondjuk, hogy egy képességhez mindazok a feladatok hozzárendelődnek, amelyek megoldásában annak a képességnek szerepe van, a definíció eléggé egyértelművé válik. De akkor számolnunk kell azzal a nehézséggel, hogy a sikeres és a sikertelen feladatmegoldások mögött egymástól rendkívül eltérő okok állhatnak, a megoldás – nem megoldás mintázatok első megközelítésben nem teszik lehetővé, hogy magának a „tisztán vett" képességnek a fejlettségét vizsgáljuk.

Vannak-e „tisztán megnyilvámuló” képességek?

Természetesen az is kérdés, hogy azok a bizonyos „tisztán vett” képességek léteznek-e. E kérdés akárcsak modellszerű megválaszolása is egy újabb könyvet igényelne, saját megoldásomat, válaszomat itt nem közölhetem teljes részletességgel. Valamit azonban mondanom kell, mert különben nem tudunk továbbhaladni.

A *képességek természete, mibenléte* régi probléma a pszichológiában és a pedagógiában. E probléma súlyának érzékeltetésére induljunk ki abból, hogy feladataink megoldásának belső feltételei vannak, tudások (a szót most tág értelemben használva), amelyek működnek egy-egy feladat megoldása során. A tudományt is – nem utolsó sorban a gyakorlat ösztönzésére – érdekli, hogy mégis mik ezek a belső feltételek, elkülöníthetők-e működések, amelyek a feladatok egyes csoportjaiban mindig ugyanúgy részei a feladatmegoldásnak. Az agyba természetesen nem látunk bele. Pontosabban lényegében most, a 21. század elején kezdődik csak e terület fejlődésének az a szakasza, amelyben már ez az „agyba betekintés” lehetővé válik, ám ez a kutatás egyelőre nem szolgáltat átütő eredményeket a képességeink értelmezésével kapcsolatban. Minden ilyen szituációban a modellalkotás a feladat. Modellt alkotunk arról a működésről, amely a feladataink megoldása mögött van.

Hagyományosan az agyunkba, a feladataink megoldásának feltételeiként képességeket „látunk bele” (modellezünk). Ezeket egyfajta *programoknak* tekintjük (Csapó 2003), amelyek a hozzájuk rendelhető adatokon működve produkálnak eredményeket, amelyeket aztán az agy felhasznál. A „fejben szorzás képességünk” például – ebben a konstrukcióban – egy olyan működés, amely egyetlen feladattípusra áll rendelkezésre. A legtöbb felnőttben és az idősebb gyerekekben működik az egyjegyű egész számokkal, ezt a szorzótábla ismeretének is mondhatjuk. Néhányan képesek kis erőfeszítéssel egyjegyűt és kétjegyűt összeszorozni fejben, már jóval kevesebben birkóznak meg két kétjegyű összeszorzásával, az ennél „hosszabb” számok esetén a feladatot általában már csak fejszámolóművészek vagy egyes autisták képesek megoldani. Természetes feltételezés, hogy van valahol az agyunkban egy idegsejtekből álló hálózat, amely ezt a feladatot elvégzi. Hogy ez tényleg így van-e, azt nem tudjuk. Kétségeket ébreszthet, hogy az egyjegyűek összeszorzását valószínűleg csak a memóriánk felhasználásával végezzük el, míg a továbbiak ugyan igényelnek memóriamunkát, de nem mondhatjuk, hogy csak azt. Az összetettebb fejben szorzási feladatok elvégzéséhez ismerni kell a szorzótáblát, tudni kell összeadni számokat, kezelni kell a helyiértéket. Mindez még nem cáfolja, hogy a munkát valóban egy jól körülhatárolható idegsejthálózat végzi, de fel kell tételeznünk, hogy aktuálisan ez a hálózat olyan elemekből tevődik össze, amelyek más feladatok esetében, más idegsejthálózatok részeként is szerephez jutnak. És az is fontosnak tűnik, hogy amikor nem két egyjegyű számot szorzunk össze fejben, akkor szerepet játszik egy algoritmus is, amely lényegében a többjegyűek összeszorzásának algoritmusa. Ez nagy valószínűséggel egy az agyban, a hosszútávú memóriában (ha van ilyen) tárolt ismeret. A 8 × 23 feladatot például valószínűleg a legtöbb ember így oldja meg:

$$8 \times 23 = 8 \times (20 + 3) = 8 \times 20 + 8 \times 3 = 160 + 24 = 184.$$

Még az is lehet, hogy a 8 × 20, valamint a 160 + 24 feladatok is felbontandók egyszerűbb összetevőkre. A 8 × 3 eredménye valahol tárolva van az agyban, azt elő kell keresni és „munkába kell venni”. A 23-nak 20 + 3 -ra való felbontása, hogy tudniillik ezt így érdemes csinálni, ismét egy ismeret, valamikor megtanultuk. Ugyan a legtöbb ember nem lenne képes világosan elmagyarázni, hogy miért jó, de automatikusan ezt teszi. Ugyanez érvényes a disztributív törvény alkalmazására is (a második lépés). Fogalmunk sincs, hogy milyen módon van tárolva az agyunkban az a szabály, hogy „ha két szám összegét 'látod', ami szorozva van egy harmadik számmal, akkor ezzel a harmadikkal külön megszorozhatod a két számot, és összeadhatod az eredményeket". Ezt így, vagy valami hasonló módon nagy valószínűséggel kevesen tudják elmondani, de a műveletet jól tudja végrehajtani az emberek nagy része. Ott van az ismeret, implicit jellegű, más szóval tacit tudás, elővehető, használható, még ha ez nem is tudatos folyamat.

Vagyis azt látjuk, hogy a fejben való szorzás explicit vagy implicit (tacit) ismeretek felhasználásával történik. Na jó, de mi „csinálja" mindezt? Van egy elkülönült idegsejthálózat az agyunkban, amely e feladatot végzi? Úgy vélem nem. A feladatot felbontottuk összetevőkre, mindegyikhez a „forrás" a memóriában található, a feladat ezeknek a „forrásoknak" a „mozgósítása", a megfelelő rendben, egymás után kell elvégezni a bennük tárolt műveleteket. Rendkívül erős pozíciói vannak a mai pszichológiában és agykutatásban annak az elképzelésnek, hogy minden ilyen feladatot egy központi „processzor", egy feldolgozó végez, bizonyos értelemben a *gondolkodásunk*. Gondolkodva oldjuk meg a feladatot, agyunknak az a része, amely folyamatosan foglalkozik a figyelmünk középpontjában álló feladatokkal, ezt is megoldja. Minden olyan feladatot, amelyhez gondolkodási erőfeszítésre van szükség, a gondolkodásunk hajt végre, a központi „processzor", a központi feldolgozó. Ezt az elképzelést nagymértékben alátámasztják azok a vizsgálatok, amelyek azt mutatják, hogy a kicsit is komplexebb feladatok esetén az agy igen nagy területei bekapcsolódnak a megoldásba, illetve azok, amelyek a homloklebeny szerepét emelik ki (ld. pl.: Edin és mts. 2009; Anderson és mts. 2008.; Corbetta, Patel és Shulman 2008; Chadderdon és Sporns 2006; Fuster 2002).

De nem csak olyan feladataink vannak, amelyeket a központi feldolgozó hajt végre a memória hathatós részvételével. Az előbb azt próbáltam érzékeltetni, hogy kellően adaptívnak tűnő elképzelés, hogy nincs olyan agyi központunk, elkülönült idegsejthálózatunk, amely kicsit is összetettebb számolási feladatokra specializálódott volna. Lényegében ugyanez lenne elmondható ennél sokkal összetettebb feladatainkra is, mint amilyen a problémamegoldás, a tanulás, a kommunikáció (ahogyan az ténylegesen zajlik, mondjuk egy beszélgetésben, vagy levélváltásban), a szövegértés, a matematikai szöveges feladatok megoldása. Mindezek gondolkodást igénylő feladatok, reális feltételezés, hogy bennük rendkívül fontos szerephez jutnak az ismeretek (nem csak a tények, adatok, hanem az algoritmusokkal kapcsolatosak is), ezeknek az ismereteknek a felhasználása, műveletekbe való befoglalása, a műveletek elvégzése. Nem ilyen viszont például a nyelvi feldolgozás sok elemi összetevője. Még inkább egyértelmű példa a térlátás. De említhetnénk az arcfelismerést, a fizikai hatások értelmezését (a testek mozgásával kapcsolatos információk feldolgozását). Az agykutatások és különösen a kognitív- és evolúciós pszichológiai megfontolások eredményei azt valószínűsítik, hogy e feladatok esetében viszont *vannak* elkülönült feldolgozó központok, vagy az ezen a területen leginkább elterjedt megnevezéssel élve *modulok* (Tooby és Cosmides 1992; Fodor 1983). Bár a viták ma is tartanak (Barrett és Kurzban 2006; Prinz 2006; Buller és Hardcastle 2000; Hirschfeld és Gelman 1994; Karmiloff-Smith 1992), de nagyon erős elképzelés az, hogy bizonyos speciális feladatokra külön információfeldolgozó „egységek" alakultak ki a törzsfejlődés során. A kutatási eredmények már régóta megerősítik, hogy a nyelvfeldolgozás részterületeinek vannak központjai (de nincs kommunikációs központunk). Továbbá: a térlátással is hasonló a helyzet, sőt, bizonyos értelemben ez az egyik legjobb példa. Senki nem gondolkodik azért, hogy térben lássa a környezetét. Teljes mértékben automatikus folyamat, az agy szinte zseniálisnak mondható produkciója. A retinára érkező, két dimenzióban felfogott fényingerek feldolgozása zajlik a gondolkodás teljes kizárásával, úgy hogy egy háromdimenziós, „jelentésteli" kép alakul ki az agyban. Van látóközpontunk, elég jól meg tudjuk mondani, hol található, nagyszerű tudományos leírások születtek erről a képességünkről (Marr 1982).

Ahogy már írtam: az ilyen működéseket, pontosabban az azokat végző központokat *Jerry Fodor* nyomán *moduloknak* nevezzük. A modulok teljes mértékben tartalomspecifikus működést végeznek (a látási rendszer például csak látással kapcsolatos információkat dolgoz fel), tudatosan nem befolyásolhatók, kötött működésűek, tehát nem flexibilisek, *Fodor* szerint nem tanulnak, velünk születettek (Fodor 1983). A térlátást szolgáltató modul nagyon sok feladatunk végrehajtásában szerepet kap, és így van ez számos más modullal is. A nyelvhasználattal kapcsolatos moduljaink (valószínűleg több is van) részt vesznek minden kommunikációs feladatunk végrehajtásában, de részesei lehetnek a matematikai szövegesfeladat megoldásoknak is, természetes módon a tanulás legtöbb feladatában is jelen vannak, és így tovább.

Vagyis a leírtak alapján megalkotható az a modell, amelyben a feladatainkat valójában a központi feldolgozó hajtja végre, e munkája során jelentős mértékben támaszkodik a memóriára, vagyis a tárolt ismeretekre, valamint a specifikus információfeldolgozást végző modulokra. Ez a modell jelentős mértékben memória alapú kognitív működést feltételez.

Természetesen szó sincs arról, hogy ez az én találmányom lenne. A kognitív tudományban erős paradigmának számít az információfeldolgozás memória-alapú modellje. E kutatások közül sok a mesterséges intelligencia kutatásához köthető, és elsősorban olyan gépi megvalósítást jelentenek embert is jellemző információfeldolgozási folyamatokkal kapcsolatban, amelyek alátámasztják a memória alapú modellek létjogosultságát (Stanfill és Waltz 1986). Szintén mesterséges intelligencia kutatási megközelítés az „eseteken alapuló gondolkodás" koncepciója (ld. pl. Aamodt és Plaza 1994), amely a problémamegoldást képzeli el a korábbi, hasonló problémamegoldások során gyűjtött tudás „újrahasznosításával". A memória-alapú gondolkodás (Memory-Based Reasoning) különböző jelenségek magyarázatában kapott szerepet, mint amilyen az angol kiejtés (Stanfill 1987). *Joaquín Fuster* (1997) vázolta fel a memória olyan koncepcióját, amelyben mind az észlelés, mind a motoros funkciók valójában a memóra struktúráiba, hálózataiba integrálódnak. *Bradley Postle* (2006) a munkamemória szerepével és működésével kapcsolatban alakított ki olyan modellt, amely szintén a memória alapvető funkcióit emeli ki. *Paul Baxter* és *Will Browne* (2010) az autonóm működésű robotokra vonatkozó elmélet megalapozásaként elemzik az információfeldolgozás memória-alapú modelljét. A hitek, elképzelések nem logikus következtetések kialakulásában játszott szerepére, tehát ismét a memória jelentőségére mutatott rá több kutatás (ld. pl. Steegen és DeNeys 2012). Magyarul is olvasható *Mike Anderson* könyve (1992), amelyben a neves szakember egy olyan kognitív működés architektúrát vázol fel, amelyhez nagymértékben hasonlít az általam leírt modell.

Mindez azért fontos a számunkra, mert érdemes feltenni a kérdést, hogy mi is egy képesség ebben a modellben. Bizonyos képességekkel kapcsolatban – alkalmazva a modellt – elkülöníthetünk agyi idegsejthálózatokat, amelyek azonosíthatók az adott képességhez rendelt feladatok végrehajtó apparátusaként. Az ilyen feladatoknak a lényege általában az ingerek közvetlen feldolgozása. Vizsgálhatók e működések, ha van egyértelmű, csak e működéstől függő válasz azokra a helyzetekre, amelyek a feladatokban szerepelhetnek. A fizikai változásokkal kapcsolatos információfeldolgozási folyamatok produkálhatnak például ilyeneket. Azok a modulok, amelyek ebben részt vesznek, azonnali, „reflexszerű" válaszokat eredményezhetnek, s ezek jól vizsgálhatók.

Az esetek döntő többségében azonban nem beszélhetünk valamifajta modulok „tiszta” megnyilvánulásáról, a feladat megoldása során minden felsorolt tényező (központi feldolgozó, memória, több modul) szerepet kap. Különösen a pedagógia számára fontos *összetett képességek*, az azokhoz tartozó feladatok ilyenek. Ha ilyen feladatok megoldása során nem vagyunk sikeresek, vajon azt mire kell visszavezetnünk? Túl sok a szereplő.

A képességek, mint társadalmi konstrukciók

Vajon miért sorolunk együvé egymástól sokszor igencsak különböző feladatokat, mondván, hogy ezek mind problémamegoldások, vagy tanulási feladatok, vagy kommunikációs feladatok? Mint láttuk, ebben a modellben nem azért, mert ugyanannak az agyi rendszernek a működését igénylik. Az ok egészen máshol, nem az egyes agyakban keresendő. *A kultúrában*. A képességeket feladathelyzetek jellemzik, és az emberi társadalmak hosszú fejlődésük során megkonstruáltak egy csoportosítást. Vannak a feladatoknak jellegzetességeik, amelyek egymáshoz hasonlókká teszik őket. Ez nem osztályozás, például egy feladat lehet egyszerre tanulási és problémamegoldási jellegű, azonban viszonylag egyöntetűen, és nagy hatékonysággal meg tudjuk mondani, hogy egy feladatot egy kategóriába besorolunk, vagy sem. Kulturális, társadalmi *konstrukcióról* van szó, amely működik anélkül is, hogy bármit tudnánk az agyról. Szó sincs itt kezdetben agyi működésekről, azokat a gondolkodó ember már a pszichológia magyarázat iránti igényeinek felmerülése után igyekezett társítani ezekhez a feladatcsoportokhoz.

Ez a konstrukció azonban – ahogy már többször jeleztem – azzal a következménnyel járt, hogy egy-egy feladatunk több kategóriába is besorolható. És ennek alapvetően az az oka, hogy a feladatok olyan részfeladatokra bonthatók, amelyek valamilyen módon együtt nyilvánulnak meg, egyszerre (egy időben), vagy/és szekvenciálisan hozzájárulva a megoldáshoz. Egy-egy képesség tehát továbbiakra bontható, amíg el nem jutunk valamifajta „elemi képességekhez” – gondolhatnánk. De ez az állítás is rendkívül bizonytalan. Például az írásbeli szorzás képességének részeként az összeadás (egy- vagy többjegyű pozitív egész számhoz egy egyjegyű pozitív szám hozzáadása) vajon „elemi képesség”, nem tudjuk már tovább bontani részműveletekre? Amikor a 37-hez 8-at adunk, mi történik? Valószínűleg minden az emberek által ténylegesen használt algoritmusnak része, hogy valójában előbb a 7-et és a 8-at adjuk össze, a többjegyű szám tizeseit, itt a 30-at „leválasztjuk”, majd később adjuk hozzá az eredményhez, mert az könnyebb feladat. A 7-hez a 8-at pedig úgy adjuk hozzá (nem elemezve most külön, hogy van, aki azonnal megfordítja a feladatot, és a 8-hoz adja hozzá a 7-et), hogy a „tízes átlépés” módszerét alkalmazzuk. Tudjuk, hogy a 7-hez hármat kell adni, hogy tízet kapjunk (ez a memóriánkban van valószínűleg), majd a 8-at bontjuk fel 3 + 5-re (ez is a memória segítségével végezhető művelet), és így már a 7 + 8-at 10 + 5 -ként számolhatjuk. De az is lehet, hogy valaki a két egyjegyű pozitív szám összegének számítását teljes mértékben, és azonnal a memóriája segítségével hajtja végre, mindegyik összeadás ott van tárolva, csak elő kell hívni az eredményt. A 30-at és a 15-öt már könnyű összeadni (de akár még ezt a feladatot is további elemekre bonthatnánk!). Mindez természetesen spekuláció, hiszen semmilyen információnk nincs arról, hogy a 37 és a 8 összeadásakor mi történik az agyunkban. Számokkal sokat foglalkozó emberek agyában követhetetlen gyorsasággal születik meg az eredmény, és inkább egy holisztikus, sőt, szinte „képies” megoldás lehet jellemző, a tapasztalt emberek szinte „látják” a végeredményt.

Láthatjuk tehát, hogy még egy első ránézésre eleminek gondolt számítási feladat is további részekre bontható (ráadásul többféleképpen is), és az is „paradigmatikusnak” tűnik, hogy végül is elemi memóriaműveletekig jutunk el. Természetesen az itt leírt műveletek összeszervezése, algoritmussá alakítása is feladat, az elemi memóriaműveletek sikere még egyáltalán nem garantálja, hogy az egész feladatot meg tudjuk oldani. De vajon az algoritmus hol van? Két eset lehetséges véleményem szerint: (1) vagy van egy olyan specifikus idegsejtcsoport, amely hálózati működése során fogadja az adatokat (a 37-et és a 8-at a példában), és mint egy számítógép programja végrehajtja a feladatot, vagy (2) a feladatot a központi feldolgozó oldja meg, mint minden feladatunkat. Az (1) verzió a modulszerű működésre lenne jellemző, és elfogadásával azt feltételeznénk, hogy az emberiség evolúciós genetikai fejlődése során tett szert erre a modulra. Ennek alapvetően ellentmond, hogy az ilyen algoritmusokat kínnal-keservvel tanuljuk, szó sincs, mondjuk, a térlátás könnyedségéről, arról, hogy már az újszülöttnek is rendelkezésére áll. Az is ellentmond annak, hogy egyjegyűt kétjegyűvel összeadni az evolúció során tanult meg az agyunk (vagy inkább a génjeink?), hogy a 100 000 évvel ezelőtt élt ősembernek még nem kellett ilyen feladatokat megoldania, de a mi génállományunk lényegében megegyezik az övével. Vagyis a (2) sokkal inkább hihető.

A képességek „munkadefiníciója"

Csak kerülgetem a forró kását. A képességnek a mérések során használhatónak tűnő meghatározásába bele-, belekapok, ám újabb és újabb gondolati nehézségekkel találom magam szemben. Végül is definiálhatjuk-e a képességeket úgy, hogy azok az emberek által végezhető, megoldható feladatok bizonyos, társadalmilag megkonstruált részhalmazai? Miért ne tehetnénk? A definíciók alkotásában az embert csak a logika, valamint egy a definíciónak megágyazó paradigma (igencsak jó esetben egy formalizált axiómarendszer) előírásai korlátozzák. Az írásbeli szorzás feladatai specifikálják az írásbeli szorzás képességét, és szemet hunyunk azon probléma fölött, hogy így nem pusztán magának a szorzásnak az algoritmusára koncentrálunk, sőt, nagyon sok teljesítményt határozhat meg jelentős mértékben negatívan más jellegű tudáselemek színvonala (összeadás, szorzótábla). A tanulás képessége mindazokkal az emberi feladatokkal határozható meg, amelyekben nyilvánvalóan tanulás történik. Abból indulunk ki, hogy nem vizsgáljuk, mi okozhatja a tanulási feladatok megoldása során a sikertelenséget. Vagy sikerül, vagy nem sikerül a vizsgált személynek megoldania az adott tanulási feladatot, csak ezt vesszük figyelembe.

Úgy látom, a képességfejlettség mérésével foglalkozó szakemberek – valószínűleg nem tudatosan – ezt a stratégiát választották, kivonva belőle a társadalmi konstrukcióként történő értelmezést. Van egy képünk arról, hogy mely feladatokat tekintsük induktív gondolkodási feladatoknak, az induktív gondolkodás képességének vizsgálatára ezeket fogjuk használni. Van egy képünk arról, hogy melyek azok a feladatok, amelyek valamiképpen arányszámítást igényelnek, ezek szolgálnak majd alapként az arányszámítási képesség mérésekor. És így tovább. Következetes ez a magatartás, valójában hozzávetőlegesen tisztává is teszi a viszonyokat, azonban számolni kell a következményeivel. Úgy vélem, hogy a következményekkel való számolás elmarad. És ennek van egy alapvető oka: ha már sikerült tisztáznunk, mit értünk egy adott képesség alatt, hozzárendelve ahhoz a vele megoldható feladatokat, akkor kellene egy következetes definíció arra is, hogy mit tekintsünk e képesség fejlettségének. Még mielőtt az e kérdésre ajánlott válaszokat megvizsgálnám, még egy fontos problémára kell kitérnem: végtelen, vagy véges sok feladattal kell számolnunk?

Az egy képességhez tartozó feladatok száma

Világos, hogy végtelen sok írásbeli szorzást igénylő feladat van, mert a szerepeltethető tizedes tört alakú számok számossága végtelen. Ugyanakkor a feladatoknak minden ember csak egy véges részhalmazát tudja hiba nélkül megoldani, ez azt jelenti, hogy minden emberre igaz, hogy végtelen sok feladat esetén zérus, vagy nagyon kicsi szám a helyes megoldás valószínűsége, és csak a feladatok egy véges halmaza esetén beszélhetünk „valamirevaló" valószínűség-értékről. Vagyis a valószínűség-értékeknek egyetlen sűrűsödési pontja[16] van, a zérus, és nincs is más ilyen pont. Ha valaki úgy szeretné definiálni a képességfejlettséget, hogy az legyen a valószínűségértékek sűrűsödési pontja (feltéve, hogy egy ilyen van, de mint láttuk a példa esetén ez természetes, és mindjárt látjuk, hogy bármilyen valamirevaló képesség esetén az), akkor az írásbeli szorzás esetében kénytelen lenne azt mondani, hogy mindenkinek ugyanannyi, mégpedig zérus a képességfejlettsége. A sűrűsödési ponttal tehát nem sokra megyünk.

Miért gondolom, hogy minden feladatunk, minden képességünk esetén hasonló a helyzet? A tipikus, és biztosan a leggyakrabban előforduló tesztelések esetén állításom jól megállja a helyét. A tesztelések döntő többségével kapcsolatban az a képünk, hogy papíron, számítógép monitorán olvasható, vagy szóbeli utasításként hangzik el a feladat, amit aztán a vizsgált személynek meg kell oldania. Vagyis a feladatok (a tesztelések során) azonosíthatók szövegekkel. Márpedig a tesztfeladatok szövegeinek hosszúsága alapvetően meghatározza azok megoldhatóságát. Ha egy emberi élet kevés egy szöveg elolvasására, akkor az a feladat minden ember számára 0 valószínűséggel oldható csak meg. Nem csak az írásbeli szorzás küzd tehát a leírt problémával, hanem lényegében minden a tesztelések során vizsgálat alá vett képességünk. Csak véges sok feladat esetén különbözik érdemlegesen zérustól a helyes megoldás valószínűsége, végtelen sok feladat esetén ez a szám zérus. Vagyis a valószínűségértékek halmazának egyetlen sűrűsödési pontja van, a zérus.

Természetesen a gyakorlatban vizsgált képességeket meghatározó feladatok halmazának egy végtelen nagy része érdektelen a számunkra, és igazából bennünket csak egy véges részhalmaz feladatai érdekelnek. Ezekből adunk feladatokat a vizsgált személyeknek, mert úgy gondoljuk, hogy közöttük többen is lehetnek, akik majd meg tudják azokat oldani.

Bármifajta tesztelmélet kérdése az (nem állítom, hogy a tesztelméletek fejlődése során ténylegesen az is volt), hogy felépíthető-e méréselméleti értelemben releváns struktúra a feladatmegoldások halmazára, és ennek eredményeként a képességhez tartozó feladatok helyes megoldása valószínűségeinek számhalmazához tudunk-e hozzárendelni úgy számokat, hogy így valamelyik, statisztikai vizsgálatokra lehetőséget adó skála jöjjön létre.

Ha véges sok feladattal számolunk, akkor is felteheti természetesen bárki azt az akár nyugtalanítónak is mondható kérdést, hogy vajon meg tudjuk-e mondani, melyek azok a feladatok, amelyek egy képesség esetén így kijelölhetők? Mindig fel tudjuk sorolni a véges sok feladatot? Biztosan nem. Ez már az írásbeli szorzással így van, pedig ez az eset igencsak egyszerű. Ha az lenne a definíció – ahogy már fentebb is szerepelt –, hogy azt a feladatot kell felvenni egy ilyen felsorolásba, amely esetén van legalább egy ember, aki nem nulla valószínűséggel tudja azt megoldani, akkor képtelenség lenne megmondani,

[16] Egy végtelen számhalmaz egy sűrűsödési pontja egy olyan szám, amely esetén minden e számot tartalmazó nyílt intervallumban az adott számhalmaz végtelen sok eleme található. Kijelölhetjük e sűrűsödési pont akármilyen picike (de „összefüggő", vagyis intervallumot alkotó) környezetét, abban a halmaznak végtelen sok eleme lesz.

hogy milyen „hosszúságú” számokig kellene elmenni, képtelenség lenne bebizonyítani nagyon sok feladatról, hogy már mindenki nulla valószínűséggel tudná jól megoldani.

Teljességgel érdektelenek a nagyon hosszú számokat szerepeltető feladatok. A gyakorlatban, az iskolában a tanári feladatadás során majdnem biztos, hogy a tesztekben nem szerepelnek 10 számjegy hosszúságot meghaladó méretű számok. Akár azt is mondhatnánk, hogy határozottan definiálható egy képesség így: a maximum 10 számjegyet tartalmazó tizedes tört alakú számok írásbeli összeszorzásának képessége. Azt mondtuk, hogy egy képességet bárhogyan definiálhatunk (persze értelmes módon), feladatoknak akár egy tetszőleges halmaza is képességnek tekinthető, ezért a most leírt definíció minden igényt kielégíthet. Hogy van-e az ilyen jellegű definícióknak gyakorlati értékük, az más kérdés. Ennek az éppen leírtnak például határozottan van, az iskolai matematika tanulás egy részterületén, az írásbeli szorzás megtanulása során alkalmas arra, hogy segítse az eredményesség vizsgálatát.

Egy másik gyakorlati példa: alsó tagozaton a matematika tanulása során részfeladat elsajátítani a kétjegyűek, majd külön a háromjegyűek (pozitív egész számok) összeadását. Rendkívül jól behatárolható feladatokról van szó, még a számukat is meg tudjuk mondani, például pozitív kétjegyű egész számból van 90 db., így az ezek összeadását kérő feladatok száma 8100. Még a felsorolás is lehetséges, így kezdődne: 10 + 10, 10 + 11, 10 + 12, …

A képességek döntő többsége esetén nem lenne ilyen könnyű dolgunk. Mely feladatokat érdemes felsorolnunk, ha a szövegértés jellemzésére használandó feladathalmazt akarnánk kijelölni? Megtehetjük, hogy szabadon definiálunk feladat részhalmazokat, ám itt a feladatok specifikálása sokkal-sokkal nehezebb, mint az írásbeli szorzás esetében, talán nem is lehetséges. Ugyanez a gond a problémamegoldással, a tanulással, a kommunikációval, a megfigyeléssel, a matematikai szövegesfeladatokkal, az arányszámítással, az induktív gondolkodással, a deduktív következtetésekkel, stb. Mégis mondhatjuk azt, hogy a feladatok egy véges halmazával van dolgunk – amikor erre van szükségünk –, még ha nem is tudunk egy ilyen halmazt definiálni, nem is tudjuk megmondani a feladatok számát, és nem is tudjuk felsorolni őket.

A képesség fejlettsége – általános megfontolások

Nem mondok újat azzal, ha kijelentem, hogy a képességfejlettség definiálása nagyon erősen igényli a *valószínűségi* fogalomrendszer használatát. A feladatok megoldása során véletlenszerű folyamatok játszódnak le, érdemes kialakítanunk azt az elképzelést, hogy ha arra kérünk valakit, hogy oldjon meg egy adott feladatot, akkor annak, hogy a kiválasztott személy a kiválasztott feladattal sikeresen megbirkózik, lesz valamekkora valószínűsége, egy 0 és 1 közötti szám (beleértve a határokat is). Fogadjuk el tehát azt a modellt, hogy egy képességet jellemez a hozzá tartozó feladatok halmaza, és bármely ember esetén e halmaz minden egyes eleméhez hozzárendelhető egy 0 és 1 közötti szám, amely az ő esetében az adott feladat helyes megoldásának a valószínűsége.

A képességről, annak fejlettségéről nyilván az árul el sokat, hogy ezek a valószínűségértékek hol helyezkednek el a 0 és az 1 között. Ha valaki esetében ezek igen nagy, tehát 1 közelében lévő értékek, akkor annak az embernek az adott képessége fejlettnek tekinthető, mert a kérdéses feladatokat nagy valószínűséggel képes helyesen megoldani. Ha valaki esetében viszont kis értékekről van szó, (0 közeliek többségükben), akkor gyenge képességfejlettségről kell beszélnünk. Ez, idáig még – bármennyire is hivatkoztam számértékekre – kvalitatív magyarázat, érdemes tehát azon elgondolkodnunk, hogy mégis, miképpen lennénk képesek e számhalmazok (az egyes embereket jellemző valószínűségek halmazai) között olyan relációkat konstruálni, amelyek segítségével rendezési-, különbségi- vagy aránystruktúrát hozhatnánk létre e halmazok halmazán.

Nem ismerek a szakirodalomban általános és az igényeket maximálisan kielégítő megoldást e feladatra, és magam sem tudok ilyet ajánlani. Maguk a konkrét próbálkozások (klasszikus- és modern tesztelmélet, együttes mérés) mindig nagyon erős feltételezésekkel élnek a valószínűségek e halmazaival kapcsolatban, működnek, ha ezek a feltételezések érvényesülnek, de az én véleményem az, hogy a valamirevaló esetekben nem ez a helyzet.

Három komolyan vehető próbálkozást ismerünk tehát a képességfejlettség mérésére vonatkozó elmélet és gyakorlat kialakítására: a *klasszikus tesztelméletet*, a *modern tesztelméletet* és az *együttes mérés elméletét* (conjoint measurement). A három elmélet státusa azonban nem egészen azonos. A klasszikus- és a modern tesztelmélet – ha most analógiát használok – azt a szerepet játssza, mint a hosszúság mérésénél a geometria elmélete, axiómarendszere. Vagyis e két elmélet azoknak a fogalmaknak a koncipiálásához kell, amelyek a reprezentációs méréselmélet követelményeinek kielégítéséhez szükséges fogalmakként jelennek meg. Az együttes mérés elmélete viszont sokkal inkább a rendezési, a különbségi és az extenzív struktúrák elméletével hozható párhuzamba. Ugyanis egyfajta keretelmélet ez is, nagyon sok mindennek a mérése történhet az együttes mérés elméletének felhasználásával. Többek közt a képességfejlettség esetén is hasznosítható, de nem a képességfejlettség elmélete. Ha úgy tetszik, inkább a tesztelméletek fölött áll, azoknál általánosabb. Vizsgáljuk meg ezt a három elméletet, s a hozzájuk köthető gyakorlatot közelebbről!

Klasszikus tesztelmélet

A klasszikus tesztelmélet státusa

A tárgyalt elméletek, így a klasszikus tesztelmélet is feltételezi, hogy létezik egy objektív, a mérőeszköztől, a mérés körülményeitől független, valós számmal kifejezhető, legalább intervallumváltozó segítségével mérhetővé tehető *képességfejlettség szint* egy adott képesség és egy adott vizsgált személy esetén, és erre dolgoz ki becslési eljárásokat (Steyer 2001; Csapó 2000; Horváth 1991, 1993; Lord és Novick 1968; Loevinger 1957). Ezek az eljárások azon alapulnak, hogy egy vizsgált személy képességfejlettsége egy korrekt módon kidolgozott teszttel, a teszt feladataira kapott pontszámok összegével (vagy ezen összeg függvényeként meghatározható számmal, ha olyan a definíció) becsülhető.

Nem az a célom, hogy bemutassam a klasszikus tesztelmélet egészét. Itt és most pusztán azt a kérdést vizsgálom, *vajon a szokásos eljárások megfelelnek-e a reprezentációs méréselméletben lefektetett követelményeknek*. Nem tárgyalom részletesen a validitás kérdését[17], az itemanalízis és a tesztfejlesztés témáit pedig egyáltalán nem érintem. Az elemzés kritikai jellegű, azokat a problémákat helyezem előtérbe, amelyek a jól megoldott feladatok száma vagy annak valamilyen meghatározott függvénye értékének kiszámítása és képességfejlettségként történő értelmezése kapcsán vetődnek fel.

A szakirodalomban – elsősorban a modern tesztelmélettel foglalkozó munkákban – számtalan leírás létezik a kidolgozott eljárások kritikájára (pl. Borsboom 2005, 2006a,

[17] A reliabilitással, a megbízhatósággal viszont foglalkoznom kell, mert szoros kapcsolatban áll az itt tárgyalt problémákkal.

b; Alagumalai és Curtis 2005; Baker 2001; Horváth 1993; Wright és Stone 1979). Ezek mély meggondolásokon alapulnak, a kritika reális. Ha valaki csak kevéssé járatos a témában, ismerkedve a kérdéskörrel, gondolhatja úgy, hogy az alapprobléma a következő: a klasszikus tesztelméleten alapuló vizsgálatok esetén a mérési eredmények erősen függnek magától a teszttől, attól, hogy abban milyen, elsősorban milyen nehézségű feladatok találhatók. Ha egy teszt feladatai inkább nehezeknek mondhatók, mondja egy felszínes kritikus, akkor a vizsgált személyek alacsonyabb pontszámokat érhetnek el; viszont ha könnyebbek a feladatok, akkor a pontszámok magasabbak lesznek.

Ez a kritika azonban nem állja meg a helyét (önmagában). Ha a hosszúságot méterrúddal mérem, és *méterben* fejezem ki, egészen más adatokat (mértékeket) kapok ugyanazon hosszúságértékekre, mintha az angolszász országokban széles körben használt *láb* mértékegységgel kifejezve határozom meg azokat. De ez nem probléma, senkinek nem jut eszébe, hogy emiatt a hosszúságmérést rossz mérésnek tartsa. Szerepelt már korábban is: egy a reprezentációs elméletnek megfelelő mérés esetén általában sokféleképpen el lehet végezni a mért objektumokhoz a számok hozzárendelését (van kivétel, az abszolút skála, amellyel itt nem foglalkozom). Ez nem probléma akkor, ha a különböző hozzárendeléseket valamilyen szabály összeköti egymással (nem túl precíz matematikai fogalmazásban), ha például bármely két hozzárendelést kiválasztva, az azokban szereplő szám-hozzárendelések esetén az egymásnak megfelelő számok halmazai között pozitív lineáris leképezés[18] van (intervallumskála). Az már persze rendkívül fontos kérdés, hogy tényleg összeköti-e valamilyen „kívánatos” szabály a leképezéseket.

Egy példával hadd illusztráljam ezt az – elismerem – nem túl könnyen átlátható összefüggést. Van két tesztünk, és tegyük fel, hogy az ezekkel mérhető képesség fejlettségét kifejező pontszámok olyan kapcsolatban vannak egymással, hogy az első teszten szerzett pontszámot 0,8-del kell megszorozni (fontos, hogy ennek a szorzónak mindig pozitívnak kell lenni, ld. 18. lábjegyzet), az eredményhez 24-et kell hozzáadni, és akkor megkapjuk a második teszttel meghatározható értéket, méghozzá a vizsgált populáció bármely tagja esetén. Ebben az esetben nincs probléma. A két teszt ugyanazt méri, csak két különböző skáláról van szó. Formalizáltan a kapcsolat így néz ki:

$$x_2 = 0{,}8x_1 + 24,$$

ahol x_1 az első, x_2 a második teszttel becsülhető pontszám. Például, ha valakinek az első teszttel mért pontszáma 10, a második teszten $0{,}8 \cdot 10 + 24 = 32$ pontot szerez, és bárki esetében hasonló a helyzet, akkor a két teszt ugyanazt méri, intervallumskálán[19]. Ugyanarról a háttérben meghúzódó konstruktumról van szó, amelynek a fejlettségét ha különböző tesztekkel igyekszünk megbecsülni, más és más értékeket kapunk, mert a fejlettség ezekkel a tesztekkel eleve más és más mértékű. Mint egy léc hossza, ha

[18] Adatok két sorozata között lineáris leképezés van, ha van olyan α és β valós szám, hogy a két sorozatban szereplő, egymásnak megfelelő adatok között érvényes a következő összefüggés: $x_{i2} = \alpha x_{i1} + \beta$, ahol x_{i1} az egyik, x_{i2} a másik adatsor *i*-edik eleme. Pozitív lineáris leképezzésnek mondjuk a lineáris leképezést, ha $\alpha > 0$. Az intervallumskálák nagyon fontos tulajdonsága, ahogy ez már szerepelt, hogy ugyanazon mért mennyiség esetén bármely két skálán mért adatok sorozatainak egymással pozitív lineáris kapcsolatban kell állniuk.

[19] Természetesen a tesztek kidolgozása során, amikor több teszt összehasonlítása, egymáshoz igazítása a feladat, soha nem teljesülhet ez az összefüggés száz százalékos pontossággal a vizsgálatba bevont személyeknél. Minden esetben a szabályostól való eltérések mértékét kell értékelni, ennek matematikai statisztikában kidolgozott módszerei (hipotézisvizsgálatok) jól ismertek. Ennek az összefüggésnek a részletes bemutatása itt nem célom, mert eltereli a figyelmet a lényegről.

centiméteres, és ha inch beosztású skálával mérjük. A hossz értéke ugyanaz mindkét esetben, ám e hossz mérőeszközökkel megbecsülni szándékozott mértékei eltérők.

Vagyis azt kell mondanunk, hogy a klasszikus tesztelmélettel szemben megfogalmazott kritika, hogy tudniillik tesztfüggő az eredmény, önmagában nem állja meg a helyét, csak akkor probléma ez, ha a különböző mérési eredmények nem függnek össze egymással lineárisan. E feltétel érvényesülését meg kell vizsgálni.

A reprezentációs méréselméletnek való megfelelés azon áll vagy bukik, hogy a követelmények teljesítéséhez szükséges, a tulajdonsággal kapcsolatos értékekhez köthető empirikus (az ERR-ben szerephez jutó) *relációk* definiáltak-e. Tudjuk-e, hogy két ember képességfejlettségét mikor tekintsük egyformának? Tudjuk-e, hogy egy ember képességfejlettsége mikor nagyobb, mint a másiké? Össze lehet-e adni (konkatenálni) két képességfejlettséget? Vagy általánosan: az ide sorolható mérések során született-e valamikor is értelmes javaslat a képességfejlettségek halmazán definiálandó relációkra? Erre nem válasz, hogy megoldatjuk a teszt feladatait, és akkor majd meglátjuk, hogy a tesztpontszámok milyen relációkban állnak egymással. Ez az operacionalizmus színtisztán, a reprezentációs méréselméletet elfogadó tudományos közösség ettől az értelmezéstől elzárkózik (Trendler 2009; Barrett 2008; Cervone és Caldwell 2008; Borsboom 2005; Michell 1999, 2000). Éppen, hogy azt kellene „tudnia" a mérésnek, hogy a tőle teljesen függetlenül definiált értelemben egyenlő képességfejlettséggel rendelkező emberek az eljárás során ugyanazt a számot kapják fejlettségük mértékeként (csak egy példát említve a lehetséges relációk közül). Azt, hogy az egyik szakasz hosszabb-e, mint a másik, a geometria axiómarendszerével definiáljuk, és szó sincs róla, hogy valamifajta konkrét mérést végeznénk el, és mondjuk milliméterekben megadott hosszúságmértékeket hasonlítanánk össze. A relációk definícióinak matematikai elméletek részeként kell állniuk előttünk, ezek segítségével fel kell építenünk egy egész elméletet, s ha ez sikeres volt, akkor elgondolkodhatunk a számok hozzárendelésének gyakorlati, empirikus feladatán is.

Van-e a klasszikus tesztelmélet mögött ilyen elmélet? Ha lenne, annak mindenképpen azzal kellene kapcsolatban állni, hogy az egyes vizsgált személyek mekkora valószínűséggel tudják megoldani a képességhez tartozó feladatokat. Egy személy egy adott képesség tekintetében meglévő tudásának szintjét, e képessége fejlettségét valamilyen módon pontosan ezek a valószínűségek fejezik ki, és semmi más. Csakis e valószínűségértékek felhasználásával kell tudnunk definiálni a különbségi struktúra létezéséhez elengedhetetlen, és az ide illő axiómarendszerben megadott szabályoknak megfelelő *D* négyváltozós különbségi relációt (amint már korábban is ígértem, a reláció tulajdonságait, az axiómákat bemutatom a modern tesztelméletről szóló fejezetben). Ilyen reláció azonban semmilyen a klasszikus tesztelmélettel foglalkozó írásban nincs megadva ismereteim szerint.

Nem az a probléma, hogy a mérés egésze nem produkál számokat, nem az a probléma, hogy nem tudunk hozzárendelni az emberekhez, a képességükhöz, annak fejlettségéhez numerikus mértékeket. Hogy megtegyük, az csak kreativitás kérdése, s aztán már a számok közti relációk megállapítása igazán nem nagy kunszt. Az a baj, hogy azok a relációk, amelyeknek éppen a számok közti relációk felelnének meg, egyszerűen nem léteznek, nincs elméleti meghatározásuk, definíciójuk az ERR-ben. *Valójában nincs is ERR*.

Az a „nagyvonalú" kijelentés, amely minden klasszikus és modern tesztelméleti leírásnak a kezdete, hogy tudniillik létezik egy pontosan nem megmérhető, objektív, látens értéke a képességfejlettségnek, legalábbis elemzést igényel. Magam nem állok be abba a sorba, amelyben szakemberek azt állítják, hogy a klasszikus tesztelmélet mögött nem áll elmélet, és hogy valójában a látens képességfejlettség definiálatlan, meghatározatlan mennyiség. A klasszikus tesztelméletet gyakran éri az a vád, hogy operacionalizmus jellemző rá (ld. pl. Borsboom 2005), hiszen magát a mérendő mennyiséget operacionálisan, magához a mérési folyamathoz kötötten határozza meg. Azt mondja például – és ezt nemsokára részletesen is megmagyarázom –, hogy egy adott képesség fejlettsége nem más, mint a képesség mérésére megalkotott tesztben szerezhető pontszám várható értéke. És látszólag ez tényleg operacionális definíció, magával a mérési folyamattal való meghatározás, azonban ez mégsem így van. A tesztbe kerülő feladatok jó megoldásának egy vizsgált személy esetén van valamekkora valószínűsége, így valójában a méréshez, annak technológiájához, eszközeihez a képességfejlettség definíciója nem kötődik. A feladatok jó megoldásának akkor is van valószínűsége, ha nem mérünk, ha egészen más eszközöket és eljárást használunk, vagyis a definíció független magától a méréstől. Bár erre felvethető, hogy dehogy független, hiszen éppen a mérés kivitelezésétől függ, hogy milyen feladatok keülnek be a tesztbe. Ez ugyan igaz, ám éppen azt állítjuk majd, hogy a klasszikus tesztelméletnek megfelelő eljárások akkor lennének összhangban a reprezentációs méréselmélettel, ha a feladatválasztástól sem függnének az eredmények. Pontosabban akkor kerülne minden a helyére, ha a különböző tesztek eredményei egymással pozitív lineáris kapcsolatban állnának, és így már nem lehetne vádolni a klasszikus tesztelméletet és a rá épülő mérési eljárásokat operacionalizmussal. A probléma tehát nem az, hogy a klasszikus tesztelméletnek nincs, vagy hiányos, vagy operacionalista meghatározottságú a háttérelmélete. Igenis adható definíció a látens képességfejlettségre, nincs arról szó, hogy az csak operacionális definícióval rendelkezne. Más probléma van: a konkrét mérésekkel összefüggésben soha nem kapunk eligazítást abban a kérdésben, hogy az intervallumskála létezésének legalább az a szükséges feltétele fennáll-e, hogy az egyes skálákon mért eredmények egymással pozitív lineáris kapcsolatban állnak. Általánosabban: a valamirevaló esetekben (kicsit is komplex képességek esetén) *nem tudunk különbségi struktúrát konstruálni a képességfejlettségek halmazán*.

Vagyis a hamis recept a klasszikus tesztelmélet keretei között így hangzik: alkoss olyan algoritmusokat, amelyek számokat produkálnak (számokat rendelnek hozzá az emberekhez), ezeket tekintsd a képességfejlettség látens értéke becslésének, és abban a hitben ringathatod magad, hogy mértél valamit. Csak azt nem tudod, hogy mit, mert ha megkérdezik tőled, akkor nem tudsz úgy válaszolni, hogy az a reprezentációs méréselmélet szerint elfogadható legyen. Nem tudod azt mondani, hogy kérem szépen, az emberek között (az adott képességükkel, annak fejlettségével összefüggésben) ilyen és ilyen relációk vannak, ezek így és így felelnek meg az emberekhez rendelt számok közti, a matematikában ismert relációknak, és valójában ez a megfelelés alapozza meg, hogy a számokkal csuda dolgokat vigyünk végbe (a szórások kiszámításától a faktoranalízisig és azon túl), mert a reprezentációs méréselméletben leírt különbségi vagy extenzív struktúra fedezhető fel a képességfejlettségek halmazán.

Legyünk azonban óvatosak! Kijelentéseim elhamarkodottak is lehetnek. Az, hogy a tesztek megalkotói és felhasználói, akik a klasszikus tesztelméleti sémát igyekeznek alkalmazni, nem írnak le következetes matematikai (méréselméleti) definíciókat, még nem jelenti azt, hogy a háttérben ilyenek nem is létezhetnek. Most megkísérlek egy elemzést e kérdésben, vagyis keresem az ismert megfontolásokban azt a magot, amely különbségi struktúra alkotását tenné lehetővé. Közben természetesen arra törekszem, hogy minél teljesebben figyelembe vegyem a szakirodalomban rendelkezésre álló megfontolásokat. Az elemzés azonban maximum részsikereket hoz, csak bizonyos nagyon egyszerű, és a pedagógia számára valószínűleg érdektelen esetekre vonatkozóan alkothatók matematikailag is korrekt rendszerek (s még azok esetében is sok empirikus vizsgálatra lenne szükség, hogy bebizonyítsuk, valóban teljesülnek a reprezentációs méréselméletben szereplő szükséges feltételek).

A klasszikus tesztelmélet alapjai

A *klasszikus tesztelmélet* magyarázata során szokás kiindulni ezen elmélet *alapegyenletéből*. Ez az egyenlet azt mondja, hogy egy adott tesztben mért pontszám egyenlő két további, nem megfigyelhető mennyiség összegével: $X = T + E$. Itt X a megfigyelhető mennyiség, a tesztben szerzett pontszám, míg T az a pontszám – ahogyan ezt a klasszikus tesztelmélettel kapcsolatos szokásos, ám nem túl pontos magyarázatok leírják –, amit akkor szerezne a vizsgált személy, ha nem lépne fel semmilyen hiba, zavaró, vagy segítő véletlen körülmény. Ez az igazi pontszám (True Score), vagy látens érték. E a hibatag, ami azt fejezi ki, hogy milyen mértékben tér el egymástól a valódi érték (T) és a mért érték (X). *A klasszikus tesztelméletben ez a bizonyos T a képességfejlettség*. T és E látens értékek, sohasem ismerhetjük pontos értéküket. Amit ismerünk, az a szerzett pontszám, vagyis az X.

Itt azonban álljunk meg egy kicsit! Amit most leírtam, az a klasszikus tesztelméletet bemutató írások szokásos gondolatmenete. Mint látható, szó sincs benne méréselméleti követelményekről, relációkról, relációs struktúrákról, homomorfizmusról, különbségi struktúráról. Anélkül, hogy ezek az alapfogalmak a tesztelés feladatára vonatkozóan definiáltak lennének, az elmélet megalkotói és későbbi leírói azonnal a számok képességfejlettségekhez való hozzárendelésének feladatát végzik el. Fogalmunk sincs még (nem is lesz) milyen D négyváltozós, az embereken az adott képességükkel kapcsolatban értelmezett relációval van dolgunk, azonnal nekiugrunk a számok hozzárendelésének. Így igencsak nehéz lesz bebizonyítani, hogy ez a szám-hozzárendelés *homomorfizmus*. Ahhoz szüségünk lenne a mért valamiken értelmezett *relációkra* is. Márpedig, ha nincsenek relációk, és ezért nem állíthatjuk, hogy a számok hozzárendelése homomorfizmus, akkor semmi jogunk nem lesz azt állítani, hogy a hozzárendeléssel előállított *skálák intervallum jellegűek*. És ha nem tudhatjuk, hogy itt legalább intervallumskálákról van szó, akkor még egy „nyamvadt” átlagot sem számolhatunk a kapott adatokon.

Ezen a ponton megállíthatnánk az elemzést, és azt mondhatnánk: *a klasszikus tesztelmélet nem követi a reprezentációs méréselmélet logikáját*, amely ma az egyetlen fejlett statisztikai eszközök alkalmazását legitimáló elméleti alap, tehát a klasszikus tesztelmélet illegitim. Csakhogy – s ez a megfontolás már fentebb is szerepelt – a klasszikus tesztelméletet leíró szerzők nem szólnak ugyan a reprezentációs méréselméleti háttérről, ám ez a háttér ettől még létezhet. Egyelőre egyáltalán nem bizonyítottuk be, hogy nem is lehet előállítani a megfelelő elméleti alapokat. Az ezután következő elemzés logikája a következő lesz: követjük a klasszikus tesztelmélet kifejtésének szokásos rendjét. Elfogadjuk átmenetileg azt a – mint láttuk – problematikus eljárást, hogy a mérés vonatkozásában azonnal a számoknak a mért egyedekhez való hozzárendelésével foglalkozunk, anélkül, hogy lenne empirikus relációs rendszerünk. Úgy képzeljük, mintha mégis lenne, sőt, mintha igaz lenne, hogy minden szám hozzárendelés homomorfizmus. Úgy teszünk tehát, hogy minden rendben van az elméleti alapokat tekintve (miközben persze semmi nem lesz rendben), és feltesszük a kérdést, hogy vajon az így előállt skálák egymással pozitív lineáris kapcsolatban állnak-e. És, ahogy pestiesen mondják: „itt ugrik majd a majom a vízbe”. Triviális lesz, hogy a kicsit is összetett képességek esetén ez nem áll. Ezzel szinte matematikai precizitással láttathatjuk, hogy a klasszikus tesztelmélettel az a baj, hogy nem csak, hogy az eddigi munkálatok nem produkáltak ERR-t, hanem a kicsit is összetettebb képességek esetén egyszerűen azért nem formáltak a szakemberek relációs rendszert vagy rendszereket, mert a klasszikus tesztelmélet által ajánlott módon *nem is lehet*.

De hogy ez a „szép program" kivitelezhető legyen, kell egy nagyon fontos megjegyzést tennem. Alapvető jelentősége van annak – és ez az elemzés során szépen látszik is majd –, hogy az érvelésben végig az adott képességhez tartozó *összes feladatra* gondolunk. A klasszikus tesztelméleti problémák kiküszöbélésével kapcsolatos gyakorlati tevékenységek, a „javítgatások" során (és ugyanez a helyzet a modern tesztelmélet alkalmazásával is) számtalanszor kezdenek a kutatók „válogatni a feladatok között". Bizonyos vizsgálatok elvégzése arra a következtetésre vezeti őket, hogy egyes feladatok úgymond nem felelnek meg, nem illeszkednek a modellbe. Ekkor e feladatokat elhagyják, nem veszik figyelembe őket a tesztelés során. Ez egy önkényes, semmivel sem igzolható eljárás. Ha az adott kiszűrt feladat a vizsgált képességhez tartozik, akkor milyen joggal zárjuk ki? Ha minden „deviánsnak" minősülő feladatot figyelmen kívül hagyunk, akkor nem arról van-e szó valójában, hogy kiválogatjuk azokat a feladatokat, amelyek alkalmasak rá, hogy segítségükkel a tesztekhez rendelhető, elméletileg definiált képességfejlettség adatsorok egymással pozitív lineáris kapcsolatban álljanak? És nem arról van-e szó, hogy már valójában nem is az eredetileg vizsgálni kívánt képességről beszélünk? Az egész eljárás, pontosabban annak eredménye nem azt támasztja-e alá inkább, hogy a lehetséges skálák mégsem intervallumskálák, vagyis az egész mérés megalapozatlan? Az én véleményem szerint ez a helyzet. Igencsak tiszteletlenül az eljárást elneveztem a *tesztfejlesztés trükkjének*. A tesztekkel komoly feladatokat végző kutatók általában nagyon büszkék rá, ha jól elsajátították a tesztelemzés, tesztértékelés és tesztfejlesztés metódusát. És valóban, ez a metódus korrekt a tesztekben szereplő itemek nagy részétől különböző módon viselkedő, „deviáns" itemek megtalálásában. És amikor e korrekt módszerrel megtalálják e nagy tudású kutatók a sorból kilógó itemeket, és azokat száműzik a tesztből, akkor éppen azt mondják, csak nem hangosan, hogy a képességhez tartozó feladatokból összeállítható tesztek nem ugyanazt mérik. De hogy ezt az is lássa, aki eddig még nem foglalkozott ilyen kérdésekkel, nem ússzuk meg, ki kell fejtenem a klasszikus tesztelmélet ezen elemzés számára releváns gondolatait.

Kezdjük (vagy inkább folytassuk) a képességfejlettség, általánosabban a teszt által mérni kívánt konstruktum kvantitatív definíciójával, vagyis annak a meghatározásával, hogy mit mérünk! Ezt a feladatot a klasszikus tesztelmélet, ahogy már korábban is kiemeltem, igencsak tisztességesen megoldja. Ám a továbbiak megértéséhez szükségünk lesz a *valószínűségi változó* és a *várható érték* fogalmának ismeretére. Nem precíz matematikai leírás következik, inkább igyekszem szemléletesen kifejteni e fogalmak jelentését.

A valószínűségi változó a véletlen jelenségek leírására használt fogalom. Egy-egy véletlen jelenség esetén az eredményt, a kimenetelt jellemzi valamilyen megkülönböztető sajátosság, nagyon sok esetben egy szám. Például a tesztfeladatok megoldása során, egy konkrét személy esetén a szerzett pontszám (x) egy valószínűségi változó egy értéke. A véletlennek van befolyása arra, hogy hány pontot szerez az illető (legalábbis ezt a modellt alakítjuk ki). Ha képesek lennénk többször megoldatni a teszt feladatait az adott személlyel, mégpedig úgy, hogy minden tesztelés kezdetén ugyanabban az állapotban lenne (erre persze a legtöbb esetben nem vagyunk képesek, lényegében soha), akkor lenne sok x értékünk, amelyek kisebb vagy nagyobb mértékben különböznének egymástól, és eltéréseiket a véletlen határozná meg[20]. Ezért a tesztváltozó (a pontszám,

[20] Az, hogy a véletlen határozza meg, hogy a mért tesztpontszám a valóságos pontszámtól mennyire tér el, egy kissé elnagyolt kijelentés. Nagyon sok esetben tudunk számtalan tényezőt mondani, amely befolyásolja ezt a

vagyis X) valószínűségi változó, ahogyan az E, a hiba is az. Ha X az egyik mérésben erősebben eltér a valódi-, látens értéktől, vagyis a T-től, akkor E abszolút értéke nagyobb lesz, ha az eltérés kisebb, akkor E értéke is kicsi lesz. T viszont állandó (egy adott személy esetében), gondolhatunk rá, mint valószínűségi változóra, amely azonban csak egyetlen értéket vehet fel.

Beszélhetünk arról, hogy egy valószínűségi változó valamekkora valószínűséggel vesz fel valamilyen értéket. Amely értékek kisebb valószínűségűek, azok kevesebbszer fordulnak elő a kísérlet többszöri elvégzése során, és fordítva (a valószínűségszámításban szokás kísérletnek mondani a vizsgált jelenséget, folyamatot, így tehát e szóhasználat szerint egy ember tesztelése is egy kísérlet).

Furcsa lehet, hogy e megfontolásokat egy amúgy elvégezhetetlen feladatra, egy ember többszöri, ugyanolyan kezdőállapotban történő tesztelésére hivatkozva fogalmazom meg. Ez azonban inkább a szemléletességet szolgálja. Ugyanis valószínűségi változók, s a lehetséges események valószínűségei akkor is léteznek, ha nem tömegjelenségről van szó, vagyis ugyanazt a kísérletet valójában nem, hogy sokszor, de még kétszer sem tudjuk elvégezni, csak egyszer. Annak, hogy egy ember egy konkrét eredményt ér el, vagyis konkrét pontszámot tudhat magáénak egy tesztelésben, van valamekkora esélye akkor is, ha a kísérletet nem tudjuk többször végrehajtani. Számtalan olyan jelenség van, amely megismételhetetlen, az egyes lehetséges eseményekhez valószínűségek hozzárendelésének mégis van értelme. Például, a Föld létrejöttekor annak a valószínűsége, hogy kialakul élet a bolygón, márciusban annak a valószínűsége, hogy egy tanuló történelemből ötöst kap majd tanév végén, stb. Fogadjuk el tehát, hogy egy ember adott pontszáma, mint tesztelési eredmény egy valószínűségi változó egy értékének megvalósulását jelenti, illetve az ilyen eseménynek igenis van valószínűsége.

A *várható érték* is egy valószínűségszámítási fogalom. Ha egy kísérletnek többféle kimenetele lehet, akkor elvileg bármelyik lehet ezek közül egy konkrét esetben az eredmény, mégis, bizonyos eredmények gyakrabban, mások kevesebbszer jönnek ki, amennyiben ugyanolyan körülmények között sokszor meg tudjuk ismételni a kísérletet. Vagy ha nem tudjuk sokszor elvégezni a kísérletet, akkor bizonyos kimeneteleket nagyobb eséllyel várhatunk, mert azoknak nagyobb a valószínűségük. Ha a kimeneteleket számok jellemzik, mint a tesztelés esetében is, akkor kiszámíthatjuk e számoknak a valószínűségekkel súlyozott középértékét. Milyen érték lehetne az elvégzett kísérletek esetén kapott számok számtani közepe? Ez a kérdés akkor, ha sokszor el tudjuk végezni a kísérletet. Ami gyakrabban jön ki eredményként (mert nagyobb a valószínűsége), az többször szerepel a számtani közép kiszámításában, s ezt a súlyt az adott eredmény valószínűsége határozza meg. Summa summárum, a várható érték *a pontszámokból és a valószínűségeikből képzett szorzatok összege*. Amelyik pontszám valószínűbb, az egy nagyobb szorzótényezővel van jelen az összeadásban, amelyiknek kisebb a valószínűsége, az csak kisebb mértékben járul hozzá az összeghez.

A klasszikus tesztelmélet fontos axiómája, hogy a hiba várható értéke nulla. Ahogy mondani szokás: nincs szisztematikus hiba. A teszteredmény, vagyis X úgy alakul ki, hogy az állandónak tekintett T látens értékhez képest vagy több pontot, vagy kevesebbet szerez a vizsgált személy, vagy véletlenül pontosan annyit. Itt azonban van a kifejtésben egy logikai probléma. A klasszikus tesztelméletnek van három kiinduló „feltételezése”, alapállítása, amelyek azonban nem függetlenek egymástól:

---

különbséget, vagyis valójában mégsem egészen a véletlenről van szó. Ám úgy képzeljük, ez a modellünk, hogy a hatások sokfélék, egyenként kicsik, így együtt – a tesztpontszám mértékét tekintve – 0 „átlagú” (pontosabban zérus várható értékű), normális eloszlású értékekkel nyilvánulnak meg. És ez nagyon hasonlít ahhoz, amit a „tisztán” véletlenszerű hatás produkálna.

- Az alapegyenlet: $X = T + E$.
- A hiba (E) várható értéke zérus.
- A „true score", a valódi, látens képességfejlettség a konkrétan mért tesztpontszámok várható értéke.

Azt állítom, hogy a harmadik állítást definícióként kell megfogalmazni, és akkor a másik kettő már a valószínűségszámítás törvényszerűségei alapján levezethető állítás. Ez azért fontos, mert az alapegyenletet és a hiba várható értékének zérus voltát axiómaként szokás kezelni, pedig valójában arról van szó, hogy a harmadik állítás egy definíció, és ebből az alapegyenlet is, és a hiba várható értékének zérus volta is már következik. *Lord* és *Novick* (1968) definíciója a látens, a valóságos képességfejlettséggel kapcsolatban valóban a kiindulópont: *a valódi, látens érték a kapható tesztpontszámok várható értéke*. Az már valószínűségszámítási következmény, hogy akkor minden egyes konkrét tesztelésnél maga a konkrétan nyert pontszám (X) ettől eltér valamennyire, de akkor az azt jelenti, hogy a mért értéket a valódiból úgy kapjuk meg, hogy hozzáadunk egy hibatagot (E), és mindezekből adódik, hogy ennek a hibatagnak a várható értéke 0[21].

Fontos még, hogy a hiba (E) nem függ össze a T-vel. Ez valódi axióma, vagyis olyan jelenségeket vizsgálunk, amelyekre ez igaz. Azt jelenti (példát használva a magyarázathoz), hogy nem fordul elő, hogy a nagyobb igazi pontszámmal (T) rendelkező személy hibái nagyobbak, vagy éppen kisebbek lennének. (Ezt az összefüggést azonnal egy kissé másképpen is leírom egy újabb reprezentációban.)

Nagyon fontos, hogy pontosan értsük: ezek a fogalmak – egyelőre – az egyes vizsgált személyekre vonatkoznak. Egy adott embernek van látens T képességfejlettsége, ehhez képest ő ér el egy konkrét mérésben X pontot, és a kettő között az ő esetében lesz a különbség, a hiba az E. Ezt azért kell kiemelnem, mert még a szakirodalomban is előfordul az egyénre jellemző átlag-, vagy szórás-, standard hiba értékek összemosása, összekeverése az egy mintán adott mérés esetén jellemző átlagokkal, szórásokkal, standard hibákkal. Amit eddig leírtam, az még nem tenné indokolttá, hogy megfogalmazzam a következő kijelentést: a teszteredmények varianciája[22] egyenlő a valódi értékek varianciájának és a hibavarianciának az összegével. Ugyanis ha az elemzést egy személy esetében végezzük, akkor – ez is remélem könnyen belátható – a szerezhető pontszámok varianciája megegyezik a hibavarianciával, és a valódi érték varianciája pedig zérus, hiszen ez az érték állandó. És bár igaz, hogy az első (X varianciája) egyenlő a másik két variancia összegével, de ez egy érdektelen kijelentés.

Rögzítsük a lényeget! A klasszikus tesztelmélet kiindulópontja: *egy személy képességfejlettsége nem más, mint a tesztben általa szerezhető pontszám, mint valószínűségi változó várható értéke.* Hogy a Kedves Olvasó ezt az elvont meghatározást jobban értse, bemutatok egy rendkívül egyszerű szemléltető példát.

Szemléltető példa

Tegyük fel, hogy a mérésre használt tesztben van négy feladat. Ha valóságosan elvégezzük a tesztelést, akkor mindenféle körülmények befolyásolhatják a vizsgált személy

[21] Ha $M()$ jelöli a várható értéket, zárójelben azzal a valószínűségi változóval, amelynek a várható értékét éppen meghatározzuk, akkor $X = T + E$ alapján $E = X - T$, $M(E) = M(X - T) = M(X) - M(T) = T - T = 0$.

[22] A variancia a szórás négyzete (önmagával vett szorzata). A szórás (a statisztikai, matematikai szakkönyvekben, tankönyvekben a precíz képletet, definíciót is megtalálhatja a Kedves Olvasó) az egyes mért értékek, az adatok átlagtól való eltérésének átlagához van nagyon közel, az adatok szórtságára jellemző mérték. A variancia is ezt az átlagos eltérést jellemzi, csak annak a négyzetéhez van közel, és azért van jelentős szerepe, mert az egymástól független változók adatai összegének, mint új adatsornak a varianciája egyenlő az összetevő adatrendszerek varianciáinak összegével.

teljesítményét, lehet, hogy egyetlen feladatot sem tud megoldani, de lehet, hogy egyet, vagy kettőt, vagy hármat, vagy négyet. Bármelyik eset előfordulhat, sőt, az esetek egy része többféleképpen is. Például, az, hogy pontosan egy feladatot sikerült jól megoldani, akár négyféleképpen is előfordulhat, attól függően, melyik az az egy feladat, amellyel való foglalkozás sikeres volt.

Tegyük föl, hogy tudjuk, a vizsgált személy mekkora valószínűséggel tudja jól megoldani a feladatokat. Meg is adok egy konkrét példát:

$$p_1 = 0{,}8;\ p_2 = 0{,}6;\ p_3 = 0{,}5;\ p_4 = 0{,}4.$$

Itt p_1-gyel jelöltem annak valószínűségét, hogy az első feladatot jól meg tudja oldani a vizsgált személy, és így tovább (nagyon fontos, hogy ezeket a valószínűségeket a konkrét mérésekben sohasem tudjuk, legfeljebb becsléseket szolgáltathat a mérés egésze az értékekre). Mindegyik feladat megoldása 1 pontot ér. Mindegyik feladat megoldásához, mint kísérlethez hozzárendelhetünk egy valószínűségi változót. Ezeknek vagy 0, vagy 1 az értéke, aszerint, hogy az egyes feladatokban valóságosan hány pontot szerzett a vizsgált személy. E valószínűségi változók várható értéke éppen az adott feladat helyes megoldásának valószínűsége. Ez könnyen belátható:

$$\textit{várható érték} = 1 \cdot p_{\textit{jó a megoldás}} + 0 \cdot p_{\textit{rossz a megoldás}} = p_{\textit{jó a megoldás}}.$$

A teszt egészében szerezhető pontszám is egy valószínűségi változó, méghozzá a négy most bevezetett valószínűségi változó összege. Egyszerű valószínűségszámítási összefüggés, hogy ha egy valószínűségi változó más valószínűségi változók összege, akkor a várható értéke az egyes valószínűségi változók várható értékének összege. Ahhoz tehát, hogy megtudjuk, a tesztben szerezhető pontszám várható értéke mennyi, az egyes feladatok jó megoldásához tartozó valószínűségeket kell összeadni.

$$\textit{A pontszám várható értéke} = p_1 + p_2 + p_3 + p_4 = 0{,}8 + 0{,}6 + 0{,}5 + 0{,}4 = 2{,}3$$

A várható érték tehát 2,3. Ennyi az adott teszttel mérhető képességfejlettség. Ez az adat önmagában még persze semmit nem mond. Viszont ha meghatározzuk másoknak is a képességfejlettségét, akkor a kapott adatok összehasonlítása már hordoz információt.

A kapott eredmény azt jelenti egyébként, hogy a vizsgált személy a feladatok felénél valamivel többet tud megoldani. Persze a megoldott feladatok száma csak 0, 1, 2, 3 vagy 4 lehet, nem lehet „egy kicsivel többet megoldani 2-nél”.

Vigyázat! Ilyen számítást valóságosan soha nem lehet végezni. Ennek oka, hogy soha nem ismerjük pontosan az egyes feladatok jó (és rossz) megoldásának valószínűségét (azt csak becsülni tudjuk). Ezt a 2,3 pontot, vagyis a *T*-t tessék úgy tekinteni, mint amit amúgy nem ismerünk, de ott van a jelenség mögött, mint a teszt megoldása során a pontszám várható értéke. Mint látható, nem egyezik meg egyik valóságosan mérhető pontszámmal sem (0; 1; 2; 3; 4). Azoknak a valószínűségekkel súlyozott közepe.

Most azonban váltsunk nézőpontot! Eddig egy személy esetében vizsgáltuk a teszteléssel kapcsolatos értékeket, valószínűségi változókat. Most tegyük összetettebbé magát a kísérletet, közelítsünk a statisztikai vizsgálatok tipikus szituációjához! A vizsgált kísérlet legyen az, hogy egy populációból véletlenszerűen kiválasztunk egy személyt, és vele oldatjuk meg a teszt feladatait. A kapott, szerzett pontszám, a valódi érték és a hibatag továbbra is valószínűségi változók, a véletlentől (is) függnek, azonban a „true score”, a *T* most már nem lesz állandó. Szintén a véletlentől függ, hogy kit választunk ki a populációból a mérésre. A pedagógiai, pszichológiai tudományos, valamint a nem az egyes tanulókra vonatkozó értékelési szituációk éppen ilyenek. Ha így értelmezzük a kísérletet magát, akkor már nem igaz, hogy a T varianciája nulla, és hogy az X és az E varianciái megegyeznek. Alapvető jelentőségűvé válik az ugyan már korábban kimondott, de ott még nem igazán fontos axióma, vagyis az E és a T függetlensége. Ezt most úgy is fogalmazhatjuk, hogy az E és a T Pearson-féle korrelációs együtthatója nulla.

A szakkönyvekben a klasszikus tesztelmélet kifejtése mindig ebben az utóbbi modellben történik. Fontosnak tartottam azonban megmutatni az egyes személyekre vonatkozó megfontolásokat is, mert azok, akik eddig nem foglalkoztak a kérdéssel elmélyültebben, elveszhetnek a definíciók között, ha e két szintet nem tudják megkülönböztetni. Ráadásul a megfontolások mindig vissza-visszatérnek az egyes vizsgált személyekkel kapcsolatos mennyiségekhez. Azt kell jól megérteni, hogy a tsztelések esetén a valószínűségi változók két szintjével kell számolnunk. Már az egyén szintjén is valószínűségi jelenségekről van szó, és amikor egyének egy populációját vizsgáljuk, akkor egy erre épülő szintet elemzünk, és itt is valószínűségi megfontolásokra van szükségünk.

Azonos eredetű tesztek

Mi az értelme a valódi értéknek, vagyis a konkrét tesztben elérhető pontszámok adott személyre vonatkozó várható értékének? Az, hogy bizonyos mértékig jellemzi a vizsgált személy képességfejlettségét. Gyorsan tegyük hozzá: úgy jellemzi, *ahogyan az éppen használt konkrét teszt mutatja*. Ha másik tesztet készítenénk ugyanezen képesség fejlettségének mérésére, akkor még a valódi érték, a T is megváltozna nagy valószínűséggel. Ehhez elég csak annyi, hogy abban a tesztben mások legyenek az egyes feladatok jó megoldásának valószínűségei. Ennek bekövetkezése igencsak valószínű. De ez nem probléma, ezt korábban már megmutattam, amikor a klasszikus tesztelmélettel szemben felvetett kritikák közül a felszínes elemzéssel megfogalmazható ellenvetésről volt szó.

Most azonban e problémát részletesebben is be kell mutatnom. *Lord és Novick* klasszikus könyvükben (1968) már foglalkoznak ezzel a kérdéssel, ők vezették be a *párhuzamos* és a *tau-ekvivalens* teszthalmazok fogalmát (a részletek az itt is hivatkozott szakirodalomban megtalálhatók). Nekünk most egy harmadik fogalomra van elsősorban szükségünk, az *azonos eredetű tesztek* fogalmára, amit *Karl Gustav Jöreskog*, svéd statisztikus alkotott meg 1971-ben megjelent művében. Egy részletes, a tesztek közötti kapcsolatok matematikai bemutatását tekintve pontos, kimerítő leírást közölt *Rolf Steyer* 2001-ben, én is most erre az utóbbi forrásra támaszkodom. *Steyer* bemutat a témában több fogalmat (mintegy szintetizálva a már korábban született elemzéseket) az ugyanazon konstruktum méréséhez összeállítható tesztek egymáshoz való viszonyának jellemzésére. Ebből bennünket most csak az „azonos eredetűség” fogalma érdekel[23]. Az azonos eredetűség első megközelítésben nem más, mint a pozitív lineáris összefüggés érvényesülése. Tesztek egy halmazában van valahány, de egynél több tesztünk. Ha kiválasztunk közülük kettőt, és a vizsgált populáció tagjainak a két teszthez tartozó valódi (látens) képességfejlettségei között pozitív lineáris összefüggés van, akkor esély van arra, hogy e két teszt azonos eredetű. Az azonos eredetűséghez azonban még az is kell, hogy bármely két teszt esetén a populáció tagjain elvégzett két méréskor a hibák (az E értékek) ne függjenek össze egymással (korrelálatlanok legyenek). Vagyis ne forduljon elő például

---

[23] A fordítás nem a legszerencsésebb, de nem találtam jobbat. Az eredetiben *Steyer congeneric* tesztekről ír.

az a helyzet, hogy akik az egyik tesztben a valódi (látens) képességfejlettségüknél magasabb pontszámot szereznek, azok a másik tesztben is jellemzően így teljesítenek, és fordítva. Ha tehát e két feltétel érvényesül tesztek egy halmazában (a pozitív lineáris összefüggés és a hibák korrelálatlansága bármely a halmazból kiválasztott tesztpár esetében), akkor e teszteket azonos ertedetűeknek mondjuk.

A megadott tulajdonságok igencsak erős megkötések, a párhuzamos tesztekre, illetve a tau-ekvivalens tesztekre vonatkozók még ennél is szigorúbbak, és ráadásul az azonos eredetűség speciális eseteit jelentik. Az azonos eredetű tesztek fogalmára részben éppen azért volt szükség, mert a klasszikus tesztelmélet reprezentációs méréselméletnek való megfelelése már régóta kérdés, és természetes volt az igény, hogy olyan tesztek viselkedését vizsgálják, amelyek mögötti valódi pontértékek egymással pozitív lineáris kapcsolatokban állnak. Csak ekkor lehet reményünk rá, hogy a vizsgált problémát tekintve (képességfejlettség mérése) léteznek méréselméleti értelemben megfelelő intervallumskálák. Ha már az sem igaz, hogy az összes tesztből kiválasztva bármely kettőt, az azoknak megfelelő két valódi, látens pontérték halmaz egymással pozitív lineáris kapcsolatban áll, akkor ezen a halmazon nem létezhet különbségi struktúra, a kétségkívül megkonstruálható skálák nem intervallum szintűek.

Ahogy én látom, a szakemberek a problémát mindig úgy fogalmazták meg, hogy miképpen lehet kiválasztani az azonos eredetű teszteket, vagy ha vannak már tesztjeink, akkor miképpen mondhatnánk meg róluk, hogy azok azonos eredetűek. Átfogó leírást ad, és a statisztikai vizsgálatok lehetőségét mutatja be szintén *Rolf Steyer* (1989). Nem kétlem, hogy egy képességhez tartozó számos feladatból alkothatók olyan tesztek, amelyeknek a halmaza azonos eredetű (ha például találunk két azonos eredetű tesztet, akkor már van is egy – kételemű – teszthalmazunk, amely megfelel). Ezek esetén felmerülhetne, hogy mivel a mérhető képességfejlettségek (a valódi értékek) pozitív lineáris kapcsolatban vannak egymással, esetleg állhat az egész mögött egy a reprezentációs méréselméletnek megfelelő struktúra. De – ha van is – ez a struktúra milyen képességhez tartozik? Az eredetileg vizsgálni szándékozott képességhez biztosan nem, mert a reprezentációs méréselméletnek – esetleg – megfelelő tesztek halmaza nem tartalmazza az összes lehetségest.

Persze mesterségesen „előállíthatunk" egy képességet. Ehhez veszünk egy tesztet, és tekintjük az összes többi olyat, amelyekkel ez a kiválasztott teszt együtt azonos eredetű teszthalmazt alkot. E teszthalmaz összes tesztjében szereplő összes feladat reprezentálja ekkor a „képességet". *Rolf Steyer* igencsak alaposan bemutatott, és részleteiben is bizonyított tételei (Steyer 1989) az ilyen „képességekre" vonatkoznak, ezért gyakorlatilag semmit nem mondanak a tényleges, a pedagógiában is tárgyalt képességekről[24].

A valódi érték tehát a klasszikus tesztelméletben – elvont értelemben – a tesztelés során szerezhető pontszám várható értéke. A méréselmélet alkalmazása során jól meg kell érteni, hogy a mérés tárgya általában nem jellemezhető egyetlen számmal. Egy szakasznak a geometriában van hossza, de ez kiinduláskor egy kvalitatív tulajdonság, és nem egy „a szakaszban rejtőző, csak kezdetben nem ismert" szám. Lehet hozzárendelni a szakaszhoz számot, de az *egy adott reprezentációban fejezi ki a hosszát*, ha

[24] Az itt tárgyalt eset, tehát a „mesterségesen előállított képesség" azonban a vázolt probléma ellenére is nagyon fontos szerephez jut. Ugyanis – ezt később még részletesebben is tárgyalom – a nagy nemzetközi tudásmérések (PISA, stb.), valamint a hazai kompetenciamérések éppen ezt a helyzetet valósítják meg. Valójában nem valamifajta matematikai- vagy szövegértés képesség fejlettségének méréséről van szó, hanem olyan feladatok alkalmazásáról, amelyeknek a tanulók általi megoldása vagy nem megoldása fontos információkkal szolgál számunkra, de amelyekből azonos eredetű tesztek építhetők fel. Ehhez azonban annyi még hozzátartozik, hogy e tesztelések esetében már a modern tesztelmélet alkalmazásáról van szó, ezért is tárgyalom ezt a lehetőséget részletesebben majd az azt leíró fejezetben.

megváltoztatjuk a mértékegységet, más lesz ez a reprezentáló érték. A szám azonban már nem a szakasz *hossza*, hanem a hossz *mértéke*. Önmagában az még nem probléma, ha más tesztek egy képességfejlettséggel kapcsolatban eltérő eredményt produkálnak. Akkor probléma csak, ha a szóba jövő populáció tagjain használva a különböző teszteket, az azokkal becsülhető látens értékek nem pozitív lineáris összefüggésben állnak egymással, vagyis nem azonos eredetű tesztek segítségével becsülhetők. Ezt a követelményt azért kell teljesíteniük a teszteknek, mert azt gondoljuk, hogy a látens képességfejlettség értékek intervallumskálán helyezkednek el. Ez olyan, mint a már többször idézett hőmérsékleti skálák esete. Van abszolút-, Celsius-, Fahrenheit-, Reaumur skála a hőmérséklet méréséhez, de ha sok hőmérséklet adatot a skálák közül tetszőlegesen választva kettőben megadunk, akkor az egyikből a másikba az átszámítás *minden adat esetén* ugyanazon lineáris ($y = ax + b$ alakú) összefüggés segítségével végezhető el. A klasszikus tesztelmélet keretei között e követelmény teljesülését semmi sem garantálja.

Egy kissé másképpen magyarázva: a hőmérsékletmérésben a hőmérők megfeleltethetők a képességfejlettség mérés tesztjeinek. Ha éppen egy Celsius fokokra kalibrált hőmérőnk van, de mi Fahrenheit fokokban szeretnénk megkapni a hőmérsékletet, akkor nem kell mást tennünk, mint a Celsius fokokban kapott mérési eredményt behelyettesíteni a megfelelő képletbe, és kis számítás eredményeként megkaphatjuk, hogy mekkora a hőmérséklet Fahrenheit fokokban. Ezt azért tehetjük meg, mert tudjuk, hogy bármekkora is az egyik skálán a hőmérséklet, mindig ugyanazt a (lineáris) képletet kell használni, ha tudni akarjuk, hogy a másik skálán mekkora a hőmérséklet. Lehetne mérni úgy hőmérsékletet, hogy az egyes értékek ne legyenek mind ugyanabban a lineáris kapcsolatban mondjuk a Celsius-skálán ugyanazon testeken mért hőmérsékletekkel? Nyilván lehetne, ha a fizikai törvényeknek nem megfelelő, az elméleti háttér által adott követelményeket nem teljesítő, hibás mérést bonyolítanánk le. Mivel a képességfejlettség mérés esetén ilyen elméleti háttér nincs, valójában nincs ERR, ezért nem mondhatjuk, hogy a mérés hibás, hiszen nincs olyan elméleti alap, vagy másképpen háttér, amelynek meg kellene felelnie. Nem hibás, „csak" nem lehet különbségi struktúra a hátterében. Adott esetben, ha erre jó okunk van, azt mondhatjuk: biztos, hogy nem állhat különbségi struktúra a hátterében. Akkor viszont a kétségtelenül létező szám-hozzárendelés ellenére sem végezhetünk az adatokon olyan statisztikai vizsgálatokat, amelyeket nagyon sok szakember elvégzett, és még ma is elvégez az adatokon.

Kétségek a klasszikus tesztelmélet megalapozásával kapcsolatban

A tesztek alkalmazása általában a következő „logikával" történik: Áttekintve a téma szakirodalmát kijelöljük azokat a feladattípusokat, amelyeket szerepeltetnünk kell a tesztekben. Feladatokat konstruálunk a vizsgálandó képességgel kapcsolatban, mindegyikben valamilyen, a képesség alkalmazásával összefüggő teljesítményre van szükség. Megformálunk a lehetséges feladatokból egy tesztet. Bármely vizsgált személy esetén vannak valamilyen nem ismert valószínűségei a tesztbe került feladatok jó megoldásának, ez pedig objektíve meghatároz minden vizsgált személyre egy-egy a konkrét teszthez tartozó T értéket, ami a fentiek alapján nem más, mint a tesztben a vizsgált személy által jól megoldható feladatok gyakoriságának (vagy ha olyan a számítás, akkor a relatív gyakoriságának), a szerezhető pontszámnak a várható értéke. Készítsünk ugyanakkor egy másik tesztet is, amely ugyanannak a képességnek a fejlettségét méri, vagyis szintén az előbb körülhatárolt feladathalmazból választjuk az itemjeit, de részben vagy teljes mértékben másokat. Itt is vannak jó feladatmegoldásokhoz tartozó valószínűségek, itt is kiszámíthatóak lennének minden személy esetén a T' értékek, ha ismernénk e valószínűségeket. A vizsgált populáció egyedeihez tehát hozzá lenne rendelve egy-egy *T* és egy-egy *T'* érték (sokadszor ismétlem, ezek nem meghatározhatók, most csak „elképzeljük" őket). Van tehát két elméleti adatsorunk. Ha a vizsgált képesség fejlettsége reprezentációs méréselméleti értelemben intervallumskálán mérhető lenne, akkor annak mindenképpen teljesülnie kellene, hogy e két adatsor egymással pozitív lineáris összefüggést mutatna. Az egyik adatsor egy elemét úgy kell kiszámolni a másik megfelelő eleméből, hogy utóbbit mindig ugyanazzal a pozitív valós számmal meg kell szorozni, és az eredményhez mindig ugyanazt a valós számot hozzá kell adni. Ezen összefüggés fennállását közvetlenül nem tudjuk ellenőrizni még két teszt esetében sem, nem is beszélve az összes lehetséges tesztről. De az elkészített tesztekkel lehet mérni, vagyis a populációban szereplő személyek a tesztelések során tudnak *X* és *X'* valódi tesztpontszámokat produkálni. Ezek kisebb vagy nagyobb mértékben eltérnek a megfelelő *T* és *T'* értékektől. Nem várhatjuk, hogy a valóságosan meghatározható, *X* és *X'* értékeket tartalmazó adatsorok között is a *T* és *T'* értékek köztivel megegyezően pozitív lineáris összefüggés alakul ki, de egy bizonyos értelemben a helyzetnek ehhez közel kell állni.

Használjuk ki most azt a fontos „tényt", hogy a képességfejlettség mérése csak akkor működik (legalábbis a lehetséges tesztek halmaza akkor közös eredetű), ha az itt meghatározó követelmények *minden teszt* esetén fennállnak. Ha a kvalitatív értelmű képességfejlettséghez minden vizsgált személy esetében a tesztben szerezhető pontszám várható értékét rendeljük hozzá (és erre adunk becslést a mérés során, amikor a teszt feladatait a vizsgált személyek megoldják), akkor *bármely* két teszt esetén érvényesülnie kell az eddig is sokszor hangsúlyozott linearitásnak. *Az egy feladatot tartamazó tesztek esetén is(!)*. Ahogy erről már volt szó: egy adott felmért személy esetén az egy feladatos tesztben szerezhető pontszámok (0 vagy 1) várható értéke éppen annak valószínűsége, hogy az illető jól tudja megoldani azt a feladatot. Ez azt jelenti, hogy a képességhez tartozó (a képességet meghatározó) feladatok felmért személyek általi jó megoldásának valószínűségei, mint adatsorok, páronként pozitívan lineárisan összefüggők[25]. Most képzeljük el az összes itt szóba kerülő valószínűségből alkotott mátrixot! Soraiban az egyes személyekhez tartozó valószínűségek szerepeljenek, egymás után az egyes feladatokat tekintve, vagyis az oszlopokban találjuk az egyes feladatokhoz tartozó valószínűségeket. Ha *N* személy van a vizsgált populációban, és ha *L* feladat alkotja a képesség „feladattestét", akkor ez *NL* darab, 0 és 1 közötti szám téglalap alakú sémába (mátrixba) foglalását jelenti. Az előbb leírt feltétel azt mondja, hogy bármely két oszlopban az adatok egymással pozitív lineáris kapcsolatban kell, hogy álljanak (2. ábra).

---

[25] A klasszikus tesztelmélet fogalmait használva ez a követelmény azt jelenti, hogy a valódi pontértékeket tekintve (*T* értékek) bármely két item esetén az item-item korrelációknak 1-nek kell lenni, hiszen pozitívan lineárisan összefüggő adatsorok korrelációs együtthatója 1.

$$\begin{array}{ccccccccc}
p_{11} & p_{12} & p_{13} & p_{14} & p_{15} & p_{16} & p_{17} & p_{18} & p_{19} \\
p_{21} & p_{22} & p_{23} & p_{24} & p_{25} & p_{26} & p_{27} & p_{28} & p_{29} \\
p_{31} & p_{32} & p_{33} & p_{34} & p_{35} & p_{36} & p_{37} & p_{38} & p_{39} \\
p_{41} & p_{42} & p_{43} & p_{44} & p_{45} & p_{46} & p_{47} & p_{48} & p_{49} \\
p_{51} & p_{52} & p_{53} & p_{54} & p_{55} & p_{56} & p_{57} & p_{58} & p_{59} \\
p_{61} & p_{62} & p_{63} & p_{64} & p_{65} & p_{66} & p_{67} & p_{68} & p_{69} \\
p_{71} & p_{72} & p_{73} & p_{74} & p_{75} & p_{76} & p_{77} & p_{78} & p_{79}
\end{array}$$

$$p_{i6} = \alpha p_{i3} + \beta$$

2. ábra: A jó feladatmegoldások valószínűségei mátrix alakba rendezve. Az ábra azt szemlélteti, hogy a 3. és a 6. feladatok jó megoldásához tartozó valószínűségek egymással lineáris összefüggésben állnak

Az NL számú valószínűség azonban – ahogy láttuk – összefügg egymással. Ha például megadjuk az első oszlop számait (számuk N, jelölésük p_{i1}), és tudjuk, hogy a k-adik ($k > 1$) oszlopban szereplő értékek milyen módon függnek az első oszlop számaitól, vagyis

$$p_{ik} = \alpha_k p_{i1} + \beta_k \ (i = 1, 2, \ldots, N, k = 2, 3, \ldots, L),$$

akkor kiszámítható az összes többi valószínűség is. Persze az α_k, β_k értékeket nem ismerjük. Egyértelművé tehetjük azonban őket, ha megadjuk még két sorban, akár az első két sorban a valószínűségeket. Megadjuk tehát a

$$p_{11}, p_{12}, \ldots, p_{1L},$$

$$p_{21}, p_{22}, \ldots, p_{2L},$$

$$p_{31}, p_{41}, \ldots, p_{N1}$$

értékeket (első két sor és az első oszlop), azért, mert ezekből már az összes többi valószínűség kiszámolható (ezt szemlélteti a 3. ábra).

$$\begin{array}{ccccccccc}
p_{11} & p_{12} & p_{13} & p_{14} & p_{15} & p_{16} & p_{17} & p_{18} & p_{19} \\
p_{21} & p_{22} & p_{23} & p_{24} & p_{25} & p_{26} & p_{27} & p_{28} & p_{29} \\
p_{31} & p_{32} & p_{33} & p_{34} & p_{35} & p_{36} & p_{37} & p_{38} & p_{39} \\
p_{41} & p_{42} & p_{43} & p_{44} & p_{45} & p_{46} & p_{47} & p_{48} & p_{49} \\
p_{51} & p_{52} & p_{53} & p_{54} & p_{55} & p_{56} & p_{57} & p_{58} & p_{59} \\
p_{61} & p_{62} & p_{63} & p_{64} & p_{65} & p_{66} & p_{67} & p_{68} & p_{69} \\
p_{71} & p_{72} & p_{73} & p_{74} & p_{75} & p_{76} & p_{77} & p_{78} & p_{79}
\end{array}$$

3. ábra: Az első két sor és az első oszlop adatai megadásának szemléltetése

Egyszerű egyenletrendszer megoldásokról van szó ugyanis. Ha például szeretném tudni, hogy az 5. oszlop valószínűségei milyen α_5, β_5 értékek segítségével számolhatók ki, akkor csak fel kell írnom a már ismert értékekkel a következő egyenleteket:

$$p_{15} = \alpha_5 p_{11} + \beta_5$$

$$p_{25} = \alpha_5 p_{21} + \beta_5,$$

amelyekben α_5 és β_5 az ismeretlenek. Az eredményt könnyű kiszámítani, de nem annyira érdekes, hogy most leírjam[26]. A lényeg, hogy ha megadunk az NL darab valószínűségből $2L + N - 2$ –t (ennyi adat van az első két sorban és az első oszlopban), akkor a többi már meghatározható. Talán ez a tény is jó annak szemléltetésére, hogy mennyire speciálisnak kell lennie a valószínűségek mátrixának, hogy valóban intervallumskálák alakuljanak ki.

Miért baj az, hogy a valószínűségek mátrixának ilyen sajátosnak kell lennie? Hátha arról van szó, hogy az emberi képességek éppen úgy működnek, hogy a leírt feltétel mindig teljesül? Ha a klasszikus tesztelmélet alkalmazható egy képességre, akkor a képességhez tartozó bármely két feladat esetén a populációhoz tartozó vizsgált személyek jó feladatmegoldásainak valószínűségei, mint két adatsor, pozitív lineáris kapcsolatban kellene, hogy álljanak egymással. Ám ha az a helyzet, hogy a populáció egy részhalmazához tartozók az egyik feladat esetén felkészültebbek, míg a másik feladat esetén ez éppen a többiekkel van így, akkor a két adatsor között nem létezhet pozitív lineáris kapcsolat. Vagyis minden olyan képességgel baj van, amelyben különböző jellegű, különböző felkészültséggel jól megoldható feladatok vannak együtt. Márpedig a komplex képességek ilyenek.

Vagyis „a majom most ugrott a vízbe". Ha mégis igaz lenne, hogy egy képességhez van a vizsgált személyek populációján értelmezett különbségi jellegű empirikus relációs rendszer, és minden személyhez egy adott teszt esetén homomorf módon a pontszámnak az abba a tesztbe bekerült feladatok jó megoldása valószínűségeiből számolható várható értéke lenne hozzárendelve, akkor bármely két feladathoz tartozó valószínűség adatoknak pozitív lineáris kapcsolatban kellene állniuk egymással. Mivel szinte triviális, hogy a komplexebb képességek esetén ez nem teljesülhet bármelyik két feladat esetén, ezért ki kell mondanunk, hogy *a pedagógia számára akárcsak kicsit is fontos képességek vizsgálata során a klasszikus tesztelmélet alapján intervallumskálán nem lehet mérni*.

Lehetséges-e, hogy ordinális a skála? Ha jogos az előbbi helyzetleírás, vagyis az a lehetőség a kicsit is összetettebb képességek esetén, hogy tudniillik a jó megoldások valószínűségeiből alkotott adatsorok milyen eltéréseket produkálnak, akkor sorrendet sem lehet felállítani. Az egyik teszt esetén A személy képességfejlettsége (a jó megoldások számának várható értéke) nagyobb, mint a B személyé, ám egy másik teszt pontosan fordított relációt mutatna. A sorba rendezés sem egyértelmű, ordinális skála sem lehetséges.

Talán vannak olyan nagyon egyszerű képességek, amelyek esetén mégiscsak érvényesülnek a leírt feltételek, de egyrészt ezek a nevelési folyamatok egészének vizsgálata szempontjából egy nagyon szűk területet jelentenének, másrészt magam nem nagyon hiszek benne (mutasson valaki egyet!), harmadrészt a linearitás, vagy más szóval a monotonitás fennállása még mindig nem jelentené azt is, hogy valóban létezik megfelelő ERR.

A klasszikus tesztelmélet szerinti „mérések" nem mérések, ha csak a reprezentációs méréselméletnek megfelelő eljárásokat tekintjük méréseknek. Számok persze minden ilyen esetben adódnak, és az érzéketlen matematika tud azokkal mindenféle csuda műveleteket végezni. A pszichológiának és a pedagógiának számtalan tudományosan hitelesnek tartott megállapítása, tétele alapszik ilyen „méréseken". Éppen ezért óriási jelentősége volt annak, hogy a szakma túllépett a klasszikus tesztelméleten, létrejött a modern tesztelmélet, az „itemválasz elmélet" (IRT) számtalan modellje. Az új

---

[26] Az egyenletrendszernek nincs mindig megoldása. Akkor nincs, ha $p_{11} = p_{21}$, viszont $p_{15} \neq p_{25}$. Ekkor azonban már a megadott adatokra vonatkozik, hogy az első és az ötödik oszlop adatai nem pozitívan lineárisan összefüggők, vagyis eleve nincs értelme megoldásokat keresni, a klasszikus tesztelmélettel a helyzet nem kezelhető.

tesztelméletről kiderült, és ezt magam is bemutatom majd, hogy matematikai méréselméleti szempontból korrekt. Itt is lesz azonban korlátja az alkalmazhatóságnak. És itt lesz a baj, „itt ugrik majd a majom másodszor a vízbe”.

Konkrét példák problematikus kutatásokra

Hadd adjak konkrét példákat is ténylegesen megvalósult, publikált mérésekkel kapcsolatban arra, hogy a feladathalmazok összetétele kizárja, hogy a klasszikus tesztelméletnek megfelelő „mérések” méréselméleti értelemben is korrektek legyenek. Hiszen azt is a szememre lehetne vetni, hogy életidegen, hajuknál előrángatott, szélsőséges, nem létező helyzetekkel akarom megkérdőjelezni a klasszikus tesztelmélet szerinti „méréseket”. Találunk még a hazai kutatások közt is olyanokat, nem is kis számban, amelyekben ez a probléma nyilvánvalóan felmerül. Csak három példát említek: a *mértékváltások* készségének, az *arányosságszámítási* készségnek és az *induktív gondolkodási* képességnek a vizsgálatát. A mértékváltás esetében (Nagy J. 2000a, Józsa 2007) lehet szó hosszúságmértékek, területmértékek, űrmértékek, tömegmértékek, és még számtalan más típusú mérték átváltásáról. Vizsgált személyeink között nagyon sokan lehetnek olyanok, ha nem minden felmért tanuló ilyen, akik a különböző típusú átváltásokat eltérő valószínűségekkel tudják megoldani. Lehetnek olyanok, akik a hosszúságmértékek átváltása feladatok esetén teljesítenek jól, lehetnek olyanok, akik számára az űrmértékek átváltása „fekszik”, stb. Nyilvánvaló, hogy az eltérő típusokba sorolt feladatok eltérő nehézségűek lehetnek a vizsgált személy számára. Pontosan az a helyzet állhat elő (nagy valószínűséggel), amit fentebb jeleztem.

Az *arányosságszámítási készség* (Varga, Józsa és Pap-Szigeti 2007) hasonló. E készség komponensei az arányos osztás, az egyenes és fordított arányossággal kapcsolatos számítások, a mértékváltás (mint láttuk, önmagában is több dimenziót tartalmaz) és a százalékszámítás. Ismét jól látható, hogy az e típusokba tartozó feladatok eltérő valószínűségekkel oldhatók meg valószínűleg nagyon sok vizsgált személy, ha nem mindenki számára. A vizsgált populáció egyes személyei inkább az egyik típusú feladatot tudják megoldani, míg kiválaszthatók olyanok, akik ezekben sikertelenek, és ez a két csoport éppen ellentétesen teljesíthet feladatok egy másik csoportjával kapcsolatban.

Az *induktív gondolkodási képesség* fejlettségének mérésére *Csapó Benő* határozottan kiválaszt három feladattípust (Csapó 2001): a szóanalógiákat, a számanalógiákat és a számsorozatok folytatását. Talán már érvelnem sem kell: nehéz elképzelni, hogy azok, akik az egyik feladattípusban jól teljesítenek, minden feladattípus esetén hasonlóan produkálnak, sőt, hogy itt határozottan pozitív lineáris összefüggések legyenek.

Miért „működnek” mégis a mérések?

Rá kell mutatnom valamire, ami *a konkrét mérések „csapdája”*. Az itt is említett kutatók, és nyilván nagyon sokan mások, akik hasonló méréseket végeznek, joggal szegezhetnék nekem a kérdést, hogy ha ez a helyzet, miért nem látszik ez a konkrét mérések során, miért jeleznek a *reliabilitás* becslések nagyon jó tulajdonságokat a mérések számára. És valóban, az említett publikációkban is szerepeltek azok a reliabilitásértékek (általában a 0,75-0,95 tartományban), amelyek a tesztek klasszikus tesztelméletben értelmezett kiváló megbízhatóságát jelzik. Bár eredetileg szerettem volna elkerülni, ám nem tudom megtenni, ezért e helyen foglalkoznom kell a reliabilitással.

A reliabilitás definíció szerint *a valódi értékek varianciájának és a mért értékek varianciájának hányadosa* (Cronbach 1947). Annak „mérésére” szolgál, hogy a hiba mennyire kicsi. A klasszikus tesztelmélet axiómarendszere szerint a hiba (E) és a valódi érték (T) független mennyiségek, ezért (mivel összegük a mért érték: X) a mért érték varianciája egyenlő a másik két valószínűségi változó varianciájának összegével:

$$D^2(X) = D^2(T) + D^2(E).$$

A reliabilitás[27]:

$$\rho = \frac{D^2(T)}{D^2(X)}$$

Tekintve, hogy a hibát leíró valószínűségi változó várható értéke 0, ezért a hiba szórása (és annak négyzete is) valójában arra jellemző mennyiség, hogy átlagosan milyen nagyok a hibák abszolút értékei, mekkorák a hibák. Ha a hiba várható értéke 0, és konkrét értékei kicsik, akkor szóródása is kicsi, a varianciája is kicsi, természetesen a T varianciájához képest. Ekkor $D^2(X)$ és $D^2(T)$ egymáshoz közeli értékek, hányadosuk 1-nél kisebb, de 1-hez közeli érték. Ha nagyok a hibák, akkor $D^2(X)$ és $D^2(T)$ jelentősebben különböznek egymástól, jóval kisebb a hányados. Az a „jó" mérés, amelyben a hiba kicsi, vagyis a reliabilitás, a mondott hányados nagy, 1-hez közeli érték.

A reliabilitás tehát a mérés hibájára jellemző mennyiség, ha kicsi, az azt jelzi, hogy a mérés során nagy hibák adódhatnak, s ha nagy – ennek örül a kutató –, akkor kicsik a hibák. A reliabilitás azonban egy elméleti érték, egy olyan hányados, aminek a számlálójában szereplő mennyiség, a látens képességfejlettség varianciája nem határozható meg. Éppen ezért eljárásokat kellett kidolgozni a becslésére. Sokféle ilyen becslés ismeretes (ld. pl. Csapó 2000), mindegyik az előforduló kutatási helyzetek közül egyesekben „működik jól".

A reliabilitást becslő mértékekkel kapcsolatban számtalan súlyos probléma merül fel. Bizonyos, és számunkra fontos szempontok szerint egyáltalán nem olyan jó jelzői a mérések megbízhatóságának, mint azt sokan gondolják (Dunn, Baguley és Brunsden 2014, Sijtsma 2009a, Green és Yang 2009, Shevlin és mts. 2000, Cortina 1993; Green, Lissitz és Mulaik 1977). Különösen érdekes az az eredmény, amely szerint a reliabilitás legfontosabb, legtöbbet használt becslése, a *Cronbach alfa* lehet azonos érték olyan mérések során is, amelyekben különböző számú faktorok (dimenziók) játszanak szerepet. Nekünk éppen ez a problémánk. A különböző faktorok jelenléte azonos azzal a problémával, hogy eltérő jellegű feladatok, feladatcsoportok tartoznak a képességhez, és ez a sajátosság meggátolja, hogy a feladatok teljes halmazán érvényesüljön a már sokszor leírt linearitás, másképpen: meggátolja, hogy a szóba jövő tesztek azonos eredetűek legyenek. Használjuk a továbbiakban az egydimenziós megnevezést azokra a tesztekre, amelyekkel kapcsolatban a teljesítményeket egy faktor határozza meg, és legyen egy teszt többdimenziós, ha legalább két faktor, legalább két dimenzió megjelenik. Úgyanígy a mért

[27] A reliabilitás definíciójában a varianciák hányadosa azt fejezi ki, hogy a T-k varianciája „mekkora részt" határoz meg az X-ek varianciájából (ha 100-zal megszorozzuk, akkor százalékosan fejezi ki, hogy mekkora ez az arány). Elemi valószínűségszámítási összefüggés, hogy ez a hányados megegyezik a két változó lineáris korrelációs együtthatójának négyzetével (R_{TX}^2). Ez egyébként a reliabilitás eredeti definíciója. Be lehet bizonyítani, hogy ha van két párhuzamos tesztünk, akkor a reliabilitás az azokkal mérhető teszteredmények korrelációs együtthatójával is megegyezik. Párhuzamos két teszt, ha a populációba tartozók esetén ugyanakkora a valódi érték (T), és ugyanakkora a hibák varianciája. Felmerült, hogy ez az összefüggés lehetőséget biztosít az reliabilitás meghatározására, hiszen elég csak párhuzamos teszteket készíteni, azokkal mérni, és kiszámítani a két adatsor korrelációs együtthatóját. Ez az eljárás azonban nem működik, mert párhuzamos teszteket gyakorlatilag lehetetlen előállítani, és még ha lennének is jelöltek, szinte lehetetlen lenne bebizonyítani, hogy azok valóban párhuzamos tesztek.

képességet is nevezhetjük egydimenziósnak, amennyiben minden tesztje egydimenziós, és többdimenziósnak egyébként. Ugyanazon – egydimenziós – képesség mérésére szolgáló tesztek mind azonos eredetűek, ha még a hibák függetlensége is teljesül.

Klaas Sijtsma olyan adatrendszereket konstruált, amelyek rendre 1, 2, 3 faktor jelenlétét mutatták, azonban a tesztek reliabilitásmutatói azonosak voltak (Sijtsma 2009a). *Michael Shevlin* és munkatársai kimutatták, hogy bizonyos, meglehetősen általánosan előforduló feltételek között minél több dimenzió jellemez egy mérést, annál nagyobb lesz a *Cronbach* alfa, ezért annak nagyobb értéke éppen nem az egydimenziósság bizonyítéka (de természetesen egydimenziós vizsgálatban is lehet magas a mutató értéke) (Shevlin és mts. 2000). Vagyis a közfelfogással ellentétben kimondhatjuk, hogy a magas reliabilitásmutató nem véd meg bennünket attól, hogy mérésünkben akár több, egymástól eltérő jellemző is megbújjék, meghatározza az eredményeket (*Sijtsma* kimutatja, hogy a faktorok lehetnek még akár korrelálatlanok is).

Valójában itt egy régről ismert problémáról van szó. A reliabilitás becslése (a *Cronbach* alfa) – egyik értelmezésében – a teszt belső *konzisztenciájáról* ad felvilágosítást, arról, hogy az itemek mennyire kötődnek egymáshoz, mennyire „viselkednek” azonos módon. E tulajdonságtól azonban meg kell különböztetni a *homogenitást*, másképpen egydimenziósságot, amelynek szükséges, de nem elégséges feltétele a belső konzisztencia (Panayides 2013; Green, Lissitz és Mulaik 1977). A *Cronbach* alfa (az említett hazai kutatásokban is használt reliabilitásmutató) a belső konzisztencia mértéke, a mérésben szereplő több dimenzió ellenére is lehet magas az értéke. Ahogy *Cortina* fogalmaz: „Itemek egy halmaza lehet összefüggő [interrelated] és többdimenziós” (Cortina 1993, 100. o.). De józan ésszel is beláthatjuk, miért nem „ugrik ki” a probléma egy konkrét tesztelés során. *Csapó Benő* például az induktív gondolkodás fejlettségének mérése során egy 58 itemből álló tesztet használ, rendkívül magas reliabilitásmutatóval (0,93) (Csapó 2001). A három feladattípus (szóanalógia, számanalógia, számsorozatfolytatás) képviselői egyaránt előfordulnak a tesztben. Egy konkrét vizsgált személy esetében a megoldandó feladatok valószínűségei alkothattak egy meglehetősen „kevert”, sokféle értéket felvonultató rendszert. Nem tudjuk, hogy az elkészített tesztben szereplő, az egyes típusokhoz tartozó feladatok nehézségei milyenek voltak az egyes felmért személyek számára. Nagy a valószínűsége azonban annak, hogy a kiválasztás eredményeként olyan feladatsor jött létre, amelyben mindhárom feladattípusból nagyon különböző nehézségű feladatok szerepeltek, és ezért a dimenzionalitás mintegy eltűnt a tesztben, Viszont a reliabilitásmutató magas értékű volt, nem utolsó sorban a gondos tesztfejlesztés eredményeként, amikor is a „deviáns” itemek már nem kerültek be a végső tesztbe. A képességfejlettséget, mint látens, elméleti értéket azonban a definíció szintjén a vizsgált képesség összes lehetséges feladatával kell definiálnunk, és ez esetben már nem lehetségesek a sajátos választások, „kiütközik a dimenzionalitás”.

És még egy megjegyzés: *Cortina* felhívja a figyelmünket arra, hogy a sok itemből álló tesztek (ő a legalább 40 itemes teszteket említi) reliabilitása – hacsak nem 0 közeliek vagy akár negatívak az itemkorrelációk – szükségszerűen magas, vagyis ezekben az esetekben a reliabilitás jól jelzi ugyan a mérési hiba kicsi voltát, de semmit nem mond a homogenitásról (Cortina 1993, 101. o.).

Mindez azt jelenti, hogy erősen kétséges, hogy a reliabilitás mutató kiszámítása segít a több dimenzió jelenlétének felismerésében. Érdekes, hogy valójában a reliabilitás becslésére meghatározott Cronbach-alfa még a mérési hiba megítélésével kapcsolatban sem nyújt túlságosan kielégítő információt (ha már belekezdtem a reliabilitás elemzésébe, érdemes megmutatnom ezt az összefüggést is). A szakirodalomban is találunk bőséggel a Cronbach-alfa kritikájára vonatkozó elemzéseket (a már említetteken túl: Sijtsma 2009b;

Thorsen és Bjorner 2000). Elméleti vizsgálat helyett itt egy példával igyekszem illusztrálni ezt a problémát.

Két teszt mért értékeinek korrelációs együtthatója érdekes összefüggést mutat a tesztekkel kapcsolatos reliabilitás értékekkel. Ez az összefüggés így írható fel:

$$R(X_1, X_2) = \sqrt{\frac{D^2(T_1)}{D^2(X_1)}} \cdot \sqrt{\frac{D^2(T_2)}{D^2(X_2)}}.$$

Itt $R(X_1, X_2)$ jelöli a két tesztben született teszteredmények (elméleti) korrelációs együtthatóját, nyilván X_1 az egyik, X_2 a másik teszthez tartozó, mért értékek valószínűségi változói, T_1 az egyik, T_2 a másik teszthez tartozó valódi, látens pontértékek változói, és a D^2 a megfelelő változó szórásnégyzetére, varianciájára utal. Mint látjuk, a négyzetgyökjel alatt éppen a reliabilitások állnak, így

$$R(X_1, X_2) = \sqrt{\rho_1} \cdot \sqrt{\rho_2} = \sqrt{\rho_1 \rho_2},$$

ahol ρ-val (megfelelő indexszel ellátva) a reliabilitást jelöltem. A baloldalon álló érték az, ami konkrétan meghatározható, amennyiben sikerült két, a képességhez tartozó teszttel mérni. Az elméleti reliabilitásokat a mért adatokból nem tudjuk meghatározni, csak becslésekkel rendelkezünk rájuk vonatkozóan, ezek lehetnek a Cronbach-alfák, amelyek mindig kisebbek, mint a tényleges, a nem meghatározható értékek. Két tesztet felvenni valójában nem is nehéz feladat. Egy tesztet kell készíteni, majd annak itemjeit véletlenszerűen két csoportra kell osztani, s utána már a vizsgálat két különálló teszttel folytatható.

A bemutatandó példa a hazai Országos kompetenciamérés adataira támaszkodik[28]. A 2011. évi mérés adatait használom, a matematika tesztben a 10. évfolyamosok által nyújtott teljesítményt. Az egyes itemekre kapott pontszámokat ismerjük, s ez alapján összeállíthatunk két különálló tesztet, véletlenül választva az összesen 83 itemből 41 és 42 elemet. A két teszt nyers pontszámainak átlaga 115,8 és 103,7 lettek, a szórások 39,5 és 34,4. A két teszteredmény változó korrelációs együtthatója az évfolyam több mint 90 000 tanulójának adatai alapján 0,901. Mivel a két teszt valódi (látens, pontosan nem megmérhető) reliabilitásai nagyjából megegyezhetnek, vagyis $\rho_1 \approx \rho_2$, ezért a fenti képletben a négyzetgyök alatt két egymáshoz közeli szám szorzata szerepel, így a négyzetgyök értéke a két szám nagyjából közös értéke. Nem követünk el nagy hibát, ha a két reliabilitást egyenlőnek vesszük, és így az egyenlőség jobboldalán ez a reliabilitásokat közelítő érték szerepel. Ez azt jelenti, hogy a leírt mérés során a nagyjából az itemek feléből alkotott tesztek reliabilitása 0,9–hez közeli érték lehet. A két tesztre azonban meghatározható a Cronbach-alfák értéke is. Ezek az értékek: 0,831 és 0,835. Azt látjuk tehát, hogy a Cronbach-alfa meglehetősen távol áll a reliabilitás egy józannak tekinthető becslésétől. Persze ez a gondolatmenet erősen felhasználta, hogy a két „félteszt" saját reliabilitásai csak kissé különböznek egymástól. Ha ez nem így van, és valaki azt feltételezi, hogy legalább az egyik reliabilitáshoz közeli érték a Cronbach-alfa, akkor bizonyos értelemben még rosszabb a helyzet. Ha az egyik reliabilitás 0,84 lenne, akkor a képlet alapján a másik félteszt reliabilitásának kb. 0,97-nak kellene lenni. Igen kicsinek értékelhetjük egy ilyen nagy eltérés (a két „félteszt" reliabilitása 0,84 és 0,97)

[28] Az adatokhoz az Oktatási Hivatal jóvoltából jutottam hozzá, köszönet érte.

kialakulásának valószínűségét, amikor a két „féltesztbe” a feladatokat véletlenszerűen válogattuk össze.

Mi a baj a klasszikus tesztelmélettel?

Mi is volt a fenti érvelés logikája? Kiindultunk egy súlyos hiányosságból: a klasszikus tesztelméletnek – ismereteink szerint – nincs reprezentációs méréselméleti legitimitása. Nem állított fel senki olyan elméletet, amely a képességfejlettségek (másképpen az emberek) halmazán értelmezett volna legalább különbségi relációs struktúrát. Nem ismerünk definíciót olyan négyváltozós *D* relációra az emberek halmazán egy adott képességükkel kapcsolatban, amely teljesítené a reprezentációs méréselméletben leírt követelményeket (megfelelne az axiómáknak), és lehetővé tenné, hogy számokat úgy rendeljünk hozzá az emberekhez, hogy ez homomorf leképezés legyen, és így legalább intervallumskála jöjjön létre. De még csak rendezési reláció sincs leírva sehol a szakirodalomban, hogy rendezési relációs struktúra jöjjön létre, és legalább ordinális skáláról beszélhessünk.

Ám azt mondtuk, hogy a tényleges hiány, hogy ilyet még senki nem írt le, nem jelenti azt, hogy ilyen nem is létezhet. Ha létezne, akkor igaznak kellene lenni annak, hogy a klasszikus tesztelméletben megadott szám-hozzárendelések esetén kialakuló adatsorok egymással (bármelyik leképezés-párt nézzük is) pozitív lineáris kapcsolatban vannak (illetve rendezési struktúra esetén egyértelműen kialakul egy rendezés). És elkezdtük ellenőrizni, hogy ez vajon így van-e. Rájöttünk, hogy ha ez a követelmény maximálisan érvényesülne, akkor az „egy-feladatos tesztekkel” is működne, ám világosan lehet látni, hogy a pedagógiát érdeklő komplex képességek esetén ez lehetetlenség. Vagyis nem csak arról van szó, hogy valamit még nem fedeztünk fel, hogy van egy rejtett különbségi relációs struktúra mégis a képességfejlettségek halmazán, amelyet majd valaki valamikor megtalál, hanem azt sikerült láttatni, hogy ilyen egyáltalán nem létezhet. Legalábbis a kicsit is komplex képességek esetén.

Ezzel befejeztük a klasszikus tesztelmélet vizsgálatát és bírálatát. Ennek az egész paradigmának el kellene tűnnie a tudomány „süllyesztőjében”. Egyelőre azonban itt van a körünkben, és csak nagyon lassan veszi át a helyét a modern tesztelmélet. (Kijelentésem inkább a hazai kutatásokra vonatkozik, a világban az IRT modellek – erről már szóltam korábban is – sokkal fontosabb szerepet játszanak.)

Modern tesztelmélet

Bevezető gondolatok

A modern tesztelmélet megalkotói arra törekedtek, hogy kiküszöböljék a klasszikus tesztelmélet problémáit (Guttman 1950, Rasch 1960/1980, Lord és Novick 1968, Lord 1980, Embretson és Reise 2000). És valóban, míg a klasszikus tesztelmélet reprezentációs méréselméletnek való megfelelését soha senki nem próbálta bizonyítani, addig a modern tesztelmélet esetén ez a megfelelés jól alátámasztott állítás. Előre bocsátom az itt következő vizsgálat végeredményét: miközben teoretikusan sokkal biztonságosabb kijelentéseket tehetünk a modern tesztelmélet korrekt jellegéről, szemben a klasszikus megfontolásokkal, addig a szintén az elméletből következő, de józan ésszel is belátható követelmények oly jelentősek, hogy a pedagógia számára érdekes esetekben *az elmélet alkalmazhatósága erősen megkérdőjelezhető*.

Annak bizonyítása, hogy a modern tesztelmélet alkalmazásával matematikai szempontból korrekt konstrukciót alkothatunk, és legalább intervallumskálákhoz jutunk, kétféleképpen is történhet. Mindkét érvelést megmutatom. Az elsőt azért, mert ennek keretében mutatom be a modern tesztelmélet egy meghatározó részterületét (az egész leírására sajnos itt nincs elég hely), illetve azért is, mert ez a bemutatás azt láttatja velünk, hogy *a modern tesztelmélet önmagában is alkalmas a megfelelő koncipiálásra*. A második érvelés azért lesz fontos, mert az *együttes mérés* bemutatását teszi lehetővé, s e könyvvel az egyik célom, hogy ezt a modern gondolatot a hazai érdeklődő olvasóközönség elé tárjam.

Az itt következő részek – miközben továbbra sem használok középiskolai tananyagon túlmenő ismereteket – matematikai szempontból nehezebben olvashatók lesznek. Igyekszem közérthetően fogalmazni, és valamennyire függetleníteni a kvalitatív leírásokat a precízebb matematikai érveléstől, hogy utóbbiak megértésének hiányában is viszonylag jól érthető legyen a lényeg. Ha nem is kimerítően, de az alapokat tekintve némileg teljességre törekvő módon mutatom be a legegyszerűbb modern tesztelméleti modellt, a Rasch-modellt, elhagyva, vagy különválasztott, „kisbetűs" részekben elhelyezve a matematikailag mélyebb ismereteket, valamint az esetek nagyobb részében nem közölve a bizonyításokat. A leírás már csak azért sem lehet kimerítő, hiszen e területek átfogó prezentációi vaskos könyveket töltenek meg (Molnár 2013; Alagumalai, Curtis és Hungi 2005; Embretson, és Reise, 2000; Hambleton, Swaminathan és Rogers 1991; Lord 1980; Wright és Stone 1979), illetve itt csak a modern tesztelmélet legegyszerűbb modelljével foglalkozom.

A modern tesztelmélet is (többek között) arra való, hogy *teoretikus keretet* biztosítson az emberi teljesítmények mögött meghúzódó *képességek* fejlettségének mérésére. Kiindulópontját a következőképpen jellemezhetjük. Vegyünk egy vizsgált személyt, akivel megoldatunk egy feladatot. Ennek a „kísérletnek" kétféle kimenetele lehetséges (legalábbis én itt korlátozom a bemutatást erre az esetre): a vizsgált személy vagy megoldja helyesen a feladatot, vagy nem. E két elemi esemény valószínűségeinek az összege 1. A modern tesztelmélet megalkotói bizonyos feltételezésekkel éltek azzal kapcsolatban, hogy milyen értékek ezek a valószínűségek. Többféle ilyen feltételezés, „hivatalosabb nevén" modell született. A leghíresebb, és a leginkább alkalmazott modell *Georg Rasch* (1960) modellje, amelyben a helyes feladatmegoldás valószínűségét a vizsgált személy *képességfejlettségének* objektív mértékétől, valamint az adott feladat objektív *nehézségétől* tette függővé. Hogy pontosabb legyek: *Rasch* a két érték *különbségének* függvényében írta fel a vizsgált valószínűséget.

A modern tesztelmélet alapgondolatai

Nézzük a részleteket! Egy adott képesség esetében a képességhez tartozó feladatokat a vizsgált személyek közül többen vagy kevesebben tudják megoldani. Amit kevesebben oldanak meg, az nehezebb, és fordítva. Egy adott személy többet vagy kevesebbet tud megoldani a feladatokból, aki többet, arról azt képzeljük, nagyobb a képességfejlettsége (és ez természetesen a fent már kifejtettek értelmében egyáltalán nem precíz matematikai kijelentés). Az a kérdés, hogy e két mennyiségre vonatkozóan, vagyis a *feladatnehézségre* és a *képességfejlettségre* tudunk-e létrehozni intervallumskálákat (majd látjuk, hogy e mennyiségek egyetlen skálán is elhelyezhetők).

A klasszikus tesztelméleti tárgyaláshoz hasonlóan nem azt a sorrendet választom a kifejtésre, ami logikus lenne. Logikus az lenne, hogy előbb megmutatom, hogy a vizsgált személyek (vagy másképpen képességfejlettségeik) halmazán értelmezhető egy olyan *D* négyváltozós reláció, amely kielégíti a reprezentációs méréselmélet által megkövetelt axiómákat, vagyis felépítem az empirikus relációs rendszert, majd bemutatom, hogyan lehet számokat megfeleltetni a képességfejlettségeknek, bizonyítva, hogy ez a leképezés homomorfizmus. Ezen az úton végigmenve biztosíthatnám, hogy intervallumskála jön létre. Ezután következhetne a diszkusszó, az így felállított teljes elméleti rendszer gyakorlatban való alkalmazhatóságának vizsgálata. Didaktikai okokból nem ezt a sorrendet választom. Először megmutatom, hogy miképpen feleltetjük meg a Rasch-modell szerint a számokat a vizsgált személyeknek, és amikor már érthetők lesznek az itt szóba került fogalmak, visszatérek az intervallumskála létezésének bizonyítására.

A képességmérés modern elméletében háromféle mennyiség játszik alapvető szerepet: a már említett *feladatnehézség* és *képességfejlettség*, továbbá *a feladatok helyes megoldásának valószínűsége*.

A további elemzést arra alapozom, hogy egy esemény valószínűsége jól modellezhető a valószínűségszámítás *Kolmogorov*-féle elmélete segítségével, vagyis feltételezzük, hogy a leírt eseményeknek valóban van valószínűségük (a valószínűség fogalma értelmezésének problémáját nem szeretném felvetni, maradjunk az objektivista megközelítésnél). A képességfejlettségek és a feladatnehézségek kérdésében már nem lehetünk ilyen „megengedők”. Ezen értékek létezése nem triviális, nem valamifajta objektív tulajdonságokról van szó. Az elmélet kifejtése során természetesen valamilyen definíciókra van szükségünk. Ezeket a definíciókat azonnal megadom, csak előbb vezessünk be fontos jelöléseket.

Minden egyes feladatnak van egy „saját”, a nehézségét jellemző paramétere, a jele a görög delta (δ), és mivel mindig sok feladattal van dolgunk, ezért általában indexe is van, erre a leggyakrabban a görög lambdát (λ) használom itt. Így *a λ-adik feladat nehézségét δ_λ jelöli*. A képességfejlettségek természetesen az emberekhez, a vizsgált személyekhez vannak hozzárendelve. Jelölésükre a görög bétát (β) használom (a szakirodalomban ez az egyik szokásos jelölés), szintén kap indexet, erre a görög nü-t (ν) rendszeresítem, így a ν-edik személy képességfejlettségét β_ν jelöli. A *Rasch-modell* szerinti számításokban a feladatnehézségek és a képességfejlettség mértékek nagyjából a –5-től +5-ig terjedő intervallumban helyezkednek el (az esetek többségében, miközben egyébként akármekkora értékek lehetnek). A harmadik mennyiség, a jó megoldás valószínűsége egy konkrét személy konkrét feladattal kapcsolatos sikeréhez van hozzárendelve (egy esemény valószínűségéről van szó, az esemény az, hogy a konkrét személy a konkrét feladatot jól oldotta meg). Ha a λ-adik feladatról és a ν-edik személyről van szó, akkor e valószínűséget ezután $p_{\nu\lambda}$ jelöli.

De mik vajon a *képességfejlettség* és a *feladatnehézség* definíciói? Mit értsünk e fogalmak alatt? Használjuk fel intuitív értelmezéseinket! Minél nagyobb valakinek a képességfejlettsége, illetve minél könnyebb feladatról van szó, annál nagyobb a valószínűsége a feladat jó megoldásának. Fordítva: minél kisebb a képességfejlettsége a vizsgált személynek, illetve minél nehezebb az a feladat, amellyel foglalkozik, annál kisebb a jó megoldás valószínűsége. Foglalkozzunk csak azokkal az esetekkel, amelyekben a helyes feladatmegoldás valószínűsége csak e két paramétertől függ, és csak azokkal az esetekkel, amelyekben megadható a vizsgálati populációban szereplő minden egyes ember képességfejlettségét kifejező valós szám, valamint a vizsgálatban szereplő képességhez tartozó összes feladathoz olyan valós szám, hogy ezek a számok minden személy-feladat párosítás esetén kielégítsék a *Rasch-modell* képletét. *Georg Rasch* dán matematikus volt az, aki 1960-ban megjelent, alapvető jelentőségű művében javaslatot tett olyan összefüggésre, amely megadja, hogy a képességfejlettségnek és a feladatnehézségnek milyen függvénye legyen a jó megoldás valószínűsége. A képlet a következő (miközben nem árt tudnunk, hogy amit ide leírok, és amit a szakirodalomban is a legtöbbször a *Rasch*-féle összefüggésnek neveznek, már egy matematikai átalakítás eredménye):

$$p_{\nu\lambda} = \frac{e^{\beta_\nu - \delta_\lambda}}{1 + e^{\beta_\nu - \delta_\lambda}}.$$

A $p_{\nu\lambda}$ az említett valószínűség, e a természetes alapú logaritmus alapja, egy szám, végtelen tizedes tört, az értéke három tizedesre kerekítve: 2,718. A képletben láthatjuk a vizsgált személy képességfejlettsége és a kiválasztott feladat nehézsége közötti különbséget: $\beta_\nu - \delta_\lambda$. Ettől a különbségtől függ csak a feladat vizsgált személy általi helyes megoldásának valószínűsége.

Vagyis a következőképpen értelmezzük a képességfejlettséget és a feladatnehézséget: Olyan, valós számokkal jellemzett mennyiségek, amelyekkel a megadott módon kifejezhető az összes megoldási valószínűség. Bármelyik személyt vesszük, és bármelyik feladatot, és ezek paramétereit behelyettesítjük a képletbe, megkapjuk, hogy éppen az a személy éppen azt a feladatot mekkora valószínűséggel tudja megoldani.

Mindebből a valószínűség az, ami a képlettől, tehát a saját konstrukciónktól független mennyiség. Legalábbis úgy képzeljük, ahogy már írtam, hogy bármely feladat és bármely személy esetében értelmes kijelentés az, hogy ennyi és ennyi a jó megoldás valószínűsége. A képességfejlettségek és a feladatnehézségek konstruált mennyiségek. Ilyen mennyiségek nem feltétlenül léteznek, nem feltétlenül lenne sikeres vállalkozás ilyeneknek a konstrukciója. Csak olyan esetekben, amelyekben a jó megoldások valószínűségei ezt lehetővé teszik, ha tehát *„éppen olyanok" a valószínűségek*. A matematikus azt mondaná: vizsgáljuk azokat a jelenségeket, azokat a képességeket, amelyekben érvényesül a *Rasch-modell*. A geometria axiómarendszerének (és bármilyen más elméletnek) a megalkotása is történhet ezen a módon: tekintsük az egyáltalán előforduló esetek egy részhalmazát, azon eseteket, amelyekben egy adott szabály érvényes. Mi a képességeket vizsgáljuk. Egy képesség a hozzá tartozó feladatokkal specifikálható. Mi azonban most csak olyan képességekkel foglalkozunk, amelyek esetében a feladatok jó megoldásainak valószínűségeit a *Rasch-modell* írja le. Vagyis vannak olyan a képességfejlettségeket és a feladatnehézségeket kifejező valós számok, amelyeket minden egyes személy-feladat pár esetén behelyettesítve a képletbe, a kiszámolt valószínűség pontosan megegyezik az adott személy sikeres feladatmegoldásának valószínűségével.

Képzeletben végtelen sok ember van, és általában végtelen sok feladat. Ha egy valakinek a feladatmegoldásait nézzük, akkor hozzá tartozik végtelen sok valószínűség. Ezek a valószínűségek olyan kapcsolatban vannak a kiválasztott személy képességfejlettségével, hogy ha bármelyik feladatot nézzük, s annak nehézségparaméterét, akkor a felsorolt mennyiségek a képletben megadott viszonyban vannak egymással. Az, hogy a modell „működik", érvényes az adott jelenségre, azt jelenti, hogy bárki másnak vizsgálva a helyes feladatmegoldásait, az egyes valószínűségekre felírható képletben e másik személy képességfejlettségén kívül *ugyanazokat* a nehézségparamétereket kell használnom, mint amiket ugyanezeknél a feladatoknál, de az előbb vizsgált személynél használtam. És fordítva: ha kiszemelünk egy feladatot, és figyeljük az azt megoldó összes embert, akkor a kiszemelt feladat nehézségparaméterét, valamint sorban az egyes emberek képességfejlettségét kell behelyettesítenem a képletbe, és megkapom a jó megoldások valószínűségeit. S ha veszek egy másik feladatot, és annak nehézségparaméterét, akkor a képletbe *ugyanazokat* a képességfejlettségeket kell helyettesítenem, mint az előbb, amennyiben vizsgálni akarom az összes, e feladatot megoldó ember jó megoldásainak valószínűségét.

Ne felejtsük el, egyelőre tisztán az elmélet megalapozásáról van szó. Olyan ez, mint mikor a matematikus egy új geometriai axiómarendszert hoz létre. Alkot egy világot, ami önmagában ellentmondásoktól mentes (legalábbis a rendszer megalkotásakor ezt reméli, és ha lehet, később akár be is bizonyíthatja valaki), és majd később teszi fel a kérdést, hogy amit alkotott, az vonatkoztatható-e bármire, ami a tapasztalati világában létezik, lehet-e modellje tényleges jelenségeknek, folyamatoknak, viszonyoknak. De jól tudjuk, hogy az elméletalkotás a valóságban nem ennyire steril. Már az elmélet létrehozása során figyelembe vesszük a tapasztalatainkat (Euklidesznek eszébe nem jutott – valószínűleg – hogy megfogalmazzon egy ilyen axiómát: bármely három különböző pontra illeszthető egyenes). Milyen meggondolás van a *Rasch-modell* megfogalmazása mögött? Miért éppen ezt a képletet akarjuk használni, és miért nem gondolkodunk valami másban? A részbeni magyarázat a képlet némi elemzésével adható meg.

A Rasch képlet mondanivalója

Mit mond a *Rasch* képlet? Szeretném ezt megvilágítani matematikai kifejezések részletes értelmezésétől kissé vonakodó olvasóim számára is. Leírom ide még egyszer ezt a képletet:

$$p_{\nu\lambda} = \frac{e^{\beta_\nu - \delta_\lambda}}{1 + e^{\beta_\nu - \delta_\lambda}}$$

Először is fontos, hogy egy hatványkifejezés mindig pozitív szám, ezért e tört számlálója és nevezője is mindig pozitív, tehát a hányados is, vagyis a valószínűség is mindig pozitív, ennek nyilván így is kell lenni. Az viszont érdekes, hogy 0 nem lehet, mert ahhoz a számlálónak 0-nak kellene lenni, de egy nem 0 alapú hatvány értéke soha nem lehet 0, mindig csak annál nagyobb. Érdemes továbbá megfigyelni, hogy a számláló és a nevező hasonlítanak egymásra, ugyanaz a hatványkifejezés szerepel bennük, csak a nevezőben hozzá van még adva 1. Ebből következik, hogy nem csak az igaz, hogy a tört értéke mindig pozitív, az is igaz, hogy a nevező mindig nagyobb, mint a számláló. Akkor viszont a tört értéke mindig kisebb, mint 1. Egyenlő nem lehet 1-gyel, mert ahhoz a számlálónak és a nevezőnek egyenlőnek kell lenni, de a nevező mindig 1-gyel nagyobb, mint a számláló. Megnyugtató, hogy a tört értéke kisebb, mint 1, a képlet „pontosan tudja", hogy 0-nál kisebb és 1-nél nagyobb szám nem lehet az eredmény, ahogyan ennek lennie is kell, hiszen valószínűségről van szó.

A tört értéke a $\beta_\nu - \delta_\lambda$ különbségtől függ. Érdemes megvizsgálnunk, hogy alakul a tört értéke, ha ez nagy pozitív érték, ha nagy abszolút értékű negatív szám, és mi van akkor, ha 0, vagy közel van hozzá.

Nézzük a nagy pozitív értékeket! A $\beta_\nu - \delta_\lambda$ különbség úgy lehet nagy pozitív szám például, hogy a képességfejlettség (a β_ν) nagy, és mellette δ_λ kicsi pozitív szám, esetleg az utóbbi negatív. Ilyen esetben a két érték különbsége nagy pozitív szám lesz. Igen jó képességfejlettségű személy old meg egy közepes nehézségű, vagy igencsak könnyű feladatot (a növekvő abszolút értékű negatív számok egyre könnyebb feladatot jelentenek). Ekkor nyilván azt várjuk, hogy a valószínűség nagy legyen, valószínű, hogy az adott, nagy képességfejlettségű személy megoldja a könnyebb feladatot. Nézzük a képletet! A számlálóban az *e* egy nagy pozitív kitevőre van felemelve, ami azt jelenti, hogy a kapott szám igen nagy lesz, és a nevezőben e számnál eggyel nagyobb szerepel. Például, ha a $\beta_\nu - \delta_\lambda$ különbség értéke 8 (látjuk majd, hogy a Rasch-modell keretében ez előfordulhat), akkor az e^8 értéke 3000-nél nagyobb, míg egy ehhez közel álló, nála eggyel nagyobb szám lesz a nevező. Két egymáshoz közeli, egymástól csak 1-gyel különböző, a példában 3000-nél nagyobb számot kell elosztani egymással, az eredmény nyilván 1-hez nagyon közeli szám lesz. Vagyis a leírt esetben igen nagy valószínűséggel tudja megoldani a személy a feladatot, ahogyan azt vártuk is.

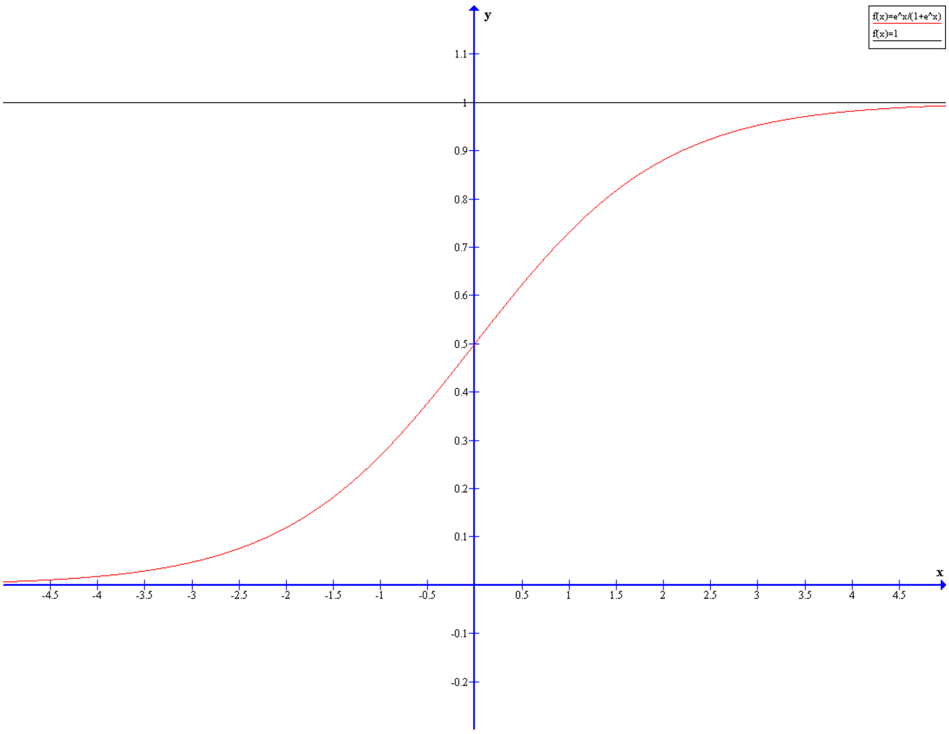
Ha most egy olyan esetet képzelünk el, amelyben β_ν nagy abszolút értékű negatív szám (pl. –4, ez is reális), továbbá δ_λ pozitív, akár relatíve nagy szám (mondjuk 4 is lehet), akkor az *e* szám egy nagy abszolút értékű negatív kitevőjű hatványa szerepel a fenti képletben. Az ilyen számok pozitívak, de nagyon közel vannak 0-hoz. Példánkban, ahol $\beta_\nu - \delta_\lambda = -4 - 4 = -8$, az e^{-8} szám 0,00033-nál kisebb, így a tört számlálója 0-hoz nagyon közeli, a nevezője viszont 1-hez közeli szám lesz (a 0,00033-hoz közeli számot adjuk hozzá az 1-hez), így a tört értéke (nagyon kicsi számláló, hozzá képest nagy nevező) 0-hoz közeli lesz. És valóban, a negatív képességfejlettség gyenge képességet jelent, a pozitív feladatnehézség nehéz feladatot, egy gyenge képességfejlettséggel rendelkező vizsgált személy old meg egy nehéz feladatot, a megoldás valószínűsége valóban 0 körül kell, hogy alakuljon.

Nézzük most azt az esetet, amikor $\beta_\nu - \delta_\lambda = 0$, vagyis a vizsgált személy képességfejlettsége, és a kiválasztott feladat nehézségparamétere pont egyenlők egymással. Ha az *e* kitevője 0, akkor a hatvány értéke 1. Ez azt jelenti, hogy a tört számlálója 1, míg a nevezője 2, vagyis a tört értéke ½, másképpen 0,5. Ha valaki egy olyan feladattal foglalkozik, amelynek a nehézségparamétere pontosan ugyanakkora, mint az ő képességfejlettsége, akkor a megoldás valószínűsége 0,5 lesz.

A jobb érthetőség kedvéért ábrázoljuk is az itt szóba került függvényt (4. ábra). A képletben a $\beta_\nu - \delta_\lambda$ különbséget jelöljük *x*-szel, a valószínűséget *y*-nal, és így már a jól megszokott jelölésben látjuk a függvényt:

$$y = \frac{e^x}{1+e^x}.$$

Az ábrázolás eredménye a 4. ábrán látszik.



4. ábra: A Rasch-modell képletének megfelelő függvény görbéje, ha a független változó (x) a képességfejlettség és a feladatnehézség különbsége, a függő változó pedig a jó feladatmegoldás valószínűsége

A „piros vonal” a függvény görbéje, s amint látjuk (és az előbb végiggondoltuk) mindig az *x* tengely fölött van a görbe, tehát a függvény értéke mindig pozitív, illetve mindig az $y = 1$ magasságban húzott, az *x* tengellyel párhuzamos egyenes alatt marad a görbe, ahogy láttuk, a függvény értéke mindig kisebb, mint 1. Minél nagyobb pozitív szám az *x*, annál jobban megközelíti a görbe az $y = 1$ egyenest (*y* értéke nagyon közel kerül 1-hez), és minél nagyobb abszolút értékű negatív szám az *x*, annál jobban megközelíti a görbe az *x* tengelyt (*y* értéke nagyon közel kerül 0-hoz). Az is látható, hogy az *y* tengelyt a görbe a 0,5 értéknél metszi, ennek is így kell lenni, hiszen láttuk, hogy ha $\beta_\nu - \delta_\lambda = x = 0$, akkor valóban ennyi a valószínűség.

A *Rasch* képlet „erejét” mutatja a következő megfontolás. Tegyük fel, hogy egy adott képesség esetén érvényes a modell, vagyis a képlet helyesen írja le a valószínűségeket, a végtelen sok emberhez és a végtelen sok feladathoz hozzá tudunk rendelni úgy megfelelő paramétereket (képességfejlettségeket és feladatnehézségeket), hogy a képlet alapján számítható valószínűségek megegyeznek a tényleges valószínűségekkel. És most tegyük fel, hogy ismerjük az összes valószínűséget. (Csak gondolatkísérletről van szó, a tényleges, valóságos valószínűségeket soha nem ismerhetjük.) Kiszámíthatók-e vajon a képesség- és a feladatnehézség paraméterek? Igen, és a megoldás voltaképpen nagyon egyszerű. Ha veszek egy konkrét embert, és egy konkrét feladatot, akkor ismerem annak valószínűségét, hogy ez az ember helyesen oldja meg az adott feladatot. A Rasch-képlet ad egy összefüggést a két ismeretlen paraméter és a valószínűség között. Ez azt jelenti, hogy két ismeretlenünk van, de csak egy egyenletünk, így még végtelen sok megoldás van. De vegyünk két embert és két feladatot! Így négyre nő az ismeretlenek száma (a két képességfejlettség és a két feladatnehézség), látszólag romlik a helyzet, de négy egyenletünk is lesz. Ugyanis összesen négy feladatmegoldásról van szó (két ember két feladatot old meg). Négy valószínűség képletét felírva négy egyenletünk lesz a négy ismeretlenre vonatkozóan. Ezt a négy egyenletből álló, négyismeretlenes egyenletrendszert megoldva[29] megkapjuk a két képességfejlettséget és a két feladatnehézséget. Ha most az egyik feladatot kicserélnénk egy másikra, vagy akár mindkettőt megváltoztatnánk, a kiválasztott két személy képességfejlettségére ugyanazokat az értékeket kellene kapnunk az új egyenletrendszer megoldásával, mint amilyen értékeket az előbb kaptunk[30]. Ez azért van, mert az adott képességre és az adott populációra érvényes a *Rasch-modell*, ebből indultunk ki. De kicserélhetnénk a személyeket is, akkor pedig a számítás ugyanazokat a feladatnehézségeket produkálná, mint az első esetben. Milyen egyszerű lenne a dolgunk, ha ismernénk a jó feladatmegoldások valószínűségeit! Sajnos éppen ezek az értékek teljes mértékben ismeretlenek.

Még egy megjegyzésre van szükség a későbbiek jó megértéséhez. A *Rasch* képlet megadja annak valószínűségét, hogy egy kiválasztott személy egy kiválasztott feladatot helyesen oldjon meg. De mennyi annak a valószínűsége, hogy nem helyes a megoldás? A valószínűségszámítás szabályai segítenek a válasz megadásában. Mivel a helyes és a helytelen válasz két egymást kizáró esemény, amelyeken kívül más nem képzelhető el, ezért a valószínűségeik összegének 1-nek kell lenni (ki kell adniuk együtt a biztos esemény valószínűségét). A jó megoldás valószínűségét, $p_{\nu\lambda}$-t már ismerjük, akkor a keresett valószínűség $1 - p_{\nu\lambda}$ lesz. Számítsuk ezt ki!

$$q_{\nu\lambda} = 1 - p_{\nu\lambda} = 1 - \frac{e^{\beta_\nu - \delta_\lambda}}{1 + e^{\beta_\nu - \delta_\lambda}} = \frac{1 + e^{\beta_\nu - \delta_\lambda} - e^{\beta_\nu - \delta_\lambda}}{1 + e^{\beta_\nu - \delta_\lambda}} = \frac{1}{1 + e^{\beta_\nu - \delta_\lambda}}$$

Vagyis a nem megfelelő feladatmegoldás valószínűségének kifejezése hasonlít a jó feladatmegoldáséhoz, a nevezője ugyanaz, csak a számlálója egyszerűbb, 1 az értéke.

A modern tesztelméleti skála intervallumskála

Megismerkedtünk a *Rasch-modell* alapképletével, tudjuk már, hogy ha ismernénk a képességfejlettségeket és a feladatnehézségeket, akkor azokból hogyan kellene kiszámolni a helyes feladatmegoldások valószínűségét. Természetesen a feladat nem ez, hanem nekünk éppen a képességfejlettségeket és a feladatnehézségeket kellene kiszámolnunk, elhelyezve őket egy-egy (látjuk majd, hogy azonos) intervallumskálán.

Most jött el az ideje annak, hogy megmutassam, az elmélet biztosítja, hogy itt valóban *intervallumskáláról* van szó. Az intervallumskálák konstrukciójának leglényegesebb mozzanata, hogy a mért entitások világában (ezek itt az emberek, illetve a feladatok, ne felejtsük el, hogy egyszerre foglalkozunk kétféle méréssel) kell találnunk egy ***négyváltozós relációt***, amely bizonyos, nemsokára itt is bemutatandó tulajdonságokkal rendelkezik, és a mérendő „dolgok” párjai elemeinek megfelelően értelmezett „távolságai” közötti rendezési jellegű relációt jelent. Kicsit konkrétabban: valamilyen módon értelmeznünk kellene a képességfejlettségek közötti „távolságokat”, és empirikusan is ellenőrizhető módon tudnunk kellene, hogy két pár esetén melyik pár között nagyobb a „távolság” (és analóg módon ugyanez szükséges a feladatnehézségek vizsgálatával kapcsolatban is). Ha ez a reláció meglesz, akkor már csak meg kell alkotnunk

[29] Ennyire mégsem egyszerű a helyzet. A négy egyenlet nem független egymástól, csak bizonyos feltételek fennállása esetén van megoldása, akkor viszont végtelen sok, a megoldásrendszerek abban különböznek egymástól, hogy két ilyen megoldásnégyesben szereplő, egymásnak megfelelő értékek ugyanannyival különböznek egymástól.

[30] Ismét pontosíthatunk: valójában csak az igaz, hogy bármelyik megoldást vesszük is a végtelen sok közül, a két képességfejlettség közötti különbség ugyanakkora lesz, mint az előbbi esetben, tehát amikor még az eredeti feladatokat oldattuk meg a két személlyel.

a képességfejlettségek valós számokra való leképezését, vagyis minden képességfejlettséghez hozzá kell rendelnünk megfelelő módon egy számot, és ha ez a megfeleltetés homomorf, akkor a reprezentációs méréselmélet biztosítja számunkra, hogy így egy intervallumskála jött létre. Ugyanez az eljárás a feladatnehézségek halmazával is.

Most egy matematikai szempontból nehezebb rész következik, a matematikai részletek iránt kevésbé érdeklődő olvasó bátran átugorhatja. A lényeg az, hogy a leírásban szereplő érvelés bizonyítja, hogy a modern tesztelmélet segítségével valóban intervallumskála jön létre.

Ahhoz, hogy bebizonyítsuk, a képességfejlettségek (és a feladatnehézségek) intervallumskálán helyezkednek el, tudnunk kellene, hogy az elmélet mit mond két képességfejlettség (ezután már csak ezt tárgyalom) különbségéről. Ehhez végezzünk el egy számítást! Ebben *Benjamin D. Wright* gondolatmenetét követem (Wright 1983).

Legyen az a személy képességfejlettségének szintje β_a, a b személyé β_b. Tekintsünk egy tetszőleges feladatot, ennek nehézségparamétere legyen δ. Mennyi annak a valószínűsége, hogy ezt a feladatot a helyesen, míg b rosszul oldja meg, és mennyi a fordított eset valószínűsége? Ezt a két kérdést azért tesszük fel, mert a két személy képességfejlettsége közötti „különbség" egy számunkra előnyös meghatározása e két valószínűség segítségével lesz lehetséges.

Egy viszonylag egyszerű, négy elemi eseményt tartalmazó eseményrendszerről van szó. Ezek az elemi események:

A: Az a személy meg tudja oldani a feladatot, és a b személy is.
B: Az a személy meg tudja oldani a feladatot, a b személy nem.
C: Az a személy nem tudja megoldani a feladatot, a b személy igen.
D: Az a személy nem tudja megoldani a feladatot, s a b személy sem.

Az az esemény, hogy a személy megoldotta, b viszont nem oldotta meg a feladatot, az itteni felsorolásban a B elemi esemény. Valószínűségét kiszámíthatjuk abból, hogy ismerjük két további eseménynek a valószínűségét:

X: a személy meg tudja oldani a feladatot.
Y: b személy nem tudja megoldani a feladatot.

X és Y események szorzata nem más, mint a B esemény, és mivel függetlenek, ezért a B esemény valószínűsége egyenlő az X és az Y események valószínűségének szorzatával. Ugyanígy a C esemény a következő két esemény szorzata:

U: a személy nem tudja megoldani a feladatot
V: b személy meg tudja oldani a feladatot.

Jelölje p_{ab} a B esemény valószínűségét, és p_{ba} a C esemény valószínűségét. Legyen továbbá p_X, p_Y, p_U, és p_V rendre az X, Y, U és V események valószínűsége. Vagyis

$$p_{ab} = p_X p_Y,$$

$$p_{ba} = p_U p_V.$$

A *Rasch-modell* szerint tudjuk, mekkorák az X, Y, U és V események valószínűségei (amennyiben a vizsgált jelenségre érvényes az összefüggés):

$$p_X = \frac{e^{\beta_a - \delta}}{1 + e^{\beta_a - \delta}}$$
$$p_Y = \frac{1}{1 + e^{\beta_b - \delta}}$$
$$p_U = \frac{1}{1 + e^{\beta_a - \delta}}$$
$$p_V = \frac{e^{\beta_b - \delta}}{1 + e^{\beta_b - \delta}}$$

Így már meghatározhatjuk a p_{ab} és a p_{ba} valószínűségeket:

$$p_{ab} = p_X\, p_Y = \frac{e^{\beta_a - \delta}}{1 + e^{\beta_a - \delta}} \frac{1}{1 + e^{\beta_b - \delta}} = \frac{e^{\beta_a - \delta}}{\left(1 + e^{\beta_a - \delta}\right)\left(1 + e^{\beta_b - \delta}\right)}$$
$$p_{ba} = p_U\, p_V = \frac{1}{1 + e^{\beta_a - \delta}} \frac{e^{\beta_b - \delta}}{1 + e^{\beta_b - \delta}} = \frac{e^{\beta_b - \delta}}{\left(1 + e^{\beta_a - \delta}\right)\left(1 + e^{\beta_b - \delta}\right)}$$

Számítsuk ki e két valószínűség hányadosának természetes alapú logaritmusát! (Hogy miért, az a számítás végén tűnik majd elő.) Ehhez tegyük fel, hogy sem p_{ab}, sem p_{ba} nem 0 (a képletek szerint nem is lehet).

$$\ln \frac{p_{ab}}{p_{ba}} = \ln \frac{e^{\beta_a - \delta}}{\left(1 + e^{\beta_a - \delta}\right)\left(1 + e^{\beta_b - \delta}\right)} / \frac{e^{\beta_b - \delta}}{\left(1 + e^{\beta_a - \delta}\right)\left(1 + e^{\beta_b - \delta}\right)} =$$
$$= \ln \frac{e^{\beta_a - \delta}}{e^{\beta_b - \delta}} = \ln \frac{e^{\beta_a}}{e^{\beta_b}} = \ln e^{\beta_a} - \ln e^{\beta_b} = \beta_a - \beta_b$$

Azt kaptuk, hogy az elmélet szerint két személy képességfejlettségének különbsége két valószínűség hányadosának természetes alapú logaritmusával egyenlő. Lényeges, hogy a levezetés során eltűnt a δ, vagyis a feladat nehézsége, ami azt jelenti, hogy az összefüggés akármelyik feladat esetén pontosan ugyanígy nézne ki, a kapott eredmény nem függ attól, melyik feladattal foglalkozunk. Vagyis a két valószínűség hányadosának természetes alapú logaritmusa ugyanaz az érték, bármelyik feladattal kapcsolatosak is a valószínűségek. Ez az érték tehát valamilyen módon a két személy egymáshoz való viszonyára jellemző, a választott feladattól teljesen függetlenül. Ez természetesen csak akkor lehet érvényes, ha az egész jelenséget (a képességhez tartozó összes feladat minden a populációba tartozó személy általi megoldására vonatkozóan) a *Rasch-modell* helyesen írja le.

Intervallumskálát szeretnénk alkotni. Úgy szeretnénk leképezni az emberek (képességfejlettségeik) halmazát a valós számok halmazára, hogy egy az emberek halmazán definiált, empirikusan ellenőrizhető relációnak, a négyváltozós D relációnak feleljen meg a számok közötti különbségek közti rendezés. Pontosabban: lennie kell egy négyváltozós D relációnak az emberek halmazán (az axiómák rögzítik a tulajdonságait, ezeket alább tárgyalom), és ha *abDst*, vagyis ha a, b, s, t képességfejlettségek ilyen sorrendben a D relációban állnak egymással, akkor a hozzájuk rendelt képességfejlettség számértékekre (β értékek) a következő összefüggésnek kell igaznak lenni:

$$\beta_a - \beta_b > \beta_s - \beta_t.$$

A fenti levezetésben éppen azt találtuk, hogy a képességfejlettségek különbsége – az elmélet szerint – két valószínűség hányadosának természetes alapú logaritmusával egyezik meg. Ez kezünkbe adja a D reláció konstrukciójának kulcsát.

DEFINÍCIÓ: Az *a*, *b*, *s*, *t* képességfejlettségek ilyen sorrendben akkor és csak akkor állnak egymással a *D* relációban, ha bármely az adott képességhez tartozó feladatra nézve annak a valószínűsége, hogy *a* megoldja a feladatot, de *b* nem, osztva annak a valószínűségével, hogy *b* megoldja a feladatot, de *a* nem, illetve véve az így kapott szám természetes alapú logaritmusát, nagyobb lesz, mint az a szám, amelyet ugyanezzel a módszerrel kapunk *s* és *t* személyekre elvégezve a számítást.

Képlettel:

$$abDst \Leftrightarrow \ln \frac{p_{ab}}{p_{ba}} > \ln \frac{p_{st}}{p_{ts}}$$

A természetes alapú logaritmus használatára valójában nincs szükség, mert az egyenlőtlenség akkor és csak akkor érvényesül, ha az argumentumok között ugyanaz az egyenlőtlenség áll fenn (a logaritmus függvény 1-nél nagyobb alap esetén szigorúan monoton növekvő). Így a *D* reláció definíciója lehet ez is:

$$abDst \Leftrightarrow \frac{p_{ab}}{p_{ba}} > \frac{p_{st}}{p_{ts}}.$$

Ez mindenképpen egy reláció a képességfejlettségek halmazán, tetszőlegesen kiválasztott *a*, *b*, *s*, *t* személyek (képességfejlettségeik) között vagy érvényesül, vagy nem, van reményünk rá, hogy megfeleljen az igényeinknek. Láttuk, hogy az $\ln(p_{xy}/p_{yx})$ mennyiség – amennyiben érvényesül a *Rasch-modell* – csakis a két személy egymáshoz való viszonyára jellemző, független a feladat választásától. Vagyis ez a mennyiség egyfajta „távolság" szerepet játszik. Nagyon fontos következménye a konstrukciónak, s ezt később még jelentős mértékben kihasználjuk, hogy a *D* reláció érvényesülésének vizsgálata valójában számok közötti szigorú egyenlőtlenség vizsgálatát jelenti.

A reprezentációs méréselmélet szerint a *D* relációnak ki kell elégíteni bizonyos igényeket. Most megvizsgálom ezen igények teljesülését (néhol egyszerűsítve – és ezzel némileg pontatlanná téve – a bemutatást). Alapként a *Fred Roberts* által leírt axiómarendszert használom (Roberts, 1979).

Még mielőtt az axiómák érvényességét vizsgálnám, be kell vezetnem három újabb relációt a fogalmazás megkönnyítése érdekében, ezek mind a *D* relációval állnak szoros kapcsolatban. Az *E* reláció szintén négyváltozós, és *a*, *b*, *s*, *t* személyek (képességfejlettségeik) akkor állnak ebben a relációban egymással, tehát akkor igaz *abEst*, ha sem ebben a sorrendben, sem a párok megcserélésével nem áll fenn köztük a *D* reláció, vagyis nem igaz *abDst* sem, és *stDab* sem. Mivel *D* egyfajta „rendezési" jellegű reláció (szemléletesen: *a* és *b* közötti „távolság" nagyobb, mint *s* és *t* közti „távolság"), ezért ha egyik párra sem igaz, hogy nagyobb lenne a „távolságuk" a másik pár elemeinek „távolságánál", akkor szemléletesen ez a „távolságok egyenlőségét" jelenti. Ha *D* egyfajta „rendezés" a párok „távolságai" közt, akkor *E* egyfajta „egyenlőség" közöttük – szemléletesen fogalmazva. Konkrét esetünkben, tehát a képességfejlettség mérésében a „távolságok" számokként adhatók meg, így itt ténylegesen is számok közötti egyenlőségről van szó.

W-vel jelöljük azt a relációt, amely szintén négyváltozós, és akkor áll fenn az *a*, *b*, *s*, *t* személyekre (képességfejlettségeikre), ha vagy *D*, vagy *E* fennáll közöttük (az elemek leírt sorrendjében). Ez a reláció pedig a „nagyobb vagy egyenlő" (számok közötti) relációhoz hasonlít, illetve a képességfejlettség mérésében, az itteni tárgyalásnak megfelelően, tekintve, hogy *D* és *E* is számok közötti relációk, valóban a „≥" relációról van szó.

Végül az *R* reláció kétváltozós, és a személyek halmazának *párjain* értelmezett, s ugyanazt fejezi ki az (*a*, *b*), valamint az (*s*, *t*) párokkal kapcsolatban, mint amit a *D* reláció az *a*, *b*, *s*, *t* személyekről (képességfejlettségeikről) mond. Vagyis (*a*, *b*)*R*(*s*, *t*) akkor és csak akkor igaz, ha *abDst*. Világos, hogy *R* fennállásának ellenőrzése is számok összehasonlítását igényli, és akkor teljesül, ha a baloldalán álló szám (*a* és *b* személyek fentebb meghatározott „távolsága"), határozottan nagyobb, mint a jobboldalán álló szám (*s* és *t* személyek „távolsága"). Vagyis *R* nem más, mint számok közti „>" reláció.

Az **első axióma** (Roberts, 1979, 136. o.) azt követeli, hogy az *R* reláció *szoros gyenge rendezés* legyen. E fogalom magyarázata a következő: egy rendezés szorosan gyenge, ha aszimmetrikus és negatívan tranzitív. Ezek persze újabb tulajdonságok, amelyeket meg kell magyarázni. Egy Q reláció aszimmetrikus, ha igaz, hogy amennyiben x és y egymással ilyen relációban állnak, vagyis xQy, akkor nem igaz, hogy „fordítva" is ilyen relációban lennének, vagyis nem igaz yQx. Rögtön nézzük meg, hogy a mi relációnk ilyen-e! A mi R relációnk, ahogy láttuk, számok közti szigorú rendezési relációt jelent. Márpedig ez a reláció rendelkezik a leírt tulajdonsággal (ha egy szám határozottan nagyobb, mint egy másik, akkor ennek fordítottja nem lehet igaz).

Egy Q kétváltozós reláció akkor negatívan tranzitív, ha minden x, y, z elemre igaz, hogy ha x és z ilyen relációban állnak egymással, akkor vagy az igaz, hogy x és y ilyen relációban állnak egymással, vagy y és z ilyen relációban állnak egymással. A vizsgálatunkban a szigorú (tehát „>" jellel jelölt) rendezés a fontos. Ha az x szám határozottan nagyobb, mint a z, akkor y nem „teheti meg", hogy nagyobb legyen, mint x, és egyben kisebb legyen, mint z. Az egyiknek mindenképpen hamisnak kell lenni. Ha y sem x-szel, sem z-vel nem egyezik meg, akkor vagy $x > y$-nak kell teljesülnie, vagy $y > z$-nek. Ha pedig valamelyikkel egyezik, akkor a másikkal való elvárt egyenlőtlenség érvényesül. A számok esetén tehát a negatív tranzitivitás teljesül. Tekintve, hogy a vizsgálatunkban az R reláció két oldalán álló „valamik" valójában számok (valószínűségek hányadosainak természetes alapú logaritmusai), és az R reláció tulajdonképpen valós számok közötti szigorú rendezési reláció, ezért R-re a negatív tranzitivitás is teljesül. Ezek szerint az R reláció aszimmetrikus és negatívan tranzitív, így szorosan gyenge rendezés, tehát az első axiómát kielégíti.

A **második axióma** (Roberts 1979, 136. o.) azt követeli, hogy ha igaz $abDst$, akkor igaznak kell lenni $tsDba$-nak is. Vagyis „fordítsuk meg" a párok elemeit, s ekkor a relációnak is „fordítva" kell érvényesülnie. A párok elemeinek „megfordítása" azt jelenti a mi esetünkben, hogy a valószínűségeket nem az eredetileg megadott sorrendben, hanem ahhoz képest „fordítva" osztjuk el egymással. Így az eredeti hányadosok reciprokait kapjuk, amelyek között az egyenlőtlenség éppen fordítva áll fenn, mint az előbb (a természetes alapú logaritmus számítása ezen nem változtat). Vagyis azt láthatjuk, hogy a 2. axióma is érvényes.

A **harmadik axióma** (Roberts 1979, 136. o.) azt követeli, hogy ha a, b, c, s, t, u hat személy (vagy képességfejlettség), és igaz, hogy $abWst$, valamint igaz, hogy $bcWtu$, akkor $acWsu$ is igaz. Írjuk fel, hogy a mi esetünkben mit is jelentenek a feltételben megadott relációk!

$$abWst \rightarrow \ln(p_{ab}/p_{ba}) \geq \ln(p_{st}/p_{ts})$$
$$bcWtu \rightarrow \ln(p_{bc}/p_{cb}) \geq \ln(p_{tu}/p_{ut})$$

És írjuk fel a következményt is:

$$acWsu \rightarrow \ln(p_{ac}/p_{ca}) \geq \ln(p_{su}/p_{us})$$

Itt p_{fg} jelöli annak a valószínűségét, hogy a kiválasztott feladatot f meg tudja oldani, de g nem ($(f, g) \in \{(a, b), (b, a), (b, c), (c, b), (s, t), (t, s), (t, u), (u, t), (a, c), (c, a), (s, u), (u, s)\}$). Az axióma érvényesül, amennyiben bármely „szóba kerülő" személy és bármely feladat esetében a jó feladatmegoldás valószínűsége nem 1 és nem is 0. Ezt a következőképpen láthatjuk be.

Hagyjuk el a természetes alapú logaritmust mindkét feltételből és az állításból is (ld. a korábbi megjegyzést)! Azt kell tehát bizonyítanunk, hogy

ha

$$p_{ab}/p_{ba} \geq p_{st}/p_{ts} \text{ és } p_{bc}/p_{cb} \geq p_{tu}/p_{ut},$$

akkor

$$p_{ac}/p_{ca} \geq p_{su}/p_{us}.$$

Alakítsuk át ezeket a kifejezéseket, szorozzunk be a nevezőkkel! A bizonyítandó állítás így a következő:

ha

$$p_{ab}p_{ts} \geq p_{st}p_{ba} \text{ és } p_{bc}p_{ut} \geq p_{tu}p_{cb},$$

akkor

$$p_{ac}p_{us} \geq p_{su}p_{ca}.$$

Most az itt szereplő valószínűségek helyébe írjuk be részletesebb kifejezéseiket, amelyek már csak olyan változókat tartalmaznak, amelyek egy-egy személy jó feladatmegoldásának valószínűségei! A bizonyítandó állítás ekkor a következő lesz:

ha

$$p_a(1 - p_b)p_t(1 - p_s) \geq p_s(1 - p_t)p_b(1 - p_a)$$

és

$$p_b(1 - p_c)p_u(1 - p_t) \geq p_t(1 - p_u)p_c(1 - p_b),$$

akkor

$$p_a(1 - p_c)p_u(1 - p_s) \geq p_s(1 - p_t)p_c(1 - p_a).$$

Most osszuk el az első feltételt kifejező egyenlőtlenség mindkét oldalát a pozitív $p_ap_bp_sp_t$–vel, a másodikat $p_bp_cp_up_t$ –vel, míg a következtetést kifejező egyenlőtlenség mindkét oldalát osszuk el a szintén pozitív $p_ap_cp_up_s$ –sel! Elvégezve egyszerűsítéseket, osztásokat a következőképpen alakul át a bebizonyítandó állítás:

ha

$$(1/p_b - 1)(1/p_s - 1) \geq (1/p_a - 1)(1/p_t - 1)$$

és

$$(1/p_c - 1)(1/p_t - 1) \geq (1/p_b - 1)(1/p_u - 1),$$

akkor

$$(1/p_c - 1)(1/p_s - 1) \geq (1/p_a - 1)(1/p_u - 1).$$

Vezessük be a következő jelöléseket: $1/p_a - 1 = A$, $1/p_b - 1 = B$, $1/p_c - 1 = C$, $1/p_s - 1 = S$, $1/p_t - 1 = T$, $1/p_u - 1 = U$! Vegyük észre, hogy mivel a p értékek 1-nél kisebb pozitív számok, ezért

$$A, B, C, S, T, U > 0.$$

Az igazolandó állítás erre egyszerűsödik: ha $BS \geq AT$ és $CT \geq BU$, akkor $CS \geq AU$. Ez az állítás már könnyen bizonyítható. Osszuk el a második feltételt kifejező egyenlőtlenség mindkét oldalát a pozitív C-vel. Ekkor ezt kapjuk: $T \geq BU/C$. Felhasználva ezt a T-re vonatkozó egyenlőtlenséget az első feltételt kifejező egyenlőtlenségben: $BS \geq AT \geq ABU/C$, majd B-vel egyszerűsítve, és C-vel megszorozva mindkét oldalt (pozitív számok!) ezt kapjuk: $CS \geq AU$. És éppen ezt akartuk bizonyítani. Kimondhatjuk tehát, hogy a harmadik axióma is érvényesül.

A **negyedik axióma** (Roberts 1979, 136. o.) nem szükséges, „csak” elégséges feltételt fogalmaz meg. Ez egyfajta megoldhatósági kritérium. A könyvben szereplő formális megfogalmazás helyett kissé szemléletesebben fogalmazva: ha az a és b elemek közti „távolság” „nagyobb vagy egyenlő”, mint az s és a t közti „távolság” (vagyis $abWst$ érvényesül), akkor találunk olyan u és v elemeket, hogy a és u közötti „távolság” megegyezzék az s és a t közti távolsággal ($auEst$), illetve a v és a b „távolsága” megegyezzék s és t „távolságával” ($vbEst$). Számunkra ez azt jelenti, hogy mindig találhatunk megfelelő személyeket, hogy a relációk fennálljanak. Mivel ideálisan (az elméleti modellben) mindig úgy gondolkodunk, hogy bármilyen képességfejlettséghez van olyan ember, akinek éppen az a képességfejlettsége, ezért triviális, hogy ez az axióma is teljesül.

Roberts rendszerében még egy axióma van (Roberts 1979, 137. o.). Megértéséhez ismernünk kell a sztenderd sorozat fogalmát. Legyen a és b két személy, akiknek nem azonos a képességfejlettsége. Korábban értelmeztük a köztük lévő távolságot. Legyen most c egy olyan személy, hogy igaz legyen $abEbc$, vgyis c legyen olyan távolságra b-től, mint amilyen távolságra b van a-tól. Most a d személyt úgy választjuk, hogy $abEcd$ legyen igaz. Ha r-rel jelöljük a és b távolságát, akkor b és c távolsága is r lesz, de az utóbb felírt reláció fennállása miatt c és d távolsága is r. Látható, hogy így egy sorozatot állíthatunk elő, bármelyik két szomszédos elem távolsága r, megegyezik az első kettőnek választott elem távolságával. Az elemek ilyen sorozatát hívjuk sztenderd sorozatnak.

Az **ötödik axióma** tartalmát kifejező állítás: *minden szigorúan korlátos sztenderd sorozat véges*. Ezt úgy is fogalmazhatjuk, hogy egy végtelen sztenderd sorozat nem lehet szigorúan korlátos. Egy képességfejlettségekből álló sztenderd sorozat két szomszédos eleme közt a „távolság” (két valószínűség hányadosának természetes alapú logaritmusa) egy szám, méghozzá bármely két szomszédos elem esetén ugyanaz a szám. Ha a sorozat végtelen, akkor az első elemétől a második valamilyen r „távolságra” van. Az egymás utáni elemek (második, harmadik, negyedik, stb.) ezért r,

$2r$, $3r$, … „távolságra” vannak az első elemtől (ezek negatív számok is lehetnek), ami viszont azt jelenti, hogy a sorozat valóban nem lehet korlátos.

A matematikai szempontból korrekt elemzés azt mutatja, hogy *a modern tesztelmélet számára felépített modell kielégíti az ún. különbségi struktúra axiómáit*, vagyis különbségi struktúra, és így a modern tesztelmélet keretében előállított skálák intervallumskálák. Ez egy fontos eredmény, a modern tesztelmélet jelentőségét bizonyítja. Itt tehát nem vethetők fel méréselméleti problémák. Amennyiben olyan képességgel van dolgunk a mérés során, amely jól kielégíti a *Rasch-modell* feltételeit, bizton számíthatunk rá, hogy a mérhetővé tett értékek intervallumskálán helyezkednek el.

A paraméterek becslésének alapjai

A valóságos mérési szituációkban nem végtelen sok emberről, és nem végtelen sok feladatról van szó. Konkrét *mintán* konkrét *teszttel* végezzük a vizsgálatot. Készítsünk a szóba kerülő valószínűségekből – képzeletben, hiszen ezeket az értékeket nem ismerjük – egy *táblázatot*. Soroljuk fel a tesztben szereplő feladatokat, legyen a számuk L, és soroljuk fel a mérésben szereplő személyeket is, legyen a számuk N. Balról jobbra haladva alkotják majd a táblázat oszlopait az egymás után következő feladatokhoz tartozó adatok, míg a sorokban felülről lefelé haladva a vizsgált személyek adatai szerepelnek majd. Pontosabban: a ν-edik sorban, annak a λ-adik helyén, vagyis a λ-adik oszlopban helyezzük el a $p_{\nu\lambda}$ értéket, azaz annak a valószínűségét, hogy a ν-edik személy helyesen oldotta meg a λ-adik feladatot. Így néz ki ez a táblázat:

$$\begin{bmatrix} p_{1,1} & p_{1,2} & \dots & p_{1,L} \\ p_{2,1} & p_{2,2} & \dots & p_{2,L} \\ \dots & \dots & \dots & \dots \\ p_{N,1} & p_{N,2} & \dots & p_{N,L} \end{bmatrix}$$

Itt minden egyes sorban a vizsgált személyek valamelyikének helyes feladatmegoldásokhoz tartozó valószínűségei szerepelnek. Nevezzük ezután a valószínűségeknek ezt a halmazát és elrendezését a *jó feladatmegoldások valószínűségei rendszerének,* illetve *táblázatának*. Érdekes lenne, ha az egyes valószínűségekre tudnánk becsléseket adni. Feladatonként és személyenként tudnánk, legalább hozzávetőlegesen, hogy mennyi az esélye a jó feladatmegoldásnak.

Hogyan lehet becsülni ezeket a valószínűségeket? Induljunk ki abból, hogy kezdetben, ha még semmilyen információt nem veszünk figyelembe, valójában ezek a valószínűségek tetszőleges (0 és 1 közötti) értékeket felvehetnének. Képzeljük el, hogy az előbbi táblázatban szereplő értékek mind konkrét, 0 és 1 közötti, tetszőleges értékek, és ezek fejezik ki a jó feladatmegoldások valószínűségét! Bármennyi is a $p_{\nu\lambda}$ valószínűség, ennyi annak a valószínűsége, hogy a ν-edik személy helyesen oldja meg a λ-adik feladatot. Lehet, hogy ez a tetszőlegesen választott $p_{\nu\lambda}$ valószínűség kisebb, mint 0,5, és ekkor inkább azt várjuk, hogy a ν-edik személy nem tudja megoldani a feladatot. Lehet viszont, hogy mégis meg tudja oldani, ez egyáltalán nem lehetetlen. Furcsálljuk ugyan egy kicsit, de elfogadjuk, hogy bekövetkezett egy viszonylag kis valószínűségű esemény. Ha viszont a $p_{\nu\lambda}$ valószínűség nagyobb, mint 0,5, akkor inkább azt várjuk, hogy a ν-edik személy megoldja a feladatot. Ha ennek az ellenkezője következik be, akkor ezt is kissé furcsállhatjuk, de végül is nem lehetetlen az esemény.

Meg lehetne kérdezni, hogy a jó feladatmegoldásokhoz köthető valószínűségeknek melyik az a rendszere, amely esetén a mérés során kapott konkrét megoldásmintázat a legvalószínűbb. Ugyanis a valószínűségeknek ezt a rendszerét tekinthetnénk olyannak, ami a jelenség mögött állhat. Ha tényleg ilyen a jó megoldások valószínűségeinek rendszere, akkor minden más megoldásmintázat valószínűsége kisebb lenne annál, mint ami konkrétan a feladatmegoldásokból kijött megoldásmintázat valószínűsége lenne. De hogyan lehet meghatározni egy megoldásmintázat valószínűségét? Úgy, hogy összeszorozzuk a konkrétan bekövetkezett események valószínűségeit. Ha egy adott feladatot egy adott személy meg tudott oldani, akkor a jó feladatmegoldás valószínűségét használjuk fel, ha nem tudta megoldani, akkor a rossz megoldás valószínűségét, ami természetesen így írható fel ($q_{\nu\lambda}$-val jelölve a hibás megoldás valószínűségét):

$$q_{\nu\lambda} = 1 - p_{\nu\lambda}.$$

Egyáltalán nem nyilvánvaló, hogy a tesztelés során konkrétan bekövetkezett esemény (minden egyes személy bizonyos feladatokat meg tudott oldani, bizonyosakat nem) valószínűségét, az egyes részesemények valószínűségeinek szorzataként lehet kiszámítani. Ez csak akkor igaz, ha az egyes események (egy személy megold egy feladatot, és az sikeres lesz vagy sem) függetlenek egymástól. Ha például csalás történik, akkor ez a függetlenség nem teljesül. Vagy például, ha a feladatokat rosszul állítottuk össze, és az egyik megoldásának eredménye befolyásolja a másikban a sikerességet, akkor is baj van a függetlenséggel. De ha nincs ilyen probléma, akkor azt mondjuk, hogy érvényesül a *lokális függetlenség*, a konkrétan kapott megoldásmintázat valószínűsége az egyes események valószínűségeinek ($p_{\nu\lambda}$ vagy $q_{\nu\lambda}$) szorzataként számítható.

A *jó megoldások valószínűségeinek rendszere* mellett tehát kialakíthatjuk a *tényleges események valószínűségeinek rendszerét*, amelyben egy adott helyen a táblázatban vagy változatlanul a jó megoldás valószínűsége szerepel, ha az adott feladatot az adott személy meg tudta oldani, vagy a rossz megoldás valószínűsége, ellenkező esetben. Ha ér vényes a lokális függetlenség, akkor összeszorozva az összes ténylegesen bekövetkezett esemény valószínűségét, megkapjuk annak valószínűségét, hogy a mérés során az adott mintán (személyek), az adott tesztet használva (feladatok) mennyi a valószínűsége annak, hogy éppen a kapott *megoldásmintázat* jöjjön ki.

Ha tetszőlegesen hozzárendelhetünk a jó feladatmegoldáshoz valószínűségeket, akkor meg tudjuk mondani, hogy e valószínűségrendszerek között melyik nyújtja a legnagyobb valószínűséget a mérés során kapott megoldásmintázatra. Az, amelyben a valószínűség 1, amikor arról van szó, hogy az adott személy meg tudta oldani a feladatot, és 0 a valószínűség, ha nem tudta megoldani. Ugyanis így minden szereplő esemény valószínűsége 1 lesz. Hiszen ha megoldotta, akkor eleve 1 a valószínűség, ha nem oldotta meg, akkor 1-ből ki kell vonni a jó megoldás valószínűségét, de mivel ez utóbbi ilyen esetben 0, ezért az előállt eseménynek, a nem megoldásnak is 1 a valószínűsége. Ha minden esemény valószínűsége 1, akkor a szorzatuk is 1, márpedig ez a maximális lehetséges valószínűség. Elég természetes is a kapott eredmény, az a valószínűségrendszer adhatja az eredményhez legjobban illeszkedőt, amely esetén maximális a jól megoldott feladatoknál a jó megoldás valószínűsége, és maximális a rosszul megoldott feladatok esetén a nem megoldás valószínűsége. Így aztán a szorzat is a lehető legnagyobb lesz.

Ha tehát csak annyit tudnánk, amennyit eddig felhasználtunk (lényegében csak a valószínűségszámítás bizonyos szabályait), akkor egyértelmű választ adhatnánk arra, hogy egy a mérés során kapott megoldásmintázat milyen tényleges eseményekhez köthető valószínűségrendszer esetén lenne maximális valószínűségű. Ez a válasz az eredeti problémánkra természetesen *érdektelen*. Mi olyan esetekben szeretnénk tudni, hogy melyik valószínűségrendszer nyújtja a maximális valószínűséget a megoldásmintázatra, amely esetekben a valószínűségek rendszere kielégíti a *Rasch* képletbe foglalt kritériumot. Nem lehetséges mégsem minden tetszőleges valószínűség az egyes eseményekkel (jó feladatmegoldásokkal) kapcsolatban, csak olyan valószínűségrendszerek jöhetnek szóba, amelyeknek elemei speciálisak, amelyek ténylegesen lehetnek kiszámított értékei a *Rasch* képletnek. A lehetséges valószínűségrendszerek halmazát leszűkítjük a *Rasch* képletnek megfelelőkre. (Az előbb leírt, a tényleges eseményeket tekintve csupa 1-esből álló valószínűségrendszer nem elégítheti a Rasch képlet követelménnyeit, hiszen a képlet nem is adhat ilyen értéket a helyes megoldásra, és 0-t sem a hibásra.)

A kérdés ezután így fogalmazható meg: melyik az a valószínűségrendszer, amely esetében (1) a mérésben konkrétan kapott megoldásmintázat a legnagyobb valószínűségű, és (2) kielégíti a *Rasch* képletet. Ez utóbbi azt jelenti, hogy van a feladatnehézségeknek és a képességfejlettségeknek egy-egy olyan sorozata, amelyeknek tagjait behelyettesítve a *Rasch* képletbe megkapjuk az adott feladat és az adott személy esetén a jó feladatmegoldás valószínűségét. A feladat tényleges megoldásának menetét csak az alábbi, keretben leírt szövegrészben adom meg, a matematikai részletek iránt nem érdeklődő olvasó bátran kihagyhatja ezt.

Használjuk az eddig már bevezetett jelöléseket! Van tehát *N* személy, *L* számú feladat, és minden egyes eseményhez, hogy ti. a ν-edik személy jól tudja megoldani a λ-adik feladatot, tartozik egy $p_{\nu\lambda}$ valószínűség. A rossz megoldás valószínűsége: $q_{\nu\lambda} = 1 - p_{\nu\lambda}$ ($\nu = 1, 2, \ldots, N$, és $\lambda = 1, 2, \ldots, L$). Feltételezzük, hogy létezik a β_ν képességfejlettségek (valós számok, $\nu = 1, 2, \ldots, N$), valamint a δ_λ feladatnehézségek (valós számok, $\lambda = 1, 2, \ldots, L$) egy-egy sorozata, hogy minden $p_{\nu\lambda}$ valószínűségre

$$p_{\nu\lambda} = \frac{e^{\beta_\nu - \delta_\lambda}}{1 + e^{\beta_\nu - \delta_\lambda}}.$$

A rossz megoldás valószínűsége:

$$q_{\nu\lambda} = 1 - \frac{e^{\beta_\nu - \delta_\lambda}}{1 + e^{\beta_\nu - \delta_\lambda}} = \frac{1}{1 + e^{\beta_\nu - \delta_\lambda}}.$$

Legyen továbbá $x_{\nu\lambda}$ értéke 1 vagy 0, attól függően, hogy a ν-edik személy jól tudja megoldani a λ-adik feladatot, vagy sem. Felírjuk a ténylegesen bekövetkezett esemény valószínűségét. A ténylegesen bekövetkezett esemény valószínűsége vagy $p_{\nu\lambda}$, vagy $q_{\nu\lambda}$, attól függően, hogy sikerült megoldani a feladatot, vagy sem. Jelöljük ezt a valószínűséget $\pi_{\nu\lambda}$-val. Kis meggondolás után felírhatjuk:

$$\pi_{\nu\lambda} = \frac{e^{x_{\nu\lambda}(\beta_\nu - \delta_\lambda)}}{1 + e^{\beta_\nu - \delta_\lambda}},$$

hiszen $x_{\nu\lambda} = 1$ esetén, vagyis amikor sikerült megoldani a feladatot, visszakapjuk $p_{\nu\lambda}$ kifejezését, ha viszont $x_{\nu\lambda} = 0$, akkor a számlálóban e^0 lesz, aminek értéke 1, vagyis valóban $q_{\nu\lambda}$ értékét kapjuk.

Ezután – feltételezve, hogy igaz a korábban már bevezetett lokális függetlenség – írjuk fel a valóságosan bekövetkezett események valószínűségeinek ($\pi_{\nu\lambda}$) szorzatát. Ez lesz annak a valószínűsége, hogy éppen a valóságos tesztmegoldásban kapott megoldásmintázat (az $x_{\nu\lambda}$ értékek) legyen az eredmény:

$$P=\prod_{\nu=1}^{N}\prod_{\lambda=1}^{L}\pi_{\nu\lambda}=\frac{\prod_{\nu=1}^{N}\prod_{\lambda=1}^{L}e^{x_{\nu\lambda}(\beta_\nu-\delta_\lambda)}}{\prod_{\nu=1}^{N}\prod_{\lambda=1}^{L}\left(1+e^{\beta_\nu-\delta_\lambda}\right)}.$$

A számlálóban azonos alapú hatványok szorzata áll, ami felírható az azonos alapnak a kitevők összegére való emelésével:

$$P=\frac{e^{\sum_{\nu=1}^{N}\sum_{\lambda=1}^{L}x_{\nu\lambda}(\beta_\nu-\delta_\lambda)}}{\prod_{\nu=1}^{N}\prod_{\lambda=1}^{L}\left(1+e^{\beta_\nu-\delta_\lambda}\right)}.$$

Jelöljük a ν-edik személy által jól megoldott feladatok számát r_ν-vel, és jelöljük a λ-adik feladatot jól megoldó személyek számát s_λ-val. P legutóbb kapott kifejezésének számlálójában a hatványkitevő a következőképpen írható:

$$\sum_{\nu=1}^{N}\sum_{\lambda=1}^{L}x_{\nu\lambda}(\beta_\nu-\delta_\lambda)=\sum_{\nu=1}^{N}\sum_{\lambda=1}^{L}x_{\nu\lambda}\beta_\nu-\sum_{\nu=1}^{N}\sum_{\lambda=1}^{L}x_{\nu\lambda}\delta_\lambda=\sum_{\nu=1}^{N}r_\nu\beta_\nu-\sum_{\lambda=1}^{L}s_\lambda\delta_\lambda .$$

Vagyis P-re a következőt kapjuk:

$$P=\frac{e^{\sum_{\nu=1}^{N}r_\nu\beta_\nu-\sum_{\lambda=1}^{L}s_\lambda\delta_\lambda}}{\prod_{\nu=1}^{N}\prod_{\lambda=1}^{L}\left(1+e^{\beta_\nu-\delta_\lambda}\right)}.$$

A képességfejlettségeknek és a feladatnehézségeknek olyanoknak kell lenniük, hogy ez a valószínűség maximális legyen. P csakis a β_ν és a δ_λ értékek függvénye ($x_{\nu\lambda}$ értékek paraméterek, adottak a mérés elvégzése után), így szélsőértéke ott lehet, ahol a P parciális deriváltjai eltűnnek. Technikailag nem is P szélsőértékét érdemes meghatározni, hanem a természetes alapú logaritmusáét. Ha az $\ln P$-nek szélsőértéke lesz, akkor P-nek is, ezt a logaritmusfüggvény szigorúan monoton növekvő volta biztosítja. Képezzük először ezt a logaritmust!

$$\ln P=\ln\frac{e^{\sum_{\nu=1}^{N}r_\nu\beta_\nu-\sum_{\lambda=1}^{L}s_\lambda\delta_\lambda}}{\prod_{\nu=1}^{N}\prod_{\lambda=1}^{L}\left(1+e^{\beta_\nu-\delta_\nu}\right)}=\sum_{\nu=1}^{N}r_\nu\beta_\nu-\sum_{\lambda=1}^{L}s_\lambda\delta_\lambda-\sum_{\nu=1}^{N}\sum_{\lambda=1}^{L}\ln\left(1+e^{\beta_\nu-\delta_\lambda}\right).$$

Írjuk fel a β_ν és a δ_λ szerinti parciális deriváltakat!

$$\frac{\partial \ln P}{\partial\beta_\nu}=r_\nu-\sum_{\lambda=1}^{L}\frac{e^{\beta_\nu-\delta_\lambda}}{1+e^{\beta_\nu-\delta_\lambda}}=r_\nu-\sum_{\lambda=1}^{L}p_{\nu\lambda}$$

$$\frac{\partial \ln P}{\partial \delta_\lambda} = \sum_{\nu=1}^{N} \frac{e^{\beta_\nu - \delta_\lambda}}{1 + e^{\beta_\nu - \delta_\lambda}} - s_\lambda = \sum_{\nu=1}^{N} p_{\nu\lambda} - s_\lambda$$

E parciális deriváltak a zérus értéket akkor veszik fel, amikor minden személyre vonatkozóan a jó feladatmegoldások valószínűségeinek összege egyenlő a jól megoldott feladatok számával, és minden feladatra vonatkozóan a jó megoldások valószínűségeinek összege egyenlő az adott feladatra érkezett jó megoldások számával.

Megoldásként azt kapjuk, hogy akkor lesz a konkrétan, a mérés során kapott megoldásmintázat a legnagyobb valószínűségű, ha a jó megoldások valószínűségei a következő kapcsolatban vannak a megoldásmintázattal: a jó megoldások valószínűségeit tartalmazó táblázatban mindegyik sorösszeg egyenlő kell, hogy legyen az adott személy által jól megoldott feladatok számával, és minden feladat oszlopában a valószínűségek összege pedig egyenlő kell legyen az adott oszlophoz tartozó feladatot jól megoldók számával.

Ez egy fontos összefüggés, ezért igyekszem részletesebben is illusztrálni a jó megértés érdekében. Egy konkrét, de fiktív példát használok. Legyen 5 feladat a tesztben, és vizsgáljunk 7 személyt. Megoldatjuk a feladatokat, és kiderül, hogy ki melyik feladatokban volt sikeres. A megoldástáblázatban 0-k és 1-esek szerepelnek, méghozzá egy adott sorban és adott oszlopban 1-es van akkor, ha a sor által reprezentált személy helyesen oldotta meg az oszlop által reprezentált feladatot, és 0 van ezen a helyen ellenkező esetben. Itt van a (fiktív) megoldástáblázat:

$$\begin{array}{l}
\begin{bmatrix}
1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0
\end{bmatrix}
\begin{array}{l}
\rightarrow r_1 = 4 \\
\rightarrow r_2 = 4 \\
\rightarrow r_3 = 3 \\
\rightarrow r_4 = 3 \\
\rightarrow r_5 = 2 \\
\rightarrow r_6 = 2 \\
\rightarrow r_7 = 1
\end{array} \\
\begin{array}{ccccc}
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
s_1 & s_2 & s_3 & s_4 & s_5 \\
\parallel & \parallel & \parallel & \parallel & \parallel \\
5 & 5 & 5 & 3 & 1
\end{array}
\end{array}$$

A táblázattól jobbra feltüntettem a sorösszegeket, a sorokban az egyesek számát, vagyis az adott személy által jól megoldott feladatok számát, ezeknek az összegeknek a jele r_ν, ($\nu = 1, \ldots, N$), jelen esetben r_1, r_2, r_3, r_4, r_5, r_6, r_7, És megadtam az oszlopösszegeket is, amelyek a feladatokat jól megoldók számát jelentik, ezek jele pedig s_λ, ($\lambda = 1, \ldots, L$), jelen esetben s_1, s_2, s_3, s_4, s_5.

Most nézzük a jó megoldások valószínűségtáblázatát! Ha összeadjuk a valószínűségtáblázat első sorának elemeit (az első személy jó feladatmegoldásaihoz tartozó valószínűségeket), akkor 4-et (az első sor sorösszegét, r_1-et) kell kapnunk eredményül. Vagyis az első személyhez, az ő jó feladatmegoldásaihoz tartozó valószínűségek összege 4 kell, hogy legyen (ha a valószínűségeknek azt a rendszerét akarjuk meghatározni, amely a feladatmegoldások során kapott megoldásmintázatra a legnagyobb valószínűséget szolgáltatja):

$$p_{1,1} + p_{1,2} + p_{1,3} + p_{1,4} + p_{1,5} = 4.$$

Ugyanígy felírhatunk egy a valószínűségek összegére vonatkozó egyenletet a második személyre vonatkozóan is, és így tovább, mind a 7 személyhez tartozik egy-egy ilyen egyenlet. De megtehetjük azt is, hogy az oszlopösszegeket számoljuk ki. Így például az első feladathoz tartozó valószínűségek összege 5 kell, hogy legyen (az első oszlopösszeg, s_1), ezért:

$$p_{1,1} + p_{2,1} + p_{3,1} + p_{4,1} + p_{5,1} + p_{6,1} + p_{7,1} = 5.$$

Ilyen egyenlőséget kifejező egyenletet az oszlopokra, vagyis a feladatokra vonatkozóan felírhatnánk ötöt (öt feladat van).

Ezek azok a követelmények, esetünkben konkrétan 7 + 5 = 12 összefüggés, amelyeket a valószínűségeknek teljesíteniük kell, miközben értékeik csak olyanok lehetnek, amelyek megfelelnek a *Rasch*-képletnek is. Utóbbi követelmény azt jelenti, ahogy már volt róla szó, hogy a valószínűségeket (esetünkben a 7×5 = 35 valószínűséget) kifejezhetjük a képességfejlettségek és a feladatnehézségek segítségével (esetünkben 7 + 5 = 12 érték). A modern tesztelméleti vizsgálódásoknak az itt leírt elvárások jelentik az alapját, abban az értelemben, hogy *a feladat mindig olyan valószínűségek, képességfejlettségek és feladatnehézségek keresése, amelyek kielégítik ezeket a feltételeket*. Itt csak egy speciális esetre fogalmaztuk meg a feltételeket (csak azt figyeljük, hogy az adott személy megoldotta vagy nem oldotta meg az adott feladatot), ám a gondolkodásmódot tekintve bármelyik modern tesztelméleti feladat hasonló módon kezelhető.

A képességfejlettségek (és a feladatnehézségek) számértékei

Ezután meg kell alkotnunk az ERR (az embereket képviselő absztrakt halmaz, amely a D relációval van „felszerelve") valós számokra (***Re***) való leképezését (s ne felejtsük el, hogy ezzel párhuzamosan ugyanezt a feladatokkal, a feladatnehézségekkel kapcsolatban is meg kell tennünk). Tudjuk, hogy ha az a, b, s, t elemek (emberek, képességfejlettségek) a D relációban állnak egymással, akkor az a-hoz rendelt és a b-hez rendelt számok különbsége nagyobb kell, hogy legyen, mint az s-hez és a t-hez rendelt számok különbsége. Képletekkel: Ha $abDst$, akkor $f(a) - f(b) > f(s) - f(t)$. Az $f(a)$, $f(b)$, $f(s)$, $f(t)$ értékek rendre a β_a, β_b, β_s, β_t képességfejlettség mértékek. Mivel itt intervallumskála létrehozásáról van szó, ezért elvileg két kiválasztott, egymással nem egyenlő képességfejlettségű ember képességfejlettségét szabadon megadhatnánk (ahogy a Celsius hőmérsékleti skála esetén az egyensúlyban lévő víz-jég keverék hőmérsékletét 0 °C-nak, a forrásban lévő víz hőmérsékletét 100 °C-nak tekintjük önkényesen). Válasszunk ki egy feladatot, aminek nehézségét 0-nak fogjuk tekinteni. Érdekes, hogy a Rasch-modell valójában ennél nagyobb szabadságot nem enged. Ugyanis ha most választunk egy tetszőleges személyt, és tudjuk, hogy ez a személy mekkora valószínűséggel oldja meg a 0 nehézségpraméterű feladatot, akkor a valószínűségre vonatkozó képlet alapján, δ helyébe 0-t helyettesítve, ki tudnánk számolni az adott személy képességfejlettségét. És ugyanígy járhatnánk el minden személlyel. Ez azt jelenti, hogy ha rögzítjük egyetlenegy feladat nehézségét, akkor a jó feladatmegoldások valószínűségének ismeretében már minden képességfejlettséget kiszámolhatnánk. De ha minden képességfejlettség adott, akkor egyben minden feladat nehézségparamétere is kiszámolható, hiszen a *Rasch* képletben mind a jó megoldás valószínűsége, mind a képességfejlettség ismert lenne[31]. Elméletileg a hozzárendelés tehát könnyen megvalósítható, a reprezentációs méréselmélet követelményei teljesíthetők.

A következő lépés: meg kell mutatnunk, hogy az elméleti modellben alkalmazott műveleteknek vannak megfelelőik a gyakorlati mérésben is. Láttuk az előző gondolatmenet során, hogy a kritikus kérdés, hogy meg tudjuk-e becsülni egy feladat helyes megoldásának valószínűségét. Ki kell jelölnünk egy feladatot, aminek a nehézégét 0-nak tekintjük. Ezután megnézzük, hogy az egyes emberek mekkora valószínűséggel képesek ezt a feladatot megoldani (majd később a többit is megvizsgáljuk). Ha e valószínűségek becslése empirikusan megoldható, akkor nyert ügyünk van.

A valószínűségek empirikus becslése relatív gyakoriságok meghatározásával történik. Ez az eljárás azonban nem mindig lehetséges, illetve különböző egyéb megfontolásokra lehet szükségünk. Itt is át kell gondolnunk alaposan, hogy milyen esemény relatív gyakoriságát vizsgáljuk. Egyszerű lenne a dolgunk, ha a feladatmegoldás mindig újra és újra elvégezhető lenne, mindig ugyanolyan feltételek között. Ekkor egy hosszú kísérletsorozatban, aminek minden egyes eleme ugyanannak a feladatnak a megoldása lenne, kellene megvizsgálnunk, hogy hányszor volt sikeres a feladatmegoldás, e számot elosztani az összes kísérlet számával, és ez a szám becsülné a helyes megoldás valószínűségét. Lehet, hogy van olyan feladat, amely esetén akár ez is lehet a helyzet (olyan tevékenységgel kapcsolatos, amelyben a tanulás nem játszik szerepet, vagy a tanulás eredménye viszonylag gyorsan elenyészik)[32], azonban a pedagógiát érdeklő, az iskolában adott feladatokhoz hasonlók nagy többsége nem ilyen, éppen azért, mert az iskolai tevékenységek pontosan a tanulást szolgálják. Esetünkben tehát nem megoldás az ugyanazon feladatok többszöri megoldása.

A *Rasch-modell*ben a helyes feladatmegoldás csak a képességfejlettségtől és a feladatnehézségtől függ. Ezért a helyes feladatmegoldás valószínűsége becsülhető több ugyanolyan képességfejlettséggel (β) rendelkező személy feladatmegoldásainak segítségével. Tegyük fel tehát, hogy nem egy, hanem több, lehetőleg minél több ugyanolyan képességfejlettségű személlyel oldatjuk meg a kiválasztott, 0 nehézségparaméterű feladatot. Honnan tudjuk, hogy a személyek ugyanolyan képességparaméterrel rendelkeznek? Onnan, hogy ha megvizsgáljuk, hogy a feladatok egy nagyobb csoportjával hogyan boldogulnak, akkor kiderül, hogy ugyanolyan arányban képesek azokat helyesen megoldani. Ha most megnézzük, hogy ezek a személyek milyen arányban oldották meg a kiválasztott, 0 nehézségű feladatot, akkor becslést kapunk arra, hogy a kiválasztott, azonos képességfejlettséggel, vagyis azonos β értékkel rendelkező személyek mekkora valószínűséggel oldják meg a 0 nehézségparaméterű feladatot. Ha ez megvan, akkor csak be kell helyettesítenünk a *Rasch*-képletben a p helyére ezt az értéket, a δ helyébe a 0-t, és kiszámíthatjuk β-t. Valahogy így:

[31] Felmerülhetne, hogy akkor itt valójában nem is intervallumváltozókról van szó. Hiszen ha a képességfejlettségek értékei intervallumváltozóhoz tartoznának, akkor a $\beta' = a\beta + b$ képlet szerint kellene transzformálódniuk a „legitim" képsségfejlettségeknek, márpedig itt csak az engedhető meg, hogy $\beta' = \beta + b$ legyen. És valóban, abból következően, hogy a képességfejlettségek és a feladatnehézségek definiálása ugyanazzal a meghatározással, ugyanazzal a képlettel történt, a lehetséges skálák szűkebb kört alkotnak, mint ami más esetekben az intervallumskálák esetén adódik. De ez semmi gondot nem jelent, mert az olyan skálák, amelyek esetében a $\beta' = \beta + b$ transzformációs szabály érvényesül, a lehetséges intervallumskálák egy speciális részhalmazát alkotják, azt, amelybe tartozó skálák esetén $a = 1$. Ezek is intervallumskálák, csak a lehetséges skálák között a számításainkban nem szerepeltethető minden skála.

[32] Csak ötletelek: egy kézügyességet igénylő feladat esetén előállhat ez a helyzet, bár még egy ilyen esetben sem zárhatjuk ki teljesen a tanulás lehetőségét.

$$p=\frac{e^{\beta-\delta}}{1+e^{\beta-\delta}}=\frac{e^{\beta}}{1+e^{\beta}}=\frac{1}{e^{-\beta}+1}$$

$$e^{-\beta}+1=\frac{1}{p}$$

$$e^{-\beta}=\frac{1}{p}-1=\frac{1-p}{p}$$

$$e^{\beta}=\frac{p}{1-p}$$

$$\beta=\ln\frac{p}{1-p}$$

Elvileg a mérés elvégezhető lenne ezen a módon. Vagyis azt kellene megvizsgálnunk, hogy az ugyanolyan képességfejlettséggel rendelkezők milyen arányban oldják meg a 0 nehézségű feladatot, és akkor az ő képességfejlettségük ebből már kiszámolható lenne. Ezen a módon azonban nem tudnánk megmérni mindenkinek a képességfejlettségét, mert a „széleken”, vagyis a jó és a gyenge feladatmegoldók körében nem találunk sok személyt, akik azonos képességfejlettséggel rendelkeznének. Ha van egy 20 feladatból álló teszt, akkor lehetnek olyanok, akik 18, 19 feladatot oldanak meg, de olyanok is, akik 1-et, 2-t. Ők azonban valószínűleg nagyon kevesen vannak, ezért a 0 nehézségű feladat megoldásában mutatott teljesítmény felhasználása érdekében nem tudunk kiválasztani az övékével megegyező képességfejlettséggel rendelkezőket nagy számban. Pl. ha 2 feladatot egy, vagy kettő, vagy három személy tudott csak megoldani, akkor az ő vizsgálatuk a 0 nehézségű feladattal nagyon gyenge lábakon állna, gyakorlatilag lehetetlen lenne.

Ám ez a módszer mégsem teljesen elvetendő. Ugyanis megtehetjük, hogy megvizsgáljuk annak a csoportnak a feladatmegoldásait, amely csoportban a legtöbben vannak, ha a csoportokat úgy képezzük, hogy azok tartozzanak össze, akik azonos számú feladatot voltak képesek megoldani. Nézzünk is egy konkrét példát: mondjuk az említett, 20 feladatot tartalmazó tesztben a legtöbben azok vannak, akik 9 feladatot tudtak megoldani helyesen, legyen a létszámuk 682 (a teljes minta ennek megfelelően igencsak nagy). Azt feltételezzük, hogy az ő képességfejlettségük azonos nagyságú (β). Megnézzük, hogy hányan oldották meg a 0 nehézségű feladatot, mondjuk 218-an voltak 682-ből. Ez azt jelenti, hogy az azonos, egyelőre nem ismert képességfejlettségű személyek a 0 nehézségű feladatot közelítőleg 218/682 ≈ 0,320 valószínűséggel tudják megoldani. Helyettesítsük be ezt az értéket a fenti levezetésben kapott végső képletbe:

$$\beta=\ln\frac{1-p}{p}=\ln\frac{1-0{,}32}{0{,}32}=\ln 2{,}126=0{,}754$$

Vagyis a 9 feladatot helyesen megoldók 682 fős mintájából a 0 nehézségű feladatot jól megoldóknak, 218 főnek már meg is határoztuk a képességfejlettségét, ez 0,754-nek adódott (három tizedes jegy pontossággal számolva). Viszont ha már ismerjük 218 fő képességfejlettségét, akkor akármelyik feladat nehézségére adhatunk becslést. Ugyanis kiszámíthatjuk, hogy ez a csoport milyen relatív gyakorisággal oldotta meg a feladatot. Ez becsli annak a valószínűségét, hogy az előbb meghatározott képességfejlettséggel rendelkező személyekből álló csoport tagjai jól oldják meg a feladatot. Ha pedig ez a becslés rendelkezésünkre áll, akkor a következő számítás segít (β a csoportba tartozók képességfejlettségének mértéke, példánkban 0,754, δ az éppen vizsgált feladat nehézségparamétere, egyelőre ismeretlen, ezt szeretnénk meghatározni):

$$p = \frac{e^{\beta-\delta}}{1+e^{\beta-\delta}} = \frac{1}{e^{\delta-\beta}+1}$$

$$e^{\delta-\beta}+1 = \frac{1}{p}$$

$$e^{\delta-\beta} = \frac{1}{p} - 1 = \frac{1-p}{p}$$

$$\delta - \beta = \ln\frac{1-p}{p}$$

$$\delta = \ln\frac{1-p}{p} + \beta$$

Ismét illusztráljuk az eljárást a példánkkal! Tegyük fel, hogy a vizsgált feladatot a 218 személy közül 38 tudta megoldani (viszonylag nehéz feladat). A relatív gyakoriság: $38/218 \approx 0{,}174$. Ez a valószínűség becslése, vagyis a képletünkben a p. A β értékét az előbbiekben már meghatároztuk, ez 0,754 volt. Helyettesítsük be ezeket az értékeket az előbbi levezetés során kapott képletbe:

$$\delta = \ln((1-p)/p) + \beta \approx \ln((1-0{,}174)/0{,}174) + 0{,}754 \approx 2{,}312.$$

Ilyen módszerrel bármelyik feladatnehézséget meghatározhatjuk, kivéve, ha a feladatot senki nem oldotta meg, vagy mindenki megoldotta. Az utóbbi esetben ugyanis 1 a relatív gyakoriság, vagyis a p becslése, s így 0-nak kellene meghatározni a logaritmusát, ami nem létezik. A másik esetben viszont a képletben a 0 relatív gyakoriság miatt 0 lenne a nevezőben. A modern tesztelméletnek megfelelő számításokban eleve nem vesszük figyelembe azokat a feladatokat, amelyeket mindenki megoldott, vagy amelyeket senki sem, illetve azokat a személyeket sem vesszük figyelembe, akik minden feladatot megoldottak, vagy egyet sem[33].

Hogyan történik a képességfejlettségek meghatározása (PROX)?

Az előbbi példában meglehetősen nagy minta adatainak rendelkezésre állását feltételeztem (682 volt a létszáma a mintán belül azon személyek csoportjának, akik a 20 feladatból éppen 9-et tudtak helyesen megoldani), ami nem tekinthető túlságosan gyakorinak a pedagógiai, pszichológiai vizsgálatokban. Ha egy kutató éppen egy ilyen „szerencsés” vizsgálatot végez, akkor működhet az itt leírt eljárás, azonban rendkívül megbízhatatlanok az eredmények. Szükség volt olyan eljárások kidolgozására, amelyek már kisebb minták, és a megoldásszámok (vagy a képességfejlettségek) szórtabb volta esetén is eredményt produkálnak. Ez megtörtént, és ma több olyan algoritmust ismerünk, amely becslést szolgáltat a képességfejlettségekre és a feladatnehézségekre, s ezek a becslések többféle matematikai szempontból is jó tulajdonságokkal rendelkeznek. Ezen algoritmusok közül egyet, a lehető legegyszerűbbet, az ún. PROX eljárást meg is mutatom. „Elhallgatom” a matematikai szempontból nehezebben érthető részek magyarázatát. Az

---

[33] Igazság szerint vannak közelítő számítások arra, hogy a számításokból ne kelljen kihagyni sem a szélsőséges megoldásokat produkálókat, sem az extrém adatokkal bíró feladatokat sem.

algoritmus így is követhető, és a Microsoft Excel program segítségével a számításokat bárki maga is elvégezheti.

A bemutatandó algoritmus a képességfejlettségek és a feladatnehézségek becslésére tehát a PROX eljárás (Wright és Douglas 1977). Valójában számítógép nélkül, számológéppel, vagy akár kézzel, egy függvénytáblázat segítségével is elvégezhetők a számítások. Minden becslési eljárás kiindulópontja a vizsgált személyek feladatmegoldásai sikerességéről informáló táblázat (megoldástáblázat), amely annyi sorból áll, ahány személy kísérletezett a feladatok megoldásával, és annyi oszlopból, ahány feladat van a tesztben. A táblázat adott sorában és adott oszlopában lévő helyen 0 vagy 1 áll, attól függően, hogy a sorhoz rendelt személy nem oldotta meg vagy megoldotta az oszlophoz rendelt feladatot. A modern tesztelméleti számításokban (legalábbis az itt tárgyalt legegyszerűbb esetben) e táblázat tartalmazza az összes információt, ami rendelkezésünkre áll. Az eredmények könnyebb olvashatósága érdekében érdemes úgy sorrendbe állítani a vizsgált személyeket, hogy azoknak az adatai kerüljenek „magasabbra" (kisebb sorszámú sorba), akik több feladatot tudtak megoldani, illetve úgy érdemes sorba rendezni a feladatokat, hogy azoknak a feladatoknak megfelelő oszlopok kerüljenek inkább balra, amely feladatokat többen oldották meg. Ez nem kötelező, csak ajánlott, mert így a képességfejlettség- és feladatnehézség becslések könnyebben átláthatók lesznek.

A PROX eljárásban fontos szerep jut az előbb bemutatott táblázat soraiban, illetve oszlopaiban szereplő számok összegének. Tekintve, hogy csak 0-k és 1-esek vannak a táblázatban, egy sorban (és egy oszlopban) a számok összege valójában az abban a sorban (oszlopban) lévő egyesek számával azonos. Vagyis egy sorösszeg nem más, mint az adott személy által jól megoldott feladatok száma, egy oszlopösszeg pedig az adott feladatot jól megoldó személyek száma. N-nel jelöljük a vizsgált személyek számát, L-lel a tesztben szereplő feladatok számát, r_ν jelöli a ν-edik sorösszeget, tehát a ν-edik személy által jól megoldott feladatok számát, és s_λ jelöli a λ-adik oszlopösszeget, vagyis a λ-adik feladatot jól megoldók számát.

Vegyük elő ismét a korábban már használt fiktív példát! 5 feladat és 7 vizsgált személy szerepel a példában, amely esetében egyik oszlopban, és egyik sorban sem szerepel csupa azonos szám:

$N = 7, L = 5$

$$\begin{array}{c} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{array}{l} \to r_1 = 4 \\ \to r_2 = 4 \\ \to r_3 = 3 \\ \to r_4 = 3 \\ \to r_5 = 2 \\ \to r_6 = 2 \\ \to r_7 = 1 \end{array} \\ \begin{array}{ccccc} \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ s_1 & s_2 & s_3 & s_4 & s_5 \\ \| & \| & \| & \| & \| \\ 5 & 5 & 5 & 3 & 1 \end{array} \end{array}$$

E kiinduló adatoknak a meghatározása könnyű feladat, s aki tud bánni az Excel programmal, vagy bármilyen más táblázatkezelővel, könnyen megoldhatja a feladatot számítógéppel is. Adottak tehát a sor és az oszlopösszegek. A PROX eljárás lépései ezután a következők (Wright és Stone 1979):

$$x_\lambda = \ln \frac{N - s_\lambda}{s_\lambda}$$

Az ln jel a természetes alapú logaritmus számítását jelenti. Az érték a λ-adik feladatra vonatkozik, természetesen mindegyik feladatra meg kell határozni (λ = 1, 2, …, L).

$$y_\nu = \ln \frac{r_\nu}{L - r_\nu}$$

Az előbbihez hasonló képlet, de ez a vizsgált személyekre vonatkozó értékek kiszámítására használatos (ν = 1, 2, …, N).

$$\xi = \frac{\sum_{\lambda=1}^{L} x_\lambda}{L}$$

A feladatok számának (L) megfelelő számú x érték átlagának a kiszámítása. Excel program használata során csak az ÁTLAG(<az x-eket tartalmazó tömb>) képlet beírását igényli.

$$\eta = \frac{\sum_{\nu=1}^{N} y_\nu}{N}$$

Hasonlóan az előzőhöz: az y értékek átlaga.

$$\sigma_x^2 = \frac{\sum_{\lambda=1}^{L} x_\lambda^2 - L\bar{x}^2}{L-1}$$

Ezzel a képlettel az x értékek szórásnégyzetét számítjuk ki. Excelben: SZÓRÁS(<az x-eket tartalmazó tömb>)^2.

$$\sigma_y^2 = \frac{\sum_{\nu=1}^{N} y_\nu^2 - N\bar{y}^2}{N-1}$$

Ugyanúgy az y értékek szórásnégyzete.

$$X = \sqrt{\frac{1 + \frac{\sigma_x^2}{2{,}83}}{1 - \frac{\sigma_x^2 \sigma_y^2}{8.35}}}$$

Ez a képlet a képességfejlettség becslések kiszámítása során felhasználásra kerülő szorzótényező meghatározására szolgál.

$$Y = \sqrt{\frac{1 + \frac{\sigma_y^2}{2{,}83}}{1 - \frac{\sigma_x^2 \sigma_y^2}{8.35}}}$$

Szerepe ugyanaz, mint az X-é, de a feladatnehézség becslések számításához szükséges.

$$d_\lambda = Y(x_\lambda - \xi)$$

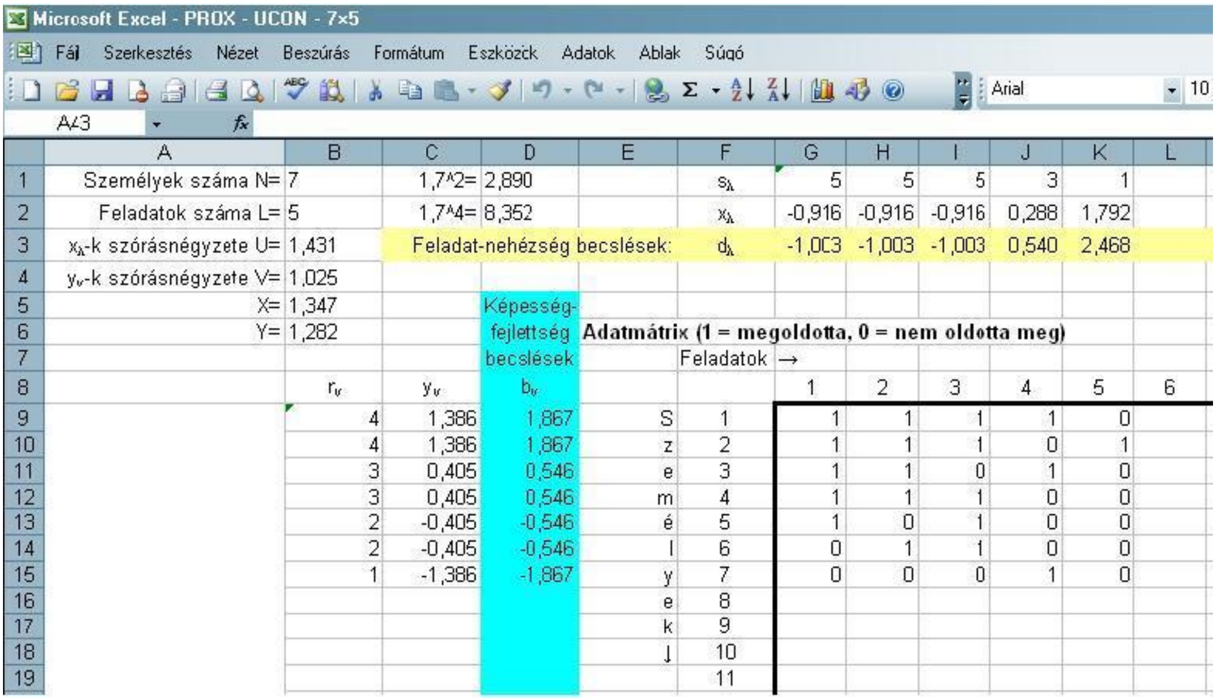
Végeredmény: ezek a feladatnehézség becslések.

$$b_\nu = Xy_\nu$$

Végeredmény: ezek a képességfejlettség becslések.

Mint látjuk, valójában elemi szintű számításokat kell végezni, még akkor is, ha sokak számára akár rémisztőek is lehetnek a képletek. Hogy miért éppen így érdemes meghatározni a becsléseket, annak mélyebb matematikai magyarázata van, és azon alapul, amit korábban már részleteztem: a képességfejlettségeknek és a feladatnehézségeknek a jó megoldásokhoz tartozó valószínűségek olyan rendszerét kell meghatározniuk, amely valószínűségek mellett a teszt feladatainak megoldása során kapott megoldásmintázat a legvalószínűbb. A matematikában az ilyen becslések egyike a *maximum likelihood* becslés (van másmilyen is). A bemutatott becslés a matematikai statisztikában leírt, a becslésekkel szemben támasztott igényeket kielégíti (Wright és Stone 1979).

Bemutatom azt az Excel táblázat részletet, amely a számításokat tartalmazza (5. ábra).



Microsoft Excel - PROX - UCON - 7x5

Fájl Szerkesztés Nézet Beszúrás Formátum Eszközök Adatok Ablak Súgó

	A	B	C	D	E	F	G	H	I	J	K	L
1	Személyek száma N=	7	1,7^2=	2,890		s_λ	5	5	5	3	1	
2	Feladatok száma L=	5	1,7^4=	8,352		x_λ	-0,916	-0,916	-0,916	0,288	1,792	
3	x_λ-k szórásnégyzete U=	1,431	Feladat-nehézség becslések:			d_λ	-1,003	-1,003	-1,003	0,540	2,468	
4	y_ν-k szórásnégyzete V=	1,025										
5	X=	1,347		Képesség-								
6	Y=	1,282		fejlettség	Adatmátrix (1 = megoldotta, 0 = nem oldotta meg)							
7				becslések		Feladatok →						
8		r_ν	y_ν	b_ν			1	2	3	4	5	6
9		4	1,386	1,867	S	1	1	1	1	1	0	
10		4	1,386	1,867	z	2	1	1	1	0	1	
11		3	0,405	0,546	e	3	1	1	0	1	0	
12		3	0,405	0,546	m	4	1	1	1	0	0	
13		2	-0,405	-0,546	é	5	1	0	1	0	0	
14		2	-0,405	-0,546	l	6	0	1	1	0	0	
15		1	-1,386	-1,867	y	7	0	0	0	1	0	
16					e	8						
17					k	9						
18					↓	10						
19						11						

5. ábra: A PROX eljárás EXCEL munkalapjának képe 7 személy által 5 feladat megoldásában mutatott teljesítményéhez tartozó (fiktív) adatokkal

A megoldástáblázat a G9:K15 tömbben található, megegyezik az itt korábban bemutatottal. A sorösszegek a B9:B15 tömbben, az oszlopösszegek a G1:K1 tömbben találhatók. Az x_λ értékeket a G2:K2 tömbben, az y_ν-ket a C9:C15 tömbben találjuk (kiszámításukra fentebb adtam meg a megfelelő képleteket). A szórásnégyzetek, valamint az *X* és *Y* értékek pedig sorban a B3, B4, B5, B6 cellákban vannak. Végül a feladatnehézség becsléseket a G3:K3 tömbben, a képességfejlettségeket a D9:D15 tömbben találjuk. Itt minden nem egész adatot három tizedessel adtam meg.

A valóságos mérési folyamatokban természetesen általában több feladat, és sokkal több vizsgált személy van. A fenti Excel táblázatot (5. ábra) nem nehéz elkészíteni, és ugyancsak könnyű feladat úgy módosítani, hogy több feladattal és több vizsgált személlyel működjék. Ehhez

- módosítani kell a B1 és a B2 cellákat (a feladatok és a vizsgált személyek száma),
- be kell írni a G9 cellától kezdődően jobbra és lefelé a megoldástáblázat adatait,
- a G1 cellában úgy kell módosítani a képletet, hogy a G oszlopban (a megoldástáblázatban) szereplő összes, G oszlopbeli adatot adja össze, majd ezt jobbra az összes, tehát a feladatok számának megfelelő számú cellára ki kell terjeszteni (jobbra „húzni”),

- a B9 cellában is módosítani kell a sorösszegre vonatkozó képletet, adja össze a megoldástáblázat 9. sorában lévő összes adatot, majd ezt a képletet ki kell terjeszteni a 9. sor alatt annyi sorra, ahány személyt vizsgálunk („le kell húzni"),
- a G2 cellában lévő képleten nem kell változtatni, viszont ezt is a feladatok számának megfelelő számú cellára ki kell terjeszteni a G2-től jobbra,
- hasonlóan a C9 cella képlete sem módosítandó, viszont „húzzuk le" annyi cellával, ahány vizsgált személy van,
- a G3 cellában lévő képletet úgy kell módosítani, hogy benne az átlagszámításnál a G2 cella és a mellette lévő összes, adatot tartalmazó cella adatainak átlagát számítsa, és ezt is természetesen jobbra kell „húzni",
- a B3 (és a B4) cellában lévő képleteket úgy kell módosítani, hogy a konkrét feladatban szereplő adatok szórásnégyzetét számolja.

Ha valaki mindezeket a rövid idő alatt végrehajtható változtatásokat elvégzi, akkor lesz egy a konkrét feladatához illesztett, a *Rasch-modell*nek megfelelő, a PROX eljárást alkalmazó Excel táblázata, vagyis modern tesztelméleti adatokkal fog rendelkezni. Eredményei a feladatnehézségek és a képességfejlettségek becslései lesznek a 3. sorban és a D oszlopban.

A PROX eljárás egyszerűsége ellenére is tűrhető becsléseket szolgáltat. Hogy mitől jó egy becslés, az egyrészt matematikai statisztikai szakkérdés, és ez is egy olyan részlet, amelynek bemutatására itt nem vállalkozom. A becslés jóságának érzékeltetése azonban másképpen is lehetséges, ezt mutatom meg a továbbiakban.

Az Interneten több, az itt bemutatott, sok önálló tevékenységet igénylő eljáráshoz képest sokkal több funkciót ellátó, sokkal átfogóbb feladatmegoldásra képes szoftvereket lehet megvásárolni, ügyesebbek korlátozott tudású ingyenes változatokat is letölthetnek.

Az eredmények használhatósága

Mire jutottunk? A modern tesztelmélet rendszerében, azon belül is az IRT módszerek (IRT = Item Response Theory = itemválasz elmélet) között a legegyszerűbb modell, a *Rasch-modell* (más megfogalmazásokban az egyparaméteres *Rasch-modell*) esetén, amennyiben a mérési feladatra valóban alkalmazható a modell, a reprezentációs méréselméletnek is megfelelő, korrekt, kivitelezhető megoldásra jutunk a feladatnehézség és képességfejlettség paraméterek meghatározásához. Már csak azt a „kis kérdést" kell megválaszolnunk, hogy vajon konkrét adataink egy konkrét mérésben mennyire támasztják alá a feltevést, hogy a *Rasch-modell* valóban alkalmazható.

A feltétel tartalmát alapvetően a Rasch-képlet határozza meg. Jelöljünk ki egy feladatot, aminek adott a nehézsége (δ). Korábbi megfontolásainkból következik, hogy a Rasch-képlet szerint a nagyobb képességfejlettségű személyeknek ezt a feladatot nagyobb valószínűséggel kell helyesen megoldaniuk. Mert a Rasch-képletből ugyanakkora δ-val számolva a nagyobb β nagyobb p-t eredményez, ez a $p(\beta)$ függvény szigorúan monoton növekedése miatt is így kell, hogy legyen. Az ugyanakkora képességfejlettségű személyek pedig ugyanakkora valószínűséggel tudják jól megoldani az adott feladatot. De egy adott képességfejlettséggel (β) rendelkező személy a nehezebb feladatot kisebb valószínűséggel tudja megoldani, ez is kikövetkeztethető a Rasch-képletből. Mint ahogy az is, hogy ugyanolyan nehézségű feladatokat egy adott személy ugyanakkora valószínűséggel tud megoldani helyesen, s ez igaz minden vizsgált személyre. Sokan már e megfontolások alapján is látják nyilván, hogy baj van, de hogy ez még inkább átlátható legyen, nézzük meg kicsit részletesebben is, miként érvényesülhetnek a Rasch modell alkalmazásának feltételei!

A Rasch-modell érvényességét nyilván lehetne ellenőrizni úgy, hogy a vizsgált személyekkel megoldatjuk ugyanazokat a feladatokat sokszor, mindig ugyanolyan előfeltételeket biztosítva a feladatmegoldásokhoz. Ezt a gondolatkísérletet vázoltam fel már a klasszikus tesztelmélet bemutatása során is. A feladat, mint már többször volt róla szó, nem kivitelezhető, valóban csak gondolatkísérlet. De ha egy pillanatra mégis kivitelezhetőnek tartjuk, akkor nyilván ki kellene számolni, hogy egy hosszú kísérletsorozatban milyen relatív gyakoriságokat kapunk az egyes feladatmegoldásokra, és utána már csak „némi" egyenletrendszer megoldás kérdése, hogy megmutassuk, a kapott relatív gyakoriságokat tekintve a valószínűségeknek, van-e a képességfejlettségeknek és a feladatnehézségeknek egy olyan rendszere, amely szerinti értékeket a Rasch képletbe helyettesítve a kiszámított valószínűségeket kapjuk. Természetesen, ha kivitelezhető lenne is mindez, akkor is minden esetben (nagyon-nagyon véletlen eseteket leszámítva) azt kapnánk, hogy nincs ilyen rendszere a paramétereknek, a valószínűségeket kifejező számok rendszerének nagyon speciálisnak kell lenni ahhoz, hogy az összefüggések mind érvényesüljenek.

Illusztrációként gondoljunk bele a következőbe! Legyen csak két vizsgált személy, és legyen csak két feladat. Ez összesen csak négy eseményt jelent. Legyenek a feladatok jó megoldásának valószínűségei p_{11}, p_{12}, p_{21}, p_{22}, az első index utaljon a vizsgált személyre, a második a feladatra. Így például p_{21} annak a valószínűsége, hogy a 2. személy jól oldja meg az 1. feladatot. A felsorolt p értékek legyenek a becsült valószínűségek, tehát a leírt gondolatkísérletben kiszámolható relatív gyakoriságok. Vannak-e vajon olyan β_1, β_2 képességfejlettségek, és δ_1, δ_2 feladatnehézségek, hogy azokat a Rasch-képletbe helyettesítve érvényesek lesznek az egyenlőségek? A következőknek kell érvényesülniük tehát:

$$p_{11} = \frac{1}{e^{\delta_1 - \beta_1} + 1}; \quad p_{12} = \frac{1}{e^{\delta_2 - \beta_1} + 1}; \quad p_{21} = \frac{1}{e^{\delta_1 - \beta_2} + 1}; \quad p_{22} = \frac{1}{e^{\delta_2 - \beta_2} + 1}.$$

Mind a négy esetben átalakíthatók az egyenletek a következő, korábban már bemutatott séma segítségével:

$$p = \frac{1}{e^{\delta - \beta} + 1}$$

$$e^{\delta - \beta} + 1 = \frac{1}{p}$$

$$e^{\delta - \beta} = \frac{1}{p} - 1 = \frac{1 - p}{p}$$

$$\delta - \beta = \ln \frac{1 - p}{p}.$$

Vagy ami ugyanaz (–1-gyel beszorozva mindkét oldalt, és alkalmazva a logaritmus függvényre vonatkozó szabályokat):

$$\beta - \delta = \ln \frac{p}{1 - p}.$$

A kapott összefüggés jobboldalán álló logaritmus kifejezést jelöljük a továbbiakban t-vel, természetesen ellátva a megfelelő indexekkel. Mivel a p értékeket, a valószínűségek becslését a gondolatkísérletben ismerjük, ezért a t értékek is kiszámíthatók. A következő egyenletrendszert kapjuk:

$$\beta_1 - \delta_1 = t_{11}$$
$$\beta_2 - \delta_1 = t_{12}$$
$$\beta_1 - \delta_2 = t_{21}$$
$$\beta_2 - \delta_2 = t_{22}$$

Ebben az egyenletrendszerben van négy ismeretlen (β_1, β_2, δ_1 és δ_2), van négy egyenlet, vagyis van remény a megoldásra. És most jön a lényeg! Ha összeadjuk az első és az utolsó egyenletet, s ebből kivonjuk a két középsőt, akkor a baloldalon 0-t kapunk. Ha e műveletek eredményeként a jobboldalon nem 0 jön ki, akkor ennek az egyenletrendszernek nincs megoldása (hiszen az elvégzett műveletekkel azt kapjuk, hogy $0 = a \neq 0$, ami ellentmondás). Vagyis a t értékeknek olyanoknak kell lenniük, hogy $t_{11} + t_{22} - t_{12} - t_{21} = 0$ legyen. Márpedig erre semmi garancia nincs. Lehetne, ha lenne olyan elmélet, amely empirikusan kellően alátámasztott, és azt mondaná, hogy a valószínűségek alakulása törvényszerűen olyan, hogy a belőlük számítható $t_{11} + t_{22} - t_{12} - t_{21}$ érték szükségképpen 0. De ilyen elméletünk nincs, semmi sem garantálja, hogy ez így legyen. Vagyis csak annyit mondhatunk, hogy a *Rasch-modell* segítségével csak olyan eseteket vizsgálhatunk, amelyekben ez valamiért mégis így van. De hogy miért, és hogy melyek pontosan ezek az esetek, azzal kapcsolatban nem tudunk mit mondani.

Most vegyük figyelembe, hogy a

$$\beta - \delta = \ln \frac{p}{1-p}$$

mennyiség is annak a valószínűségét jellemzi, hogy a β képességfejlettségű személy jól oldja meg a δ nehézségű feladatot. Szerepe tehát ugyanaz, mint a p valószínűségé, csak míg az utóbbi egy 0 és 1 közötti számmal jellemzi egy esemény bekövetkezésének esélyét, az $\ln(p/(1-p))$ a mínusz végtelen és végtelen között. Vagyis a fent bevezetett t értékek az egyes feladatmegoldásokkal kapcsolatos események valószínűségét jellemzik ezen a másik módon. Most képzeljük el, hogy a két feladat különböző jellegű („kétdimenziós a jelenség"). Az elsőt az első személy tudja nagyobb valószínűséggel megoldani, míg azt a másik személy csak kis valószínűséggel. A másik feladattal kapcsolatban éppen fordított a helyzet, azt a második személy tudja nagyobb valószínűséggel megoldani. Kis gondolkodás után könnyen belátható, hogy így a $t_{11} + t_{22} - t_{12} - t_{21} = 0$ feltétel nem teljesülhet. t_{11} és t_{22} nagy pozitív számok, t_{12} és t_{21} kis számok, vagy akár negatívak is lehetnek. Az előbbi kettő összegéből kivonva e két utóbbit, nem kaphatunk 0-t.

Vagyis legalább két dimenzió jelenléte esetén nagyon könnyen adódhat olyan szituáció, amelyben már két személyhez (illetve már két feladathoz) sem tudunk megfelelő paramétereket rendelni, ilyenek matematikailag nem létezhetnek.

Ha növeljük a vizsgált személyek és a feladatok számát, akkor az egyenletrendszer több egyenletből áll majd, de a lényeg nem változik. Mindig kiválaszthatunk két személyt és két feladatot, és ezekre a részhalmazokra felírhatjuk az egyenleteket. Ha van az egész egyenletrendszert kielégítő rendszere a képességfejlettségeknek és a feladatnehézségeknek, akkor a két személyre és a két feladatra felírt egyenletekből álló rész-egyenletrendszernek is van megoldása, vagyis az elemzés ugyanúgy elvégezhető.

Most ráadásul még annak is teljesülnie kell, hogy ha egy adott párosban, véve egy adott feladatpárt, kiszámítható volt valakinek a képességfejlettsége (olyan speciálisak a valószínűségek, hogy lesz megoldása a négy egyenletből álló egyenletrendszernek), akkor kiválasztva társnak egy másik személyt, és véve ugyanazokat a feladatokat, az új egyenletrendszernek is lesz megoldása, de az természetesen ugyanaz kell legyen a kiválasztott személy képességfejlettségére vonatkozóan, mint amit az előbb meghatározhattunk. És erre semmilyen elvi garancia nem létezik.

Amit most megmutattam, az első megközelítésben fölösleges elvieskedésnek tűnik, hiszen nincs módunk rá, hogy a gondolatkísérletet valódi kísérletté tegyük, nincs módunk rá, hogy megbízható valószínűségbecsléseket állítsunk elő. A gondolatmenet mégis mutat valamit. Azt, hogy a *Rasch-modell*, de igaz ez általában minden IRT modellre is, olyan hipotetikus szituációkra érvényes, amelyek vagy léteznek, vagy nem. Lehet, hogy a modern tesztelmélet alkalmazásával kapcsolatos eddigi törekvések mind hiábavalók voltak, mert egyáltalán nincs olyan jelenség, ami jól leírható lenne ezzel az elmélettel.

Minden tudományos modellnek, pontosabban az alkalmazásuknak vannak ésszerű korlátai. Bizonyos esetekre már nem érdemes alkalmazni egy adott modellt, mert már nem adaptív. Szolgáltat ugyan eredményt, működik az eljárás matematikája (bár még ez sem mindig igaz), de a modell explicit, vagy hallgatólagos feltételei nem vagy csak nagyon kevéssé teljesülnek, így az eredmény is teljességgel megbízhatatlan. Vagyis nem azt mondom, hogy a modern tesztelmélet hibás, erről természetesen szó sincs. Csak annyit mondok, hogy a modern tesztelmélet alkalmazása addig jogos, amíg érvényesek a mögötte álló feltételezések. Márpedig, úgy tűnik, a komplex képességek esetén ezek nem érvényesek.

A modern tesztelmélet modelljei (köztük a legegyszerűbb, az itt is tárgyalt Rasch-modell) a pszichológiában, neveléstudományban is alkalmazható modellek egy nagyon tág körébe tartoznak. E modellek klasszifikációja sokféleképpen lehetséges, azonban már viszonylag korán, a 20. század 80-as éveinek közepén született átfogó leírás e klasszifikációról, amely matematikai szempontból mélyen megalapozott, és az azóta bekövetkezett számottevő fejlődés ellenére is máig maradandó értékekkel bír. Ez *Paul Holland* és *Paul Rosenbaum* munkája (1986). *Holland* és *Rosenbaum* megadják az ún. látens változó modellek (látens változókat tartalmazó modellek) általános leírását, majd elkülönítik típusaikat. A látens változók, mint már volt róla szó, közvetlenül nem megfigyelhető értékekkel bíró változók, amelyek valamilyen kapcsolatban vannak a megfigyelhető értékekkel. Természetesen a Rasch-modell is ide tartozik, hiszen a képességfejlettség és a feladatnehézség látens változóiról van szó, valamint a feladatmegoldások sikerét vagy sikertelenségét leíró, megfigyelhető változókról[34]. Holland és Rosenbaum két alapvető tulajdonságát írják le az ilyen modelleknek, olyan tulajdonságokat, amelyekkel vagy rendelkeznek, vagy nem rendelkeznek, de amelyek megléte szükséges ahhoz, hogy sikeres vizsgálatokat tudjunk végezni. Az egyik tulajdonság a *lokális függetlenség*, a másik a *monotonicitás*. Mit jelentenek ezek, és miért fontosak számunkra?

A lokális függetlenség azt jelenti, mint már volt róla szó (és most maradjunk a mi egyszerűen kezelhető esetünknél, a legprimitívebb Rasch-modellnél), hogy a mérés során

[34] Meg kell említenem, hogy a látens változók alkalmazásához köthető modelleknek számtalan klasszifikációja létezik. Az item-válasz elméletek (IRT) körében pedig létezik az általánosított item-válasz elmélet (Generalized Item Response Theory, ld. pl. Mellenbergh 1994), amely egy átfogó keretben, és általános matematikai modellt alkotva, egységesen írja le ezeket az elméleteket. A részletek iránt érdeklődő olvasók számára különösen hasznos lehet Mellenbergh írásában a 304. oldalon található táblázat és a hozzá tartozó magyarázatok tanulmányozása.

szóba kerülő események függetlenek egymástól, illetve egy kicsivel precízebben: egy átfogó (több személyre és több feladatra vonatkozó) mérés eredménye, mint esemény valószínűsége egyenlő az azt alkotó események valószínűségeinek szorzatával. Vagyis vegyük minden egyes személy és minden egyes feladat esetén a tényleges megoldás (vagy jó, vagy nem jó megoldás) valószínűségét, tehát a ténylegesen bekövetkező $N \times L$ számú esemény valószínűségét, s ha ezek szorzata adja ki a teljes esemény valószínűségét, akkor azt mondjuk, hogy érvényesül a lokális függetlenség. A tesztelési feladatok – gondos megvalósítás esetén – viszonylag jól teljesítik ezt a követelményt. Ha lehetőség van csalásra, ha szerepet játszhat a tanulás a teszt feladatainak megoldása során, ha a feladatok nem függetlenek egymástól (például vannak közös részeik), akkor viszont sérül a lokális függetlenség (a felsorolt problémák csak példák, számos oka lehet még a lokális függetlenség sérülésének).

A másik tulajdonság, amelynek megléte szükséges a vizsgálatok sikeréhez, a monotonicitás. A legegyszerűbb Rasch-modell esetén ez azonos az ún. egydimenziósság követelményével (Karabatsos 2000, Walker és Beretvas 2000). Azt jelenti, hogy nagyobb képességfejlettségű személy jó megoldásainak valószínűségei minden feladat esetén nagyobbak kell, hogy legyenek, mint egy kisebb képességfejlettségű személy jó megoldásainak valószínűségei ugyanazon feladatok esetén. És hasonló mondható el a feladatokkal kapcsolatban is: egy nehezebb feladatot mindenkinek kisebb valószínűséggel kell tudni megoldani jól, mint egy könnyebb feladatot. A mérés tartalma szerint ez lehet racionális, adaptív feltételezés, de lehet, hogy nem az. A pedagógiát jobban érdeklő összetett emberi teljesítmények esetén semmiképpen nem az. Emlékezzünk rá! A klasszikus tesztelmélet problémája ugyanez volt. A közös probléma tehát az, hogy *az esetek (a képességek) döntő többsége esetén nincs monotonitás*.

Fit-elemzés

Hogy merem én azt állítani, hogy komoly gondok vannak nem csak a klasszikus-, hanem a modern tesztelmélet alkalmazásával is, amikor a kutatásmódszertan terén, és még számos más területen is nálam sokkal képzettebb szakemberek hittek, hisznek az alkalmazásában, illetve – s ez a fontosabb – számtalan empirikus vizsgálat találta úgy, hogy a modern tesztelmélet segítségével elvégzett számítások szerint a modellek működtek? Tévedtek volna a kutatók? E kérdés számomra is ijesztő, s az ok, hogy ennek a könyvnek a megírása nagyon sok időt igényelt, elsősorban ez a bizonytalanság volt. Egy összefüggést világosan látok: ha a jelenség nem egydimenziós, akkor a *Rasch-modell* alkalmazása hibás. Ezt azonban mindenhol lehet olvasni a szakirodalomban, az egydimenziós jelleg egyértelműen elfogadott követelmény, illetve nagyon nyilvánvaló a következtetés. Az egydimenziós jelleg a modern tesztelméleti modellek alkalmazhatóságával kapcsolatban minden oldalról körüljárt követelmény.

Ám még mindig nem válaszoltam arra a kérdésre, hogy mindezek ellenére miért találunk a szakirodalomban számtalan sikeresnek tűnő alkalmazást. Miért van az, hogy a kutatók által elvégzett ellenőrzés, a „fit-elemzés" nagyon sok esetben azt mutatja, hogy a modellek alkalmazhatók?

Először is, látni kell, hogy a szakirodalomban is sokféle megfontolás kapott szerepet, illetve számtalan kritikája létezik az alkalmazott fit-elemzéseknek. A modern tesztelméleti kérdésekkel foglalkozó kutatók kialakították azokat az eljárásokat, amelyek az előző bekezdésekben szereplő módszerekhez viszonyítva egyszerűbben teszik lehetővé az egydimenziósság vizsgálatát. *George Karabatsos* egy kifejezetten e témában írt tanulmányában felhasznál az elemzés során 36 fit-elemzést, vagyis az adatok Rasch-modellnek való megfelelését elemző módszert (Karabatsos 2003), *Rob Meijer és Klaas Sijtsma* szerint az ilyen módszerek száma 40 fölött van (Meijer és Sijtsma 2001).

A fit-elemzések között el kell különítenünk egymástól a paraméteres és a nem paraméteres módszereket. A paraméteres módszerek felhasználják a számítások során a képességfejlettségekre, valamint a feladatnehézségekre kapott becsléseket, míg a nem paraméteres módszerek csak azt, hogy az egyes kérdésekre az egyes vizsgált személyek milyen (helyes vagy helytelen) választ adtak. Foglalkozzunk először a paraméteres megoldásokkal!

Lényegében minden paraméteres megoldás abból próbál következtetésekre jutni, hogy megvizsgálja, a számítások eredményeként kapott valószínűségértékek rendszere milyen „távol" van a tényleges válaszok rendszerétől. Ha egy konkrét személy egy konkrét feladatot p valószínűséggel tud megoldani, és ha ez a személy valóban helyesen meg is oldotta a feladatot, akkor az $1 - p$ mennyiség jó mértéke lehet annak, hogy milyen távol vannak egymástól a jó megoldás és a háttérben lévő, egyébként természetesen nem ismert valószínűség. Gondoljunk bele! Ha p nagy, vagyis viszonylag közel van 1-hez, mondjuk 0,96, akkor az $1 - p$ értéke (példánkban 0,04), egy kicsi szám, legalábbis a lehetséges legnagyobb „távolsághoz", vagyis az 1-hez viszonyítva. Fordítva, ha az adott személy annak ellenére jól oldotta meg a feladatot, hogy a helyes megoldás valószínűsége az ő esetében kicsi (vagyis egy váratlan esemény következett be), legyen ez a valószínűség 0,07, akkor az $1 - p$ értéke 0,93, ami a lehetséges legnagyobb „távolsághoz", az 1-hez nagyon közeli, vagyis igencsak nagy érték, ami jól jelzi, hogy a valóságos megoldás távol van attól, amit a valószínűség alapján várunk.

Ha a személy a feladatot nem tudja megoldani, akkor a várakozásnak megfelelő helyzettől való távolságot éppen a p mérheti. Ha a megoldás valószínűsége kicsi, akkor ez a kicsi érték adja a „távolságot", megfelelve annak, hogy közel van egymáshoz a megoldás és a várakozás, míg ha p értéke nagy, vagyis 1-hez közeli, akkor a „távolság" is nagy, ahogy lennie kell, hiszen a várakozásunk (nagy a helyes megoldás valószínűsége), és ami bekövetkezett (nem tudta megoldani a feladatot) távol állnak egymástól.

Most még annyi matematikai megfontolásra van szükségünk, hogy a p érték írható $p - 0$-nak is. De hogy jobban hasonlítson a fenti másik kifejezéshez, az $1 - p$-hez, jó lenne $0 - p$-t írni, amivel azonban az érték negatívvá válik, márpedig távolságként negatív számokat nem használunk. Ezért aztán a várakozástól való eltérés jellemzésére ezeknek az értékeknek a négyzetét használjuk, vagyis az $(1 - p)^2$ és a $(0 - p)^2$ értékeket. Egységesen, és most már használva indexeket is, továbbá y-nal jelölve a különbséget:

$$y^2 = (x_{\nu\lambda} - p_{\nu\lambda})^2,$$

ahol $x_{\nu\lambda}$ jelöli a feladatmegoldás sikerét (a ν-edik személy a λ-adik feladatot jól ($x_{\nu\lambda} = 1$), vagy helytelenül ($x_{\nu\lambda} = 0$) oldja meg), és $p_{\nu\lambda}$ a helyes megoldás valószínűsége. A korábban bemutatott számítások eredménye alapján a valószínűségre van becslésünk, a mérés során kiderült, mennyi az x, tehát ezeknek az értékeknek a kiszámítása nem okoz gondot. Egy-egy ilyen, a megadott képlettel kiszámítható érték jelzi a bekövetkezett eredmény „távolságát" a várakozástól. A fit-elemzés egyik szokásos kérdése, hogy az egyes vizsgált személyek „nem lógnak-e ki a sorból", azaz a feladatmegoldásaik sikerét jelző $x_{\nu\lambda}$ értékek nincsenek-e túlságosan távol a $p_{\nu\lambda}$ valószínűségektől.

Tekintve, hogy $p_{\nu\lambda}$ éppen az adott feladat megoldása során az adott személy által szerezhető pontszám várható értéke, a fenti képletben az $x_{\nu\lambda} - p_{\nu\lambda}$ a mért érték, mínusz a várható értéke, amit ha elosztunk a szórással, standardizált értéket kapunk (az így számítható értékek várható értéke 0, szórása 1). Valószínűségszámítási összefüggésekből tudjuk, hogy a 0 és 1 értéket felvevő változók szórásnégyzete, amennyiben az 1 értéknek $p_{\nu\lambda}$ a valószínűsége, $p_{\nu\lambda}(1-p_{\nu\lambda})$. Vagyis a következő z érték (a képletben a z^2-et írjuk fel) standard normális eloszlású:

$$z_{\nu\lambda}^{\ 2} = \frac{(x_{\nu\lambda} - p_{\nu\lambda})^2}{p_{\nu\lambda}(1-p_{\nu\lambda})}.$$

A kissé „áttekinthetetlen képlet jóval egyszerűbbé válik, ha belegondolunk, hogy az $x_{\nu\lambda}$ értéke csak 1 és 0 lehet. Ha 1 az értéke, vagyis az adott feladatot az adott személy helyesen oldotta meg, akkor $z_{\nu\lambda}^2 = (1 - p_{\nu\lambda})/p_{\nu\lambda}$, ha viszont 0 az értéke (nem sikerült a feladatot megoldani), akkor $z_{\nu\lambda}^2 = p_{\nu\lambda}/(1 - p_{\nu\lambda})$, vagyis a két érték egymás reciproka.

Képezzük most az egy adott vizsgált személyre jellemző z^2 értékek átlagát (L feladatról van szó):

$$OMS_\nu = \frac{1}{L}\sum_{\lambda=1}^{L} z_{\nu\lambda}^{\ 2} = \frac{1}{L}\sum_{\lambda=1}^{L}\frac{(x_{\nu\lambda} - p_{\nu\lambda})^2}{p_{\nu\lambda}(1-p_{\nu\lambda})}.$$

(Természetesen az átlagolás egy itemre is elvégezhető lenne.) Ezt a statisztikát „külső négyzetátlagnak" nevezik („Outfit Mean Square"), és annak egyfajta mértéke, hogy a kiválasztott (a ν-edik) személy válaszai mennyire térnek el attól, amit várnánk tőle a jó megoldások valószínűségei alapján. Hasonló a másik, az adott személy válaszainak jellemzésére szánt mérték:

$$IMS_\nu = \frac{\sum_{\lambda=1}^{L} y_{\nu\lambda}^{\ 2}}{\sum_{\lambda=1}^{L} p_{\nu\lambda}(1-p_{\nu\lambda})} = \frac{\sum_{\lambda=1}^{L}(x_{\nu\lambda} - p_{\nu\lambda})^2}{\sum_{\lambda=1}^{L} p_{\nu\lambda}(1-p_{\nu\lambda})}.$$

E statisztika neve „belső négyzetátlag" („Infit Mean Square"), és bár kissé másképpen „viselkedik", mint az OMS_ν, gyakorlatilag ugyanúgy a személy feladatmegoldásai „megfelelőségének" a mértéke. Mindkét valószínűségi változó várható értéke 1. A konvenció szerint egy személy akkor „lóg ki a sorból", vagyis akkor nem fogadhatók el az eredményei, ha az itteni két statisztika (OMS_ν, IMS_ν) valamelyikének az értéke 1,3-nél nagyobb. Ennek sok minden lehet az oka, például egyszerű csalás a feladatmegoldások során, véletlenszerű válaszok adása (ha olyan a teszt), különleges, a többségre jellemző tudásstruktúrától eltérő tudásrendszer, stb. Ha a fit-elemzés problémát jelez, akkor a mérés során a személyre kapott eredmény nem hiteles, nem lehet további következtetések kiindulópontja.

Megtehetjük, hogy a felmért személy adatainak felhasználásával (az OMS_ν vagy IMS_ν értékkel) kiszámítjuk a következő értéket:

$$t = \left(\sqrt[3]{MS_\nu} - 1\right)\left(\frac{3}{\sqrt{V_{MS_\nu}}} + \frac{\sqrt{V_{MS_\nu}}}{3}\right)$$

E képletben az MS_ν az OMS_ν, IMS_ν valamelyikét jelenti, a V pedig a variancia, a szórásnégyzet jele, itt az MS_ν indexszel a ν. személyhez tartozóan az OMS_ν vagy az IMS_ν

értékek varianciája. A felírt valószínűségi változó közelítőleg student-féle t-eloszlású, és ezért a pedagógiai vizsgálatoknál szokásos p = 0,05-os szignifikanciaszint mellett azt kell megvizsgálnunk, hogy ez az érték nem haladja-e meg a 2 küszöbértéket. Ha igen, akkor az adott személyre vonatkozóan a mérés sikertelen.

A modell egészének vizsgálatára is van (számos) fit-elemzési módszer, de a most éppen általam is leírt statisztikák is jól használhatók erre a feladatra. Ehhez az OMS_v, IMS_v értékeket összegezni kell az N számú személyre (v = 1, 2, …, N). A kapott valószínűségi változó közel khi-négyzet eloszlású, $N + L - 1$ szabadsági fokú, és ezért konkrét mérés során kapott értéke vizsgálható abból a szempontból, hogy meghalad-e egy küszöbértéket. Ha igen, akkor az adatok nem elégítik ki a Rasch-modell követelményeit.

A modern tesztelméleti eljárások alkalmazásának kritikája

De mi a helyzet akkor, ha viszonylag sok személy adatai jeleznek problémát? Ekkor már felmerülhet az is, hogy maga a képesség nem elégíti ki a Rasch-modellben megfogalmazott követelményeket, tehát a modern tesztelméleti módszerek egyáltalán nem alkalmazhatók. Gondot jelent számos a szakirodalomban fellelhető megfontolásban, hogy a fit-elemzés során a „sorból kilógónak” tapasztalt személyeket vagy itemeket egyszerűen kiveszik a vizsgálatból. Ha például egy itemmel, vagyis egy feladattal (vagy egy feladat egy elkülöníthető részével) tesszük ezt, akkor azt vajon milyen jogon tesszük? Nem a vizsgálni kívánt képességhez tartozik? De akkor miért válogattuk be a tesztbe? Tévedtünk? De annak nem a fit-elemzés során kellene kiderülnie. Lehet, hogy a fit-elemzés eredménye, hogy tudniillik néhány item alkalmatlannak bizonyult, nem azt a reakciót kellene, hogy kiváltsa belőlünk, hogy ezeket az itemeket száműzzük a vizsgálatból, hanem ahhoz kellene, hogy elvezessen bennünket, hogy felülvizsgáljuk a Rasch-modell alkalmazhatóságát a problémára. (És emlékezzünk vissza, a klasszikus tesztelmélet esetén is beleütköztünk ebbe a problémába.) Általában: a tesztfejlesztés feladatát, amikor is kiszűrünk feladatokat, illetve vizsgált személyeket, én eleve értelmetlennek tartom. Ha a mérés megfelelne a méréselméleti igényeknek, akkor tetszőleges teszt alkalmazása és a populáción belül tetszőleges minta választása esetén hiteles, használható eredményt kellene kapnunk. Mint már többször érzékeltettem, a tesztfejlesztés eljárása, a feladatok és a vizsgált személyek szűrése valójában egy olyan teszt kialakítását jelenti, amellyel szemben a bevett fit-eljárások már nem lehetnek érzékenyek.

A fit-statisztikákkal szemben nagyon sokféle kritika fogalmazódott meg (csak példaként néhány szakirodalmi forrás: Emons, Meijer és Sijtsma 2002; Karabatsos 2000, Smith, Schumacker és Bush 1995; Meijer és Sijtsma 1995, Rogers és Hattie 1987). Talán a legfontosabb (és még én is szeretnék később ezzel foglalkozni), hogy az eloszlás khi-négyzet eloszlás jellege sok esetben súlyosan sérül. Vagyis így a fit-elemzés megbízhatatlan eredményt szolgáltat. Ez is az oka annak, hogy oly nagy számban állnak rendelkezésre módszerek, hiszen a kutatók igyekeznek kiküszöbölni a korábban kialakított statisztikák problémáit. Igaz, „tökéletes” megoldásra eddig nem tettünk szert, nagy valószínűséggel olyan nem is létezik. Nem tudom itt a fit-statisztikákra vonatkozó konstruktív (konstrukciós) és kritikai szakirodalom minden megállapítását, de még a legfőbbeket sem elősorolni (a források fent megadott, messze nem teljes listája némi eligazítást adhat a részletek iránt érdeklődő olvasónak), azonban a módszerek egy része, vagyis a paraméteres eljárások esetén jellemző hiányosságok gyökerére szeretnék rámutatni.

Idézzük fel újra a modern tesztelmélet konstrukciójának kiindulópontjait! Van egy képesség, amelyet jobb híján azokkal a feladatokkal igyekszünk meghatározni, amelyek kifejezetten e képesség „alkalmazásával” oldhatók meg. Ha van a vizsgált személyeinknek egy csoportja, egy minta, akkor az e csoporthoz tartozók mindegyikével kapcsolatban elmondható, hogy az egyes feladatokat valamekkora, precízen nem ismert valószínűséggel tudják megoldani. A modern tesztelmélet segítségével, pontosabban a *Rasch-modell* alkalmazásával azok az esetek kezelhetők, amelyekben tudunk úgy hozzárendelni minden egyes személyhez annak képességfejlettségét kifejező számot, valamint minden egyes feladathoz annak nehézségét kifejező számot, hogy minden egyes helyes feladatmegoldáshoz, tehát ehhez az eseményhez a *Rasch* képlet segítségével tudjuk kiszámolni annak valószínűségét. Hogy ez tényleg így van-e, hogy tehát a Rasch-képlet helyesen (persze megközelítőleg helyesen) adja meg a valószínűségeket, hogy tényleg létezik a képességfejlettségeknek és a feladatnehézségeknek egy megfelelő rendszere, nem automatikus, nem feltétlenül igaz, és alapvetően empirikus vizsgálatok eredményeivel lehet csak megállapítani.

A paraméteres fit-statisztikák azt vizsgálják – némi leegyszerűsítéssel –, hogy milyen távol van egymástól a tényleges feladatmegoldások eredményeinek rendszere (a 0-k és 1-esek rendszere), és a Rasch-modell alkalmazásával kiszámított valószínűségbecslések sora. A megfigyelt eredményeket a Rasch-modellt kielégítőnek nevezzük, ha ez a távolság egy a konkrét vizsgálati módszerben meghatározott, általánosan elfogadott mértéket nem halad meg. Itt ragadhatjuk meg az ilyen jellegű módszerek nehézségeinek egyik okát. A módszerek nem a kapott eredmények (0-k és 1-esek rendszere), valamint a valóságos, objektív valószínűségek rendszere közti távolságot vizsgálják, hanem a megfigyelések, és az *azok felhasználásával becsült*, a Rasch-modellnek már megfelelő valószínűségek rendszere közti távolságot.

A valóságos, objektív valószínűségek rendszerét nem használhatjuk semmilyen számításban, hisz azt nem ismerjük. Nem is jogos számon kérni a fit-elemzésektől, hogy miért nem az objektív, valóságos helyzethez viszonyítják az adatokat, hiszen a hipotézisvizsgálatok során soha nem ezt tesszük. Minden hipotézisvizsgálatban van valamilyen feltételezésünk az objektív értékekkel, helyzettel kapcsolatban, és azt vizsgáljuk, hogy ha az nem csak feltételezés lenne, hanem a tényleges helyzet, akkor a mérés során kapott adatok, vagy az azoknál is szélsőségesebbek valószínűsége mennyi lenne. Ha ez a valószínűség kellően kicsiny, a mi tudományunk esetében a legtöbbször 0,05 a küszöbérték, akkor úgy döntünk, hogy a hipotézisben megfogalmazott állítás nem állja meg a helyét.

A paraméteres fit-statisztikák tehát a Rasch-modellre alapozott becslési módszerrel kiszámított valószínűségek és e becslési folyamat kiindulópontját képező tapasztalati értékek (a 0-k és 1-esek) rendszerét hasonlítják össze. Rendkívül érdekes ebből a szempontból *Antal Judit Ohio Állami Egyetemen* (Ohio State University) megvédett doktori disszertációjának elemzése (Antal 2003). Antal elsősorban azzal a kérdéssel foglalkozik, hogy miképpen lehetne kiküszöbölni azt a problémát, hogy az általam is fentebb bemutatott *OMS* és *IMS* statisztikák (már az összegzésről van szó, tehát az egész modell megfelelő voltának elemzéséről) esetében csak akkor kapunk kellően megbízható eredményeket, ha e statisztikák khi-négyzet eloszlásúak, viszont az esetek egy nagy részében ez a feltétel nem teljesül. Antal e statisztikákra építve kidolgoz egy nem paraméteres próbát, sőt, hogy pontosabb legyek, négy hasonló próbáról van szó, amelyek lehetővé teszik, hogy az eloszlás „szabályostól” való eltérése ne játsszon szerepet a Rasch-modellnek való megfelelés vizsgálata során. Sokatmondó Antalnak az az eredménye, hogy olyan esetekben, amikor az eloszlásra érzékeny, eredeti hipotézisvizsgálatok azt jelzik, hogy az adatok nem elégítik ki a Rach modell követelményeit, az általa fejlesztett, szimulációt is alkalmazó módszerekkel szinte minden esetben „jó” eredményt kapunk, vagyis a Rasch-modell alkalmasnak bizonyul a vizsgálatra. Disszertációjának végén Antal ebből az eredményből a következőkre jut:

> E megfigyelés fényében akár arra a nagyon üdvös következtetésre is juthatnánk, hogy *a Rasch-modell majdnem mindig kielégíti az adatokat*. Ezzel megszüntethetnénk a Rasch-modell régóta ismert és igencsak kellemetlen paradox sajátosságát, amelyet az a tény tartott életben, hogy miközben a Rasch-modell alkalmazása elméletileg rendkívül indokolt, mégis a tesztek egy meglehetősen nagy száma esetén bizonyul nem megfelelőnek. E disszertáció egyik mondanivalója, hogy ez a fit elemzés feltételeinek gyakori nem teljesülése miatt van, és nem maga a modell a rossz. E felismerések fényében a Rasch-modell a tesztfejlesztés még vonzóbb modelljeként tűnik fel (im. 87-88. o.).

Magam a szerintem is korrekt módon „kihozott" empirikus erdményekből nem erre a következtetésre jutnék (ismét egy példa, hogy az empirikus eredmények interpretációja nagyon gyakran rendkívül különböző lehet). Az alaphelyzet szerintem az, hogy az *OMS* és *IMS* statisztikák eleve hibás, vagy legalábbis bizonyos mértékig félrevezető meggondoláson alapulnak. Ezt mutattam be fentebb, vagyis azt, hogy a mért adatokon nyugvó összehasonlítása a jó feladatok megoldása valószínűségértékeinek és maguknak a kiinduló mért adatoknak (1-esek és 0-k) „igazságtalan", vagyis csak arra ad információt, hogy a mért adatokhoz legközelebb álló, a Rasch-modellnek megfelelő valószínűségrendszer mellett nem túl valószínűtlenek a mért adatok. Antal eredményei számomra azt mondják, hogy ha korrektebbé tesszük a hipotézisvizsgálatot, akkor még inkább kiviláglik, hogy e vizsgálatok nem sokat érnek, hiszen szinte bármilyen realisztikus adatrendszer kielégíthető olyan valószínűségrendszerrel, amelyet a Rasch-modell alapján határozunk meg. Másképpen fogalmazva: a hipotézisvizsgálatok azt jelzik, hogy a Rasch-modellre alapozott számítások valóban nagyszerű eredményt hoznak, de csak abban az értelemben, hogy az adatokat kielégítő jellegűnek szánt valószínűségértékek tényleg elég jól megfelelnek az adatoknak.

Ez olyan, mintha egy intelligenciavizsgálatban azt találnánk, hogy egy minta átlaga 105 pont, majd hipotézisvizsgálatot végeznénk arra vonatkozóan, hogy lehet-e a populáció várható értéke 105 pont. Az egymintás t-próba azt az eredményt adná természetesen, hogy az ilyen, vagy a 105 ponttól még jobban eltérő eredmények valószínűsége 1, vagyis elfogadhatjuk azt a nullhipotézist, hogy a populáció átlaga 105 pont. A vizsgálat nem teljesen értelmetlen ha valóban az lenne a nullhipotézisünk, hogy a populáció átlaga 105 pont, akkor a mérés ezt nagyon jól alátámasztaná. (más kérdés, hogy magát a hipotézisvizsgálatot természetesen nem kellene elvégezni.) Nem is lenne semmi baj, ha a nullhipotézist valamilyen külső információ alapján fogalmaztuk volna meg. De nem, mi magának a mintának a felmérése alapján állítottuk a nullhipotézisben, hogy 105 pont a populáció átlaga. A minta valóban azt az állítást támasztja alá a legerősebben, hogy a populáció várható értéke 105 pont, de ezzel az információval nem megyünk semmire. A teljesség kedvéért hozzá kell tennem, hogy ez a kritika a legelterjedtebb fit-elemzési módszerekkel kapcsolatban nem a saját találmányom, már korábban is olvashattunk róla, legalábbis *George Karabatsos* 2000-ben megjelent írásában egy említés erejéig (161. o.).

Természetesen lehet olyan adatrendszereket mesterségesen előállítani, amelyek semmilyen Rash modell alapján számított valószínűségekkel nem férnének össze, még *Antal Judit* módszerei is problémát jeleznének. Két típusba sorolható feladatokat képzeljünk el, és két részmintát. Az egyikbe tartozó személyek az egyik típusba tartozó feladatokat tudják jól megoldani, míg a másik típusú feladatok esetén sokkal rosszabb a teljesítményük, és a másik részmintába tartozók pontosan fordítva lennének ezzel. A módszerek mindegyike természetesen jól jelezné, hogy egy ilyen esetben nem alkalmazható a Rasch-modell. Kérdés, hol van a „határ" (és egyáltalán miképpen lehet egy ilyen „határt" megadni), amelytől a módszereknek jelezniük kellene, hogy baj van. Magam úgy látom (miközben hallatlanul veszélyes ilyen kijelentéseket tenni egy publikációk áttekinthetetlen mennyiségét tartalmazó területen!), hogy erre a problémára nincs egyelőre megoldás.

A probléma lényegét még egy módon tehetem világosabbá. Ha van a feladatoknak – egyszerűsítsünk – egy véges rendszere, és van egy személyekből álló, véges populációnk, akkor bármilyen mérés előtt már adott azoknak az „objektív" valószínűségeknek a rendszere, a mintázata, amely valószínűségekkel a populációba tartozó személyek az egyes feladatokat helyesen tudják megoldani. Ez a valószínűségmintázat vagy megfelel a Rasch-modellnek, vagy nem[35]. Nevezzük el a Rasch-modellnek megfelelő valószínűségmintázatokat Rasch-típusúaknak. További egyszerűsítésként most képzeljük el, hogy minden személlyel megoldattuk az összes feladatot, vagyis adott a megoldások rendszere, már sokszor emlegetett, 0-kból és 1-esekből álló rendszer, nevezzük el ezt most megoldásmintázatnak. A korábbiakban vázolt megoldások valamelyikét alkalmazva a megoldásmintázat segítségével meghatározhatjuk azt a Rasch-típusú valószínűségmintázatot, amely mellett éppen a kapott megoldásmintázat a legvalószínűbb. Bizonyos értelemben a Rasch-típusú valószínűségmintázatok közül ez utóbbi áll a „legközelebb" a megoldásmintázatunkhoz. A paraméteres fit-elemzések azt mutatják meg, hogy ez a „távolság"[36] nem túl nagy-e. Mert ha túl nagy, akkor az annak a jele, hogy a megoldásmintázat oly mértékben valószínűtlen a hozzá „legközelebb" álló Rasch-típusú valószínűségmintázat érvényesülését feltételezve is, hogy megrendül a hitünk abban, hogy tényleg a kiszámolt Rasch-típusú mintázat áll-e a háttérben. Ez idáig még logikus következtetés: ha még a „legközelebbi" Rasch-típusú mintázat érvényesülését is el kell utasítani, akkor nem lehet szó a Rasch-modell alkalmazhatóságáról.

A gondjaink akkor fokozódnak, amikor pozitív döntést kell hoznuk. Tegyük fel, hogy a paraméteres fit-elemzés nem jelzi, hogy baj lenne, a megoldásmintázat nincs túl „távol" a segítségével kiszámolt, Rasch-típusú valószínűségmintázattól. Logikai hiba ebből a tényből arra következtetni, hogy akkor érdemes úgy dönteni, hogy érvényesül a Rasch-modell. Azért logikai hiba, mert a megoldásmintázatunk végtelen sok valószínűségmintázathoz van sokkal „közelebb", mint a kiszámolt Rasch-típusúhoz. A paraméteres fit-elemzés egyáltalán nem mond semmit arról, hogy a tényleges, a háttérben meghúzódó valószínűségmintázat tényleg Rasch-típusú-e. Csak azt mondja meg, hogy a lehetséges Rasch-típusú valószínűségmintázatok közül a megoldásmintázathoz „legközelebb" álló „kellően közel" van-e. De ha „kellően közel" van, akkor is rettenetesen

[35] Helyesebben: mivel egy képességhez tartozó feladatok megoldási valószínűségei soha nem elégíthetik ki a valóságban egészen pontosan a Rasch-modellt, ezért léteznie kell egy kritériumnak, hogy milyen „eltérés" esetén tekintjük még adaptívnak azt a kijelentést, hogy a valószínűségmintázat, amely a háttérben „meghúzódik", jó közelítéssel a Rasch-modellnek megfelelő. Ismereteim szerint ilyen megállapodás nincs. Mivel nem sokat tudnánk kezdeni vele (a valóságos, a háttérben ténylegesen meghúzódó valószínűségeket nem ismerhetjük pontosan), így közvetlen gyakorlati jelentősége nincs a kérdésnek. Más kérdés, hogy matematikai módszerekkel a kérdést vizsgálni kellene, és akár még bizonyos gyakorlati vizsgálatok is tervezhetők lennének a kérdés megválaszolására.

[36] Ezekben az esetekben, amikor tehát a mintázatok „távolságáról" van szó, soha nem valamilyen precíz, matematikai értelemben vett távolságra kell gondolnunk. A valószínűségek értéke jelez valamifajta „távolságot", amely azonban nem rendelkezik azokkal a tulajdonságokkal, amelyeket a matematika megkövetel az ilyen mennyiségektől.

„távol" van a valóságosan előforduló esetekben. Már a fent bemutatott, 7 személy és 5 feladat által adott, nagyon primitív példa esetén is a bemutatott megoldásmintázat valószínűsége a hozzá legközelebbi Rasch-típusú valószínűségmintázat érvényesülését feltételezve kb. $1{,}1\times10^{-7}$, vagyis alig több, mint egy tízmilliomod. Ha valaki mond egy akármilyen 0 és 1 közötti számot, ami lehet nagyon-nagyon közel 1-hez, nagyon könnyen tudunk szerkeszteni olyan valószínűségmintázatot, amely mellett a megoldásmintázatunk valószínűsége ez a megadott szám lesz. Milyen okunk van feltételezni, hogy a nagyon pici valószínűséget szolgáltató, Rasch-típusú valószínűségmintázat áll a háttérben, és nem a sokkal nagyobb valószínűséget adó másik, ami nem Rasch-típusú? A megoldásmintázat csak a Rasch-típusú valószínűségmintázatok közül a hozzá „legközelebb" állótól nincs túl „távol", de mérhetetlenül sok olyan (nem Rasch-típusú) valószínűségmintázat van, amihez sokkal, de sokkal „közelebb" van. Így ez a vizsgálat semmit nem mond arról, hogy a helyzet valóban megfelel-e a Rasch-modellnek, amennyiben az elvégzett vizsgálat alapján pozitív választ kell adnunk.

És itt kell megemlítenem egy olyan meglepő matematikai eredményt, amely rendkívüli módon megerősíti azok véleményét, akik jelentős problémákat látnak a teszteléssel kapcsolatban. *William H. Batchelder* és *Louis Narens* még 1977-ben jelentették meg a *Philosophy of Science* című folyóiratban azt a tanulmányukat, amelyben bebizonyítják, hogy a dichotóm adatokkal dolgozó kutatások során a vizsgált jelenség mögötti látens változók struktúrájára kapott eredmények jelentős mértékben függnek a kérdések megfogalmazásától. A dichotóm változó fogalma a kétértékű változót jelenti, vagyis az eredmény az általunk vizsgált tesztelésekre is vonatkozik (meg tudta oldani, nem tudta megoldani). A látens változó pedig például éppen az, amelyet mi a megelőző gondolatmenetekben kerestünk, a képességfejlettségnek az a belső, közvetlenül nem megmérhető értéke, amely a feladatok nehézségével együtt a feladatmegoldás-valószínűségekkel a Rasch-modell által leírt kapcsolatban van. *Batchelder* és *Narens* azt mutatják meg, hogy ha van két azonos számú kérdésből álló teszt, a két teszt pontosan ugyanazt a tudást feltételezi, csak abban különböznek egymástól, hogy az egyikben szereplő kérdésekre adott válaszok a másik teszt kérdéseire adott válaszok logikai függvényei (vagyis a két teszt a szerzők definíciója szerint információs szempontból ekvivalens), akkor a két teszttel akár egymástól lényegesen eltérő látens struktúrákat „tárhatunk fel".

Röviden foglalkozzunk a nem paraméteres fit-eljárásokkal is! Ezek nem használják a megfontolások során a méréskor becsült valószínűség értékeket, csakis a megoldás-táblázat csupa 0 és 1 adatait veszik tekintetbe. Ezek esetében azt a problémát nem vethetjük fel, hogy a számítások eredményeit (a jó feladatmegoldás valószínűségeket) hasonlítjuk össze azzal az adatrendszerrel, amit amúgy felhasználtunk a kiszámításukhoz. A nem paraméteres próbák arra a kérdésre keresik a választ, hogy vajon a megoldástáblázatban a 0-k és 1-esek rendszere mennyire támasztja alá azt a hipotézist, hogy itt egydimenziós jelenséggel van dolgunk. Képzeljük el, hogy a személyeket úgy soroljuk fel, hogy legfölülre kerüljön az, aki a legtöbb feladatot tudta megoldani, és így jöjjenek egymás után lefelé a vizsgált személyek, a sor végén az legyen, aki a legkevesebb feladatot tudta megoldani. A feladatokkal tegyük azt, hogy a táblázatban az első (a baloldali) oszlopba annak a feladatnak az egyesei és nullái kerüljenek, amelyet a legtöbben tudtak megoldani, és így tovább, ahogy haladunk jobbra, egyre csökkenjen az adott feladatot megoldó személyek száma. Ha a feladatok és a személyek rendszere tökéletesen megfelelne az egydimenziósságnak, akkor azt látnánk egy ilyen táblázaton, hogy minden sorban balról egy darabig csak egyesek lennének, de egy ponttól már csak 0-k, és minden oszlopban felülről lefelé egy darabig csak egyesek, majd csupa nulla. Az egyesek egy ilyen táblázatban a „bal felső sarokban" helyezkednének el egy zárt „alakzatban", s az összes többi elem „a jobb alsó sarok irányában" nulla lenne. Ilyen abszolút tökéletes megoldástáblázat nem is lenne alkalmas a vizsgálatra, mert elkerülhetetlen, hogy legyen vagy csupa egyesből, vagy csupa nullából álló sor, vagy oszlop. Eleve némi kompromisszumot kell kötni tehát.

A nem paraméteres fit-elemzési eljárások arra szolgálnak, hogy megtudjuk, a valóságos, mért megoldástáblázat milyen „távol" van ettől az ideálistól, az egydimenziós esettől. Az ilyen próbák vagy megfelelő illeszkedést mutatnak, vagy sem. A fentiekben azonban már megadtam ennek az eljárásnak az egyik problémáját. Az illeszkedés sokkal inkább a „jó" tesztfejlesztést, a „jó" személy- és feladatszelekciót dícséri, és nem arra utal, hogy az adott képesség valóban egydimenziós. Lehet úgy kiválogatni feladatokat, és lehet úgy szűrni a személyeket, hogy a konkrét vizsgálatban a nem paraméteres vizsgálatok akármelyike egydimenziósként láttassa velünk a képességet. Ha a „képesség" pusztán azokból a feladatokból állna, amelyeket beválogattunk a tesztbe, és a populáció is azonos lenne a végül is a mintába bekerültekkel, akkor mérésünk teljes mértékben korrekt lenne. Ám a feladatok köre a tesztbe bekerültekhez képest jóval nagyobb, és a szóba jövő minták is jóval nagyobbak.

A nem paraméteres eljárások azonban küzdenek azzal a bajjal is, amelyet logikai problémaként mutattam be a paraméteres próbákkal kapcsolatban. Valójában ezek is felhasználják a kapott eredményt annak az adatrendszernek a kijelölése során, amellyel összehasonlítják a kapott, a feladatok megoldására vagy nem megoldására vonatkozó eredményeket. Hiszen valójában nem tudhatjuk, hogy melyik az az „ideális" eredménymátrix, amelyhez viszonyítani kell a mérés során kapott eredménymátrixot. Az „ideális eredménymátrixban minden sorban balról jobbra van néhány 1-es, majd ezt követően 0-k jönnek. De feltételezve, hogy a vizsgált képesség a lehető legnagyobb mértékben megfelel az egydimenziósággal kapcsolatos követelménynek, akkor is még nagyon sokféle ilyen „ideális" mátrix állhat a háttérben. A mérés során kapott eredménymátrix csak egy a lehetségesek közül, és a nem paraméteres fit eljárások azzal az ideálissal hasonlítják össze, amely hozzá a „legközelebb" áll. Vagyis a logikai probléma ugyanaz, mint a paraméteres eljárásoknál.

Nagy kérdés számomra is, hogy a nagy mérésekkel összefüggésben, mint a PISA, a TIMSS, az Országos kompetenciamérés, miképpen merülnek fel ezek a problémák. A szövegértés, a matematikai eszköztudás, a természettudományos felkészültség egyenként rendkívül komplex kompetenciák, eddigi érvelésem szerint kizárt, hogy teljesüljenek az alapvető elvárások, amelyek valamely modern tesztelméleti modell alkalmazásánál elengedhetetlenek. Ám ezeknek a méréseknek az esetében okoskodhatunk másképpen is. Az tény, hogy a gondos munka eredményeként minden ilyen mérés esetében olyan tesztekkel van dolgunk, amelyek megfelelőnek mutatkoznak az alkalmazott fit-eljárásokkal szemben. Ez ellentmondásnak tűnik, összevetve az előbbi, a vizsgált kompetenciák rendkívüli összetettségére vonatkozó állításom fényében. Ezt az ellentmondást az oldja fel, hogy ezekben a mérésekben valójában nem a leírt komplex tudásösszetevők fejlettségének mérése zajlik. Ha úgy tetszik, megfizetjük az árát annak, hogy modern tesztelméleti eszközöket alkalmazunk: egy gondos tesztfejlesztés eredményeként létrejön egy olyan feladatrendszer, amely már kielégíti a megfelelő modell követelményeit, vagyis egydimenziós. De ha ilyen, akkor már nem mérheti a megcélzott komplex jellemző fejlettségét, mert az ahhoz tartozó feladatoknak csak egy „egy dimenzióba besűrűsödő" részén mér. Ez azonban nem akkora probléma, amekkorának látszik. A „nagy mérések" nem bizonyos képességek fejlettségét érintő kutatások, amelyekben a vizsgálatok tárgya határozottan az a bizonyos pszichológiai konstruktum lenne (szövegértés, stb.). Ám azok a feladatok, amelyek a vizsgálat tárgyát képező, és közelebbről nem meghatározható dimenziót alkotják, *rendkívül fontosak a számunkra*. Úgy látjuk, hogy ezekben a feladathalmazokban csupa olyan elem van, amelyekkel kapcsolatos felkészültséget érdemes vizsgálni, mert az oktatási rendszerek, részeik átfogó céloknak megfelelő működéséről tudunk meg sok mindent. *Adaptívnak bizonyul ez az eljárás*. Az más kérdés, hogy jó lenne kerülni a kommunikációban az olyan megfogalmazásokat, amelyek szerint ezek a mérések globálisan jellemzik a szövegértés, a matematikai- és természettudományos eszköztudás fejlettségi szintjét.

A gondos tesztfejlesztéshez az is hozzátartozik, hogy az egyes alkalmakkor elvégzett mérések teszteredményei egymással összevethetők. Ez azt jelenti, hogy bár részben új feladatsor jön létre minden alkalommal, arra a vizsgálatok szakemberei gondosan vigyáznak, hogy mindig ugyanannak a dimenziónak a feladatai kerüljenek be a feladatsorokba. Így minden alkalommal valóban ugyanazt a konstruktumot mérik a tesztek, az eredmények ténylegesen összehasonlíthatók egymással, van például értelme arról beszélni, hogy 2000 és 2015 között a PISA matematika teszteredményei hogyan változtak.

Ezen a ponton egy újabb ablakot kell nyitnom. A „nagy mérések" (PISA, stb.) megítélésével kapcsolatos fenti szöveget már régebben írtam, ez most egy a megjelentetés előtt közvetlenül befűzött rész. A világ gyorsan változik, egyik pillanatról a másikra kérdőjeleződhetnek meg biztosnak vélt tudásaink. Bárki tájékozatlannak tarthat, de e könyv elkészültének hosszú ideje alatt magam nem igazán figyeltem fel azokra a kritikai írásokra, amelyek elsősorban a PISA mérésről, annak akár még súlyosnak is nevezhető problémáiról szólnak. Úgy láttam – és be kell ismernem, hogy ez egyfajta vakság volt –, hogy a PISA méréssel kapcsolatban felmerült kritikák az OECD politikáját, oktatással kapcsolatos szándékait veszik célba elsősorban (lásd például Ursin 2015). Hogy ezek a kritikák mennyire jogosak, azt most és itt egyáltalán nem kívánom tárgyalni, mert messze vinne témánktól, a tesztelés inkább módszertani kérdéseitől. Ám az OECD politikáját ért kritikák „félretételével", azzal, hogy nem kívántam azokkal behatóbban foglalkozni, kizártam a tanulmányozás köréből olyan munkákat is, amelyek igenis módszertani problémákat vetettek fel. Egyenesen fogalmazva: számos kutatás megkérdőjelezte, hogy a PISA adatok megfelelő szinten kielégítik-e a Rasch-modellt. A dimenziók problémáját is felveti, de sok más problémát is vizsgál *Hervey Goldstein* (2004), és közöl negatív eredményeket. *Joachim Wuttke* (2007) amellett, hogy az egyes országkban zajló mintaválasztás problémáira hívja fel a figyelmet, kritizálja a PISA mérést több, a Rasch-modell érvényesülésével illetve nem érvényesülésével összefüggő kérdésben. *Svend Kreiner* és *Karl Bang Christensen* (2014) már négy PISA mérés eredményeinek ismeretében, és a PISA jelentések által is leírt modell megfelelést ellenörző módszerek alkalmazása alapján állítják, hogy az adatok nem megfelelően illeszkednek az egydimentziós Rasch-modellhez. Mint ahogy sok más kritikus, Kreiner és Christensen is elsősorban az országok rangsorolását kifogásolják, mondván: a Rasch-modell nem megfelelő érvényesülése miatt az országok átlageredmények alapján felállított rangsora teljes mértékben megbízhatatlan. Mindezen problémákkal kapcsolatban jó összefoglaló olvasható magyarul is, *Lannert Judittól* (2015).

A PISA vizsgálatokkal összefüggésben tehát létezik egy „masszív", erősen a metodológiát érintő kritika. A korábban általam kifejtettek alapján is érthető, miért merülnek fel a gondok. Bár a PISA szakemberei törekszenek arra, hogy a tesztek egydimenziósak legyenek, érvényesüljön a monotonitás, azonban az a módszer, amellyel ennek a feltételnek a teljesülését vizsgálják, vagyis az infit és outfit paramétereken alapuló paraméteres fit-elemzés, sajnos nem teszi eldönthetővé, hogy tényleg alkalmazható-e a

Rasch-modell. Ahogy volt róla szó: a megoldásmintázatok eltérése a hozzájuk „legközelebb” álló valószínűségmintázattól nem kérdőjelezi meg a Rasch-model alkalmazhatóságát, de egyáltalán nem támasztja alá, semmilyen mértékben, hogy e modell érvényesül.

Végül ehhez egy személyes megjegyzés: azt a véleményt, ami fentebb olvasható, hogy tudniillik a „nagy vizsgálatokkal”, mint amilyen a PISA nincs baj, valójában a könyv végső megfogalmazásából ki kellett volna hagynom (ez lett volna a logikus), és csak a legutóbbi két, sokkal kritikusabb bekezdést kellett volna leírnom. Én azonban szándékosan meghagytam ezt az ellentmondást a könyvben, a témában most két, egymástól eltérő véleményt is megfogalmaztam. A pozitívabb a korábbi, ez a kritikusabb csak a könyv megjelenése előtt kevéssel fogalmazódott meg bennem. E furcsa megoldással jelezni szeretném, hogy mennyire borotvaélen táncol a szakma, amikor a jelentős oktatáspolitikai döntéseket is befolyásoló nemzetközi (vagy országos) méréseket tervezi, végzi és interpretálja. Nincsenek még kimondva végső ítéletek, a 2018. évi PISA mérés előkészítése és szervezése már zajlik, folyamatosan minden évben megszületnek hazánkban is az Országos kompetenciamérés eredményei, miközben – látjuk – lenne mit elemezni, kutatni és vitatni ezen a területen.

És a többdimenziós elemzések?

Aki ismeri a modern tesztelmélet szakirodalmát, az tudja, hogy a számos modell között vannak olyanok is, amelyek számot vetnek a képességek többdimenziós jellegével (pl. Molnár 2013, van Abswoude, van der Ark és Sijtsma 2004, McDonald 2000, Ackermann, 1996). Nem szánok ennek a kérdésnek itt túl nagy helyet. Minden ilyen modell arról szól, hogy feltételezzük, a képesség előre meghatározott (nem túl sok) dimenzióból áll, és e több dimenzió „mentén” vizsgálunk vagy több feladattípust, vagy az egyének képességét gondoljuk „vektorosnak” (vagyis több összetevőből állónak). Ha például azt mondjuk, hogy a szöveges matematika feladatok megoldani tudását egy kétdimenziós képességnek tekintjük, ahol az egyik dimenzió a matematikai tudással, a másik dimenzió a szövegértéssel kapcsolatos, akkor talán e példából is látszik, hogy mi a probléma. A képességek igen nagy része esetén a dimenzionalitás összetettsége meg sem közelíthető néhány összetevő feltételezésével. Világos, hogy a matematikai felkészültség és a szövegértés külön-külön is rendkívül összetett képességek, és ezért a gyakorlatban csak az történik, hogy a többdimenziós vizsgálatokkal megkönnyítjük a dolgunkat a „jó” teszt feladatainak és a „jó” minta személyeinek kiválogatásában. A problémát azonban így nem tudjuk megoldani a pedagógiát elsősorban érdeklő komplex képességek esetén.

Szinte magam is megijedek, hogy a több dimenzióra épülő vizsgálatokat ennyivel „elintéztem”, de hosszas meggondolások után sem tudok mást mondani. A klasszikus tesztelmélet alkalmazása során, mint láttuk, a reliabilitás kiszámításával, és megfelelő szintjének konstatálásával véltük legitimáltnak az eljárásunkat, azonban láttuk, hogy a reprezentációs méréselmélet követelményei nem teljesíthetők, és a reliabilitás sem azt jelzi, hogy joggal végeznénk a számításokat. A modern tesztelmélet már megfelel a reprezentációs méréselméletnek, ám önmaga is eltántorít bennünket attól, hogy többdimenziós képességek esetén alkalmazzuk azt, ami csak egy dimenzió léte esetén lenne elfogadható. A „jó” teszt és a „jó” minta kialakításának trükkjével olyan teszteket és mintákat hozunk létre, amelyek esetén a fit-elemzések jó tulajdonságokat mutatnak, mint ahogy a reliabilitás értékek is jók voltak a másik esetben. De ez önbecsapás: ha arra törekszünk, hogy a fit elemzések kívánalmainak megfelelő szituációt teremtsünk, akkor sikerrel járunk. Ha viszont ki akarunk lépni ebből az önbecsapásból, és feladjuk az egyetlen dimenzió létére vonatkozó elképzelésünket, akkor – egyelőre legalábbis ez a helyzet – minden szofisztikált matematikai bűvészkedés ellenére sem vagyunk képesek megfelelő elemzést nyújtani, mert eljárásaink nagyon messze állnak a szinte végtelen komplexitás figyelembe vételétől.

E ponton szeretném leírni a tesztek alkalmazása során gyakran bekövetkező „probléma-eliminálás” sikerének egy szerintem rendkívül érdekes okát. Azt látjuk, hogy miközben komplex képességeket vizsgálnak a kutatók, a tesztfejlesztés eredményeként a legtöbb esetben nem is aránytalanul nagy munkával, nem sok itemet és vizsgálati „objektumot” eltávolítva, jó tulajdonságokkal rendelkező teszteket, vagyis lényegében egydimenziós teszteket tudnak létrehozni. Ennek az is oka lehet, hogy szinte óhatatlan, hogy a tesztbe már eleve egy a lehetségesnél jóval szűkebb körből választanak ki a kutatók feladatokat. Tipikus példa az iskolai szituációkban végzett fölmérések esete. Ha mondjuk a problémamegoldás képességét vizsgáljuk, vajon milyen problémák kerülnek elsősorban előtérbe? Nagyon valószínű, hogy inkább iskolás, vagy mondjuk a többségi társadalomban viszonylag ismertebb, hétköznapi problémák lesznek ezek. A tanulók bizonyos csoportjainak tudása eltérhet a többséghez tartozozókétól, más típusú problémákat, a saját élethelyzeteikben számukra ismertebb feladatokat tudnak nagyobb valószínűséggel megoldani. A feladatok, amelyeket a kutatók a tesztekbe beválogatnak, ezeket az eltéréseket nem tudják felszínre hozni, ehhez nincsenek, vagy nem megfelelő mértékben reprezentáltak a sajátos feladatcsoportok. Sokkal inkább érvényesül egy „normál hierarchia”, az, hogy a többségi társadalom tagjaitól elvárt, és bizonyos mértékig a tartalmukban már eleve homogén halmazt alkotó feladatok megoldásának valószínűségei úgy alakulnak, ahogy azt a Rasch modell megköveteli. Nem azért működnek – kis erőbedobással – a kialakított tesztek, mert a vizsgált képesség egydimenziós valamilyen elvont értelemben, hanem azért, mert az adott csoportba tartozó feladatok az adott populáción már egydimenziós struktúrát alkotnak egy szűken értelmezhető, a feladat sajátosságaihoz köthető „intelligencia” tekintetében. Ismert, hogy a tanulók között az iskolában is jó közelítéssel „abszolút hierarchiák” jönnek létre a tanulmányi teljesítményük tekintetében, kialakul egyfajta „sorrend” mondjuk az egy osztályba járó tanulók között, amely sorrend nem a vizsgált képesség tekintetében formálódott meg pusztán, hanem általánosan. *A* jobb tanuló, mint *B*, valójában tudható, hogy bármilyen iskolai, vagy a többségi kultúra keretei között jól ismertnek tekinthető feladatot *A* nagyobb valószínűséggel fog megoldani, mint *B*. A teszt valóban egy egydimenziós „valaminek” a fejlettségét méri, de ez a valami sokkal inkább a „normál tudás”, a többségi tudás ismerete, semmint az elvont képesség, az absztrakt pszichikus rendszer. Durván fogalmazva: a kutatások során alkalmazott tesztek nem valamifajta képességfejlettséget mérnek (hiszen az nincs is), hanem egy társadalmi folyamatokban formálódott hierarchiában elfoglalt helyet jeleznek. Így például a kutatások mindig kimutatják, hogy a rosszabb helyzetben lévő családok hátrányos helyzetű gyermekei a problémamegoldás képességét tekintve fejletlenebbek társaiknál. Az a feltételezésem, sőt, szilárd meggyőződésem, hogy az e tételt alátámasztó tesztek, az azok segítségével végrehajtott vizsgálatok azt mérik, hogy a gyerekek hol állnak egy társadalmi rangsorban. Itt egyáltalán nem a problémamegoldó képességről van szó. Ha arról lenne szó, akkor bekerülnének a tesztbe olyan feladatok is, amelyek például a szegénység körülményei között adódó problémák megoldásáról szólnának, szerepet kapnának a többségitől lényegesen különböző élethelyzetekben, vagy/és másfajta kultúrákban adódó problémák is. De ez nem így van, a tesztek a konszolidált középosztály kultúrájában jól ismert feladatokat tartalmazzák. És ez szükségképpen van így, mert ha a kapuk kinyílnának, és egészen más jellegű feladatok is bekerülnének, akkor a fit eljárások jeleznék a súlyos bajt, amit nem is lehetne orvosolni.

Együttes mérés

A feladat kijelölése

Mint már szerepelt korábban, az együttes mérés (conjoint measurement) az a másik eljárás a modern tesztelméleti módszerek mellett, amely reményt jelent a pszichológia (és így a pedagógia) számára az additív struktúrával nem rendelkező mennyiségek mérhetőségével kapcsolatban. Még egyszer: a probléma legelőször az volt, hogy a pszichológiában nincsenek konkatenálható „dolgok", a pszichológiában nem tudjuk a vizsgált tulajdonságaik szerint a „dolgokat" egymáshoz illeszteni folytatásként, vagy ugyanarra a serpenyőre helyezni. Két ember valamilyen tárgyra vonatkozó attitűdjéből nem lesz valami közös attitűd, bármilyen „konkatenáló tevékenységben" vesznek is részt, és két ember képességfejlettsége sem adódik össze semmilyen értelemben, semmilyen tevékenységük során (legalábbis ilyesmire a terület fejlődése során – ismereteim szerint – soha nem született javaslat). Úgy kell tehát a pszichológiának mérni, pszichológiai fogalmakkal leírt tulajdonságokhoz mértékeket (számokat) rendelni, hogy nem számíthat a klasszikus méréselmélet alapját jelentő értelmezésekre, mert, hogy azoknak nincsenek reprezentánsaik e tudomány (és persze ismét hangsúlyozzuk: a pedagógia) tarsolyában.

Stevens elmélete, a nyomában kialakuló reprezentációs méréselmélet, az annak megfelelő modern tesztelmélet kiutat mutat ebből a problémából, hiszen kiderült, hogy az is jó megoldás, ha „csak" intervallumskálán mérhetők az adatok. Ehhez az kell, hogy valamilyen kvalitatív módon meghatározott „távolság" legyen értelmezett a mért elemek között (a párjaik halmazán értelmezett legyen egy rendezési reláció). A klasszikus tesztelmélet nem tudott nyújtani ilyen értelmezést, a képességfejlettségek halmaza (másképpen a vizsgált személyek halmaza) nem lett felruházva olyan relációval, amely egy intervallumskála korrekt értelmezését lehetővé tenné. Ahogy már volt róla szó: egyelőre nincs ERR.

A modern tesztelmélet viszont megoldást kínál. Empirikusan meghatározható (becsülhető) valószínűségek hányadosának természetes alapú logaritmusát kiszámolva jutunk megfelelő relációhoz a képességfejlettségek, valamint a feladatnehézségek halmazán, és bebizonyítható, hogy amennyiben a Rasch-modell valóban érvényes az adott képességre (és ez empirikusan valóban be is látható), akkor ez a reláció megfelelő a reprezentációs méréselmélet követelményeinek teljesítéséhez.

Az, hogy a képességfejlettségek (és a feladatnehézségek) különbségi struktúrát alkotnak, vagyis megfelelnek a reprezentációs méréselmélet igényeinek, és alkothatók intervallumskálák, másképpen is belátható, segítségül híva az együttes mérés elméletét. Ugyanis az együttes mérés olyan „dolgokkal", azok tulajdonságaival kapcsolatos mérési eljárásokat nyújt, amelyek esetében közvetlenül csak ordinális skálákkal rendelkezünk. Csak rendezés ismert, legalábbis ezek az esetek tárgyalhatók itt, a számok intervallumskálát eredményező hozzárendeléséhez nincsenek közvetlenül megfelelő empirikus relációk. Az együttes mérés elmélete, ahogyan azt *Luce* és *Tukey* a 20. század '60-as éveinek elején megalkották (Luce és Tukey 1964) éppen azt teszi lehetővé, hogy csak rendezési struktúrával rendelkező tulajdonságértékek halmazán intervallumváltozót definiáljunk, és megadjuk a mérés menetét. Vagyis az együttes mérés elmélete valóban megoldást nyújt a számunkra, ha beválik, hiszen éppen azt a hiányosságot képes pótolni, ami a klasszikus méréselméletnek megfelelő elemzésben súlyos problémaként merült fel. Így az együttes mérés elmélete – megalkotói és hívei szerint – megoldja a pszichológia szinte valamennyi mérési problémáját, az olyan problémákat, mint hogy ki tudjuk ugyan mutatni, hogy valamely tárgy iránt valakinek erősebb attitűdje van, mint egy másik embernek, de nem tudjuk megmondani, hogy „mennyivel erősebb”. Vagy az olyan problémák is megoldhatók lesznek ezzel, hogy mértékeket, vagyis a különbségeikkel is információt hordozó számokat rendeljünk hozzá az emberek valamely képessége fejlettségéhez, és ez lehetővé tegye, hogy legitim módon, ne csak a számok rendelkezésre állása bűvöletében végezzünk legalább intervallumváltozókat igénylő matematikai vizsgálatokat (pl. szóráselemzést, faktoranalízist, klaszterelemzést, regressziós vizsgálatokat, stb.).

Az itt következő fejezetben bemutatom az együttes mérés modelljét három egymással kapcsolatban lévő változó esetében (ezen belül is egy bennünket érdeklő, speciális esetet vizsgálva), bemutatom, mit jelent a képességfejlettség mérésével kapcsolatban mindez, de azt is bemutatom, hol vannak az elmélet alkalmazásának határai, s erős kétségeket fogalmazok meg azzal kapcsolatban, hogy a pedagógiában szinte kizárólagosan szereplő, összetettebb képességekkel kapcsolatos mérésekre valójában az elmélet alkalmazható lenne.

Kiinduló megfontolások a Rasch-modellre alapozottan

Tegyük fel, hogy van három mennyiségünk, amelyek olyan viszonyban állnak egymással, hogy az egyik a másik kettőnek az összege. Ezt feltételezzük róluk, ami egyébként felveti annak a lehetőségét, hogy itt valójában egy „beledefiniálásról”, egy semmivel alá nem támasztható konstrukcióról van szó. De az elemzés érdekében ezt most tegyük meg, nyilván minden esetben konkrétan megfontolás tárgyává kell tenni, hogy vajon vannak-e jó indokaink egy ilyen feltételezés megfogalmazására.

A képességfejlettség mérésére vonatkozóan ez a modell első megközelítésben megfelelőnek látszik. A modern tesztelmélet (és persze azon belül a legegyszerűbb Rasch-modell) szerint a képességfejlettség, valamint a feladatnehézség ugyanazon a skálán mérhetők, különbségüknek a függvénye, hogy milyen valószínűsége van annak az eseménynek, hogy a β_ν képességfejlettséggel rendelkező személy jól meg tudja oldani a δ_λ nehézségű feladatot. Vagyis, ahogy ezt korábban már láttuk, a modellben a $\beta_\nu - \delta_\lambda$ különbségnek van fontos szerepe. A képlet azt adja meg – ahogy azt már többször is használtuk (Rasch 1960) –, hogy annak a valószínűsége, hogy a β_ν képességfejlettséggel rendelkező személy jól meg tudja oldani a δ_λ nehézségű feladatot:

$$p_{\nu\lambda} = \frac{e^{\beta_\nu - \delta_\lambda}}{1 + e^{\beta_\nu - \delta_\lambda}}.$$

A képletben a képességfejlettség és a feladatnehézség különbsége szerepel, ami kis zavart okozhat, hiszen az együttes mérés elméletében (legalábbis az additív modell esetén) két mennyiség összegéről van szó. A δ_λ feladatnehézség azonban olyan mennyiség, amelynek a negatívja (az ellentettje, a –1-szerese) az ellentétes tulajdonságot, vagyis a feladat könnyűségét jellemzi, ezért bátran mondhatnánk, hogy a $\beta_\nu - \delta_\lambda$ különbség nem más, mint a képességfejlettség és a feladatkönnyűség összege. Így aztán valóban két mennyiség összegéről van szó.

Érdemes a fenti képletet egy kis mértékben átalakítani. Osszuk el a számlálóban lévő kifejezéssel a számlálót is és a nevezőt is. Tudjuk, hogy ilyenkor a tört értéke nem változik. Ezt kapjuk (használva most már a feladat könnyűségét jellemző $\kappa = -\delta$ mennyiséget:

$$p_{\nu\lambda} = \frac{1}{e^{-(\beta_\nu + \kappa_\lambda)} + 1}$$

Határozzuk meg a képességfejlettség és a feladatkönnyűség összegét a feladat jó megoldása valószínűségének függvényeként (ezt is megtettük már korábban többször, csak azokban az esetekben a feladatnehézség paramétert használtuk). Ha némi átalakításokat végzünk, azt kapjuk, hogy

$$\beta_\nu + \kappa_\lambda = \ln \frac{p_{\nu\lambda}}{1 - p_{\nu\lambda}}.$$

A jobboldalon szereplő mennyiséget logit-nak is hívják (bár szokás a mennyiség mértékegységét is logitnak mondani). Értéke – amennyiben a $p_{\nu\lambda}$ valószínűség a 0 és az 1 között minden értéket felvesz (a 0-t és az 1-et kizárva) – mínusz végtelen és plusz végtelen között lehet. Ha például nagyon kicsi valószínűségről van szó, akkor az ln után álló tört számlálója nagyon kicsi, nevezője közel 1, vagyis a tört értéke nagyon kicsi, 0-hoz közeli, de akkor a természetes alapú logaritmusa nagy abszolút értékű negatív szám. Ennek megfelelően a képességfejlettség és a feladatkönnyűség összege nagy abszolút értékű negatív szám, ami előállhat például úgy, hogy mindkettő nagy abszolút értékű negatív szám, vagyis egy igencsak gyenge képességfejlettségű személlyel akartunk megoldatni egy igencsak nehéz feladatot. Nyilván kicsi a valószínűsége a jó megoldásnak. Ha viszont a valószínűség 1-hez közeli (nagyon valószínű a helyes megoldás), akkor a számláló 1-hez közeli, a nevező nagyon közel van 0-hoz, a tört értéke nagy pozitív szám, annak logaritmusa szintén nagy pozitív szám. A képességfejlettség és a feladatkönnyűség összege nagy pozitív szám, ami úgy is előállhat, hogy egy jó képességfejlettséggel rendelkező személlyel kívánunk megoldatni egy könnyű feladatot, nyilván nagy valószínűsége lesz a jó megoldásnak.

A logit, hasonlóan a $p_{\nu\lambda}$ valószínűséghez az esemény valószínűségét jellemzi, csak nem a szokásos, 0 és 1 közötti értékkel, hanem mintegy „kivetítve" ezt az intervallumot a mínusz végtelentől plusz végtelenig tartó teljes számhalmazra. A kis valószínűségnek nagy abszolút értékű negatív szám felel meg, a nagy valószínűségnek nagy pozitív szám felel meg, és a 0 logit (ami „pongyolán" fogalmazva középen van) a 0,5 valószínűségnek felel meg (ami szintén a valószínűségértékek intervallumának közepe).

Miért citáltuk ide (újra) ezeket a matematikai részleteket? Azért, hogy az olvasó lássa, itt valóban három mennyiség összefüggéséről van szó, ezek közül az egyik (a helyes megoldás valószínűsége) a másik kettő (a képességfejlettség és a feladatkönnyűség) összegének függvénye. Ez az oka annak, hogy az együttes mérésre vonatkozó különböző modellek közül a két változó összegét tartalmazót (additive conjoint measurement) vizsgáljuk meg közelebbről. Most visszatérünk az általános tárgyaláshoz, de ha szükség lesz rá, újból konkrétan, a feladatmegoldásokra vonatkozó összefüggésekkel szemléltetjük az eljárást.

A képességfejlettségek meghatározása

A következő elemzésben nem az együttes mérés precíz, axiomatikus matematikai elmélete alapján írom le a mérés logikáját, hanem fokozatosan, lépésről-lépésre kibontom az eljárást. Eközben bizonyos igények lépnek majd fel az alaphalmazokkal, az azokon definiálandó, szükséges relációkkal kapcsolatban. A leírásban ezeket az elvárásokat, szükséges összefüggéseket vastagon szedett szöveggel jelölöm, s majd a végén, mint az eljárás kivitelezésének előfeltételeit össze is foglalom. Ezeket a feltételeket kell teljesítenie egy mérésnek ahhoz, hogy az együttes mérés elmélete egyáltalán alkalmazható legyen rá. Előre jelzem – ez talán az eddigiek alapján gondolható –, az összetettebb, a komplexebb tevékenységekben realizálódó képességek fejlettségének mérésére vonatkozó ismert eljárások – meglátásom szerint – nem felelnek meg a követelményeknek. Állításom súlyos, és valójában nincs összhangban a szakirodalomban szereplő állítások döntő többségével. Még a pszichológiai mérések szokásos eljárásainak „legádázabb" kritikusai, *Michell* és *Barrett* is úgy tekintenek az együttes mérés elméletére, mint ami megoldás lehet a problémákra (Michell 1999, Barrett 2003). Hasonlóan másokhoz (pl. *Günter Tendlerhez* (2009) és *Norman Cliffhez* (1992)) én ezt nem így látom, ezért kell itt minél alaposabb érveket felsorakoztatnom.

Legyen három alaphalmazunk, az X, az Y és a Z, ezek azok a halmazok, amelyeknek elemei a mérendő entitások. Konkrét esetünkben X a vizsgált személyek halmaza, Y a feladatok halmaza, Z pedig azoknak az eseményeknek a halmaza, hogy adott vizsgált személyek jól oldják meg az adott feladatokat. E halmazokon egy-egy tulajdonság értelmezett, a konkrét esetben ezek rendre a képességfejlettség, a feladatkönnyűség, valamint az adott feladat és a vizsgált személy esetén a jó feladatmegoldás eseményének valószínűsége, illetve annak függvénye, a szintén a valószínűség jellemzésére alkalmas logit. Lesz majd egy ξ, egy η és egy ζ mennyiségünk, a tulajdonságok értékeihez hozzárendelt számok, amelyek viszonyáról az a feltételezésünk, hogy $\zeta = \xi + \eta$ (a mi konkrét esetünkben, tehát a feladatmegoldásokkal kapcsolatban a ζ a logit, ξ a képességfejlettség számértéke, η a feladatkönnyűség számértéke). Persze, egyelőre nem létezik semmilyen hozzárendelés, vagyis valójában nem tudjuk, hogy mely mért entitásokhoz milyen számokat lehet hozzárendelni, csak azt feltételezzük, hogy ha majd lesz már ilyen hozzárendelés, akkor a mennyiségek közt ez az összefüggés lesz érvényes.

Azokat az f_x, f_y, f_z függvényeket szeretnénk előállítani, amelyek az X, Y, Z halmazok elemeihez rendelik hozzá a számokat, vagyis formálisan: f_x: $X \rightarrow \boldsymbol{Re}$ (ahol $\boldsymbol{Re}$ a valós számok halmaza), vagy másképpen $f_x(x) = \xi$ ($x \in X$, $\xi \in \boldsymbol{Re}$), és az Y-beli y-okra, és a Z-beli z-kre is hasonlókat írhatunk.

Válasszuk ki az X halmaz egy tetszőleges elemét, és önkényesen rendeljük hozzá a 0 számot: $f_x(x_0) = 0$. Válasszuk ki a halmaz egy másik elemét tetszőlegesen. Arra lesz szükségünk, hogy ez a másik elem különbözzön a mért tulajdonsága tekintetében az előzőtől. Ez feltételezi, hogy **a vizsgált tulajdonsággal összefüggésben adott az *X* halmazon egy ekvivalencia reláció**, hiszen szükségünk van a mért tulajdonság szempontjából vett azonosság vagy különbözőség megítélhetőségére. Vagyis itt arról van szó, hogy az X halmaz elemeivel kapcsolatban tudjuk, hogy melyek azok, amelyek ugyanolyan tulajdonságúak (mondjuk megegyezik a képességfejlettségük, egyelőre anélkül, hogy ezt a mérték alapján állapítanánk meg, ez nyilván egy kvalitatív összefüggés, és ezért empirikusan ellenőrizhetőnek kell lennie). Éljünk ezzel a feltételezéssel! Van tehát egy másik elemünk az X halmazban, x_1, amelyhez önkényesen az 1 számot rendeljük hozzá ($f_x(x_1) = 1$)[37]. Most az Y halmazból vegyünk ki tetszőlegesen egy elemet, y_0-t, és

[37] Miért járhatunk el ilyen önkényesen? Láttuk, és a gyakorlatból is tudjuk, hogy ha egy mennyiség mérésével kapcsolatban intervallum-változót vagyunk képesek definiálni, akkor nem egyetlen skála lehetséges, hanem végtelen sok. Ezek egymástól egy lineáris transzformációban különböznek, és egy lineáris transzformáció megadásához éppen két tetszőleges valós szám megadására van szükség. Kevésbé elvontan! Gondoljunk a hőmérsékleti skálák definíciójára: két hőmérsékleti állapot esetén is a hőmérsékletet önkényesen határozhatjuk meg, a Celsius-skála esetén például önkényesen 0 °C-nak vesszük a víz és vízjég keverékének hőmérsékletét, és 100 °C-nak a forrásban lévő víz hőmérsékletét, ahogy ezt a példát már korábban is használtuk szemléltetésre.

rendeljük hozzá a 0 számot[38]: $f_y(y_0) = 0$. A Z halmaz elemei hozzá vannak rendelve az (x, y) párokhoz. Vagyis **létezik egy h: $X \times Y \rightarrow Z$ leképezés, amely kölcsönösen egyértelmű.** (Az $X \times Y$ a matematikában szokásos jelölés, az X és az Y halmazok elemeiből alkotott rendezett párok halmazát jelöljük így.) Például a feladatmegoldások esetében, ha választunk egy vizsgált személyt, valamint egy feladatot, akkor az, hogy a kiválasztott vizsgált személy sikeresen megoldja a kiválasztott feladatot, egy esemény, és az itteni jelölésben a Z halmaz egyik eleme, egyértelműen meghatározott, valamint ha megadjuk az eseményt, az csakis valamely konkrét személyhez, és egy konkrét feladathoz, annak helyes megoldásához köthető. Vagyis a tetszőlegesen kiválasztott x_0 és y_0 elemekhez adott egy és csak egy $z_{0,0}$, amelyhez, ha feltételezzük az additivitást, a 0 számot kell rendelnünk: $f_z(z_{0,0}) = f_x(x_0) + f_y(y_0) = 0 + 0 = 0$.

Most keressünk Z-ben egy olyan elemet, amely az (x_1, y_0) párhoz van hozzárendelve! Egy következő feltételezés, hogy **mindig lehet találni bármelyik halmazban a három közül olyan elemet, hogy ha adott a másik kettőből egy-egy elem, akkor közöttük érvényes az $(x, y) \rightarrow z$ összefüggés a *h* függvény szerint (vagyis $h(x, y) = z$).** Az angol nyelvű szakirodalomban ez a tulajdonság „solvability" néven szerepel, nevezzük ezt magyarul megoldhatóságnak. Jelöljük az (x_1, y_0) párhoz rendelhető Z-beli elemet $z_{1,0}$-val (vagyis $h(x_1, y_0) = z_{1,0}$). $z_{1,0}$-hoz nyilván az 1 számot kell rendelnünk, ha az additivitást akarjuk érvényesíteni: $f_z(z_{1,0}) = f_x(x_1) + f_y(y_0) = 1 + 0 = 1$.

Legyen most az Y halmazból y_1 olyan, hogy az (x_0, y_1) párhoz egy olyan $z_{0,1}$ elem legyen hozzárendelve h szerint, hogy a $z_{0,1}$ és a $z_{1,0}$ elemek ekvivalensek a vizsgált tulajdonság szempontjából. (A konkrét esetben a két logit, vagyis a két valószínűség megegyezik egymással.) Ez azt jelenti, hogy **a Z halmazon is léteznie kell egy ekvivalencia-relációnak.** Sőt, itt még azt is fel kell teleteznünk, hogy a Z-ben $z_{1,0}$-val ekvivalens elemek közt van olyan, amely (x_0, y_1)-hez van rendelve. Vagyis amikor kiválasztottuk tetszőleges módon az x_1 és y_0 elemeket, akkor egyben automatikusan kijelöltük azoknak a z elemeknek a halmazát, amelyek az (x_1, y_0) párhoz rendelt $(z_{1,0})$ elemmel ekvivalensek. Ezek közt kell lennie legalább egy olyannak, amely egy olyan párhoz van hozzárendelve, amelynek első eleme az x_0.

Hogy ez kicsit jobban érthető legyen, fordítsuk le a vizsgált személyek, feladatok „nyelvére"! Kiválasztottunk egy személyt, x_0-t, akinek a képességfejlettségét 0-nak tekintettük. Kiválasztottunk egy másikat, x_1-et, akinek a képességfejlettsége 1. Kiválasztottunk továbbá egy feladatot, y_0-t, amelynek a könnyűségét 0-nak vettük. Ahhoz az eseményhez, hogy az x_0 személy meg tudja oldani az y_0 feladatot, a 0 számot kell hozzárendelnünk (ez az esemény logitja). Most nézzük, hogy mit tud csinálni az x_1 személy az y_0-feladattal! Ha meg tudja oldani, akkor ez egy olyan esemény $(z_{1,0})$, amelyhez az additív szabály értelmében az 1 számot rendeljük hozzá. Most vegyük a másik ismert személyt, x_0-t! Kellene nekünk egy olyan y_1 feladat, amelyet ő ugyanakkora valószínűséggel tud megoldani, mint amekkora valószínűséggel x_1 személy megoldotta az y_0-feladatot. Vagyis ennek az egész gondolatmenetnek akkor van bármi értelme, ha annak,

[38] A gondos olvasó itt hibát vehet észre az eljárásunkban. Ha a képességfejlettség méréséről van szó, akkor a feladatok halmazából a 0 mértékkel rendelkező feladat kiválasztása már nem lehet önkényes. Ugyanis már korábban is láttuk, hogy ha valaki képességfejlettségének mértéke megegyezik az általa megoldott feladat nehézségparaméterének értékével, akkor a helyes megoldás valószínűségének 0,5-nek kell lenni, a logitnak 0-nak. Vagyis csak az a feladat lehet 0 mértékű, amelyet a legelsőnek kiválasztott személy 0,5 valószínűséggel tud megoldani, s ez csak a feladatok egy része esetén lehet érvényes. Élhetünk is akár ezzel a feltételezéssel a feladatok halmazából a nulla mértékű elem kiválasztásakor, és akkor nem lesz semmilyen ellentmondás, de azt is megtehetjük, hogy az ellentmondás elkerülése érdekében a feladatok halmazán a folyamatban definiált mértékeket egy lineáris transzformációnak vetjük alá, éppen úgy, hogy megfeleljünk a valószínűségek értékével kapcsolatos elvárásnak.

hogy két feladatmegoldást jelentő eseménynek ugyanannyi a valószínűsége (a két esemény ebből a szempontból ekvivalens egymással) van valamilyen definíciója és ezzel együtt empirikus ellenőrzési lehetősége. Erre természetesen vissza kell majd térnünk, ez egy kritikus pont. De előbb építsük ki az egész eljárást!

Formalizáljuk még jobban azt, amiről eddig szó volt. A Z halmazon érvényesül egy ekvivalencia reláció, jelöljük ezt E-vel. A Z egy elemét, z-t jelölhetjük a hozzá egyértelműen rendelhető x és y segítségével (x, y)-ként. Az, hogy a $z_{1,0}$ és $z_{0,1}$ ekvivalensek egymással, jelölhető kétféleképpen: $z_{1,0}Ez_{0,1}$, de ezt a jelölést is használhatjuk: $(x_0, y_1)E(x_1, y_0)$. Fogalmazzuk meg pontosabban, a szakirodalomban gyakrabban szereplő módon a megoldhatósági feltételt is! Legyen x_a, $x_b \in X$, y_c, $y_d \in Y$. Ha e négy elem közül három adott, akkor mindig létezik olyan negyedik, hogy $(x_a, y_c)E(x_b, y_d)$. Ez viszont mindig azt jelenti, hogy $f_x(x_a) + f_y(y_c) = f_x(x_b) + f_y(y_d)$. Ezek azok az összefüggések, amelyek lehetővé teszik, hogy további elemekről mutassuk meg, hogy valójában már ismerjük a vizsgált tulajdonságok szerinti mértékeiket.

Eddig már önkényes hozzárendelés, valamint logikai következtetések során sikerült definiálnunk illetve belátnunk, hogy $f_x(x_0) = 0$, $f_y(y_0) = 0$, $f_x(x_1) = 1$, $f_y(y_1) = 1$, $f_z(z_{0,0}) = 0$, $f_z(z_{1,0}) = 1$, $f_z(z_{0,1}) = 1$. A megoldhatóság tulajdonsága miatt van olyan $y_2 \in Y$, hogy $(x_1, y_1)E(x_0, y_2)$. A feladatmegoldások esetében ez azt jelenti, hogy ha az x_1 személy megoldotta az y_1 feladatot valamekkora valószínűséggel (a $z_{1,1}$ esemény tartozik a jó megoldáshoz), akkor találunk olyan y_2 feladatot, amelyet az x_0 személy ugyanakkora valószínűséggel old meg. Ha ez igaz, akkor viszont

$$f_x(x_1) + f_y(y_1) = f_x(x_0) + f_y(y_2).$$

De a baloldal értéke ismert,

$$f_x(x_1) + f_y(y_1) = 1 + 1 = 2,$$

továbbá a jobboldalon $f_x(x_0) = 0$, ezért $f_y(y_2) = 2$. Vagyis képesek voltunk egy újabb elemhez rendelhető számot megtalálni. Megtehetjük azt is, hogy olyan $x_2 \in X$ elemet keresünk, hogy $(x_2, y_0)E(x_1, y_1)$ legyen. (Most egy olyan vizsgált személy kereséséről van szó, aki az y_0 feladatot ugyanakkora valószínűséggel oldja meg, mint amekkora valószínűséggel az x_1 személy oldotta meg az y_1 feladatot.) De ha $(x_2, y_0)E(x_1, y_1)$ igaz, akkor

$$f_x(x_2) + f_y(y_0) = f_x(x_1) + f_y(y_1)$$

$$f_x(x_2) + 0 = 1 + 1$$

$$f_x(x_2) = 2$$

Felmerülhet a kérdés, hogy ha már ismertük az y_2 elemhez rendelt számot (2), akkor miért nem az (x_0, y_2) Z-beli elemet használtuk fel a számításhoz. Hiszen nem csak az igaz, hogy van olyan $x_2' \in X$, hogy $(x_2', y_0)E(x_1, y_1)$, hanem az is igaz, hogy van olyan $x_2'' \in X$, hogy $(x_2'', y_0)E(x_0, y_2)$. Nincs itt semmi probléma, megtehettük volna, hogy az $f_x(x_2)$ meghatározása érdekében nem az (x_1, y_1)-t használjuk, hanem az (x_0, y_2)-t. Az egyik módszer alkalmazása esetén kiválasztott x_2' –nek nem kell azonosnak lennie a másik módszer alkalmazása során kiválasztott x_2''-vel, az a lényeg pusztán, hogy ugyanaz legyen az új elemhez hozzárendelt szám, bármelyik eljárást alkalmazzuk is. Márpedig ezek azonosak, hiszen $f_x(x_1) + f_y(y_1) = f_x(x_0) + f_y(y_2) = 2$.

És most kereshetünk olyan y_3 elemet (feladatot), hogy $(x_2, y_1)E(x_0, y_3)$ legyen igaz. Ekkor

$$\begin{aligned} f_x(x_2) + f_y(y_1) &= f_x(x_0) + f_y(y_3) \\ 2 + 1 &= 0 + f_y(y_3) \\ f_y(y_3) &= 3. \end{aligned}$$

De kereshetünk olyan x_3 elemet (vizsgált személyt), hogy $(x_3, y_0)E(x_0, y_3)$ legyen igaz. Ekkor

$$\begin{aligned} f_x(x_3) + f_y(y_0) &= f_x(x_0) + f_y(y_3) \\ f_x(x_3) + 0 &= 0 + 3 \\ f_x(x_3) &= 3. \end{aligned}$$

Ennyiből talán látszik, és elfogadja a tisztelt olvasó (de természetesen ez korrekt módon bizonyítható is), hogy találhatunk (illetve részben kijelölhetünk) olyan $x_0, x_1, \dots, x_n, \dots \in X$, és $y_0, y_1, \dots, y_n, \dots \in Y$, elemeket, hogy érvényes mind $f_x(x_i) = i$, mind $f_y(y_j) = j$ ($i, j \geq 0$ egész számok), és $(x_i, y_{n-i})E(x_j, y_{n-j})$ (i, j, n nem negatív egész számok, és $n \geq i$, $n \geq j$). Van tehát egy-egy sorozatunk az X és az Y halmazokban, amelyek esetén a mérés nemnegatív egész számokat produkál. Ez már reményt kelt a tekintetben, hogy tetszőleges X-beli, illetve Y-beli elemekhez tudunk úgy hozzárendelni számokat, hogy az intervallumskála igényeit ki tudjuk elégíteni. De mielőtt ezt megmutatnám, érdemes még kiterjeszteni az eddig létrehozott skálát negatív irányban is. A megoldhatósági feltétel miatt kell lennie olyan x_{-1} elemnek X-ben, a már ismert x_0, y_0, y_1 elemekhez, hogy $(x_0, y_0)E(x_{-1}, y_1)$ igaz legyen. Ismét elvégezve a már jól ismert számítást:

$$\begin{aligned} f_x(x_0) + f_y(y_0) &= f_x(x_{-1}) + f_y(y_1) \\ 0 + 0 &= f_x(x_{-1}) + 1 \\ f_x(x_{-1}) &= -1. \end{aligned}$$

És ezzel ugyanazt a módszert alkalmazva, mint eddig, sorban előállíthatjuk az $x_{-1}, \dots, x_{-n}, \dots \in X$, és $y_{-1}, \dots, y_{-n}, \dots \in Y$, elemeket, amelyeket egyesítve az előbbi sorozatokkal kimondhatjuk, hogy vannak az X és az Y halmazokban olyan elemek, amelyekhez a mérés során az egész számokat kell hozzárendelnünk.

És akkor most gondoljuk át, hogyan kellene hozzárendelnünk egy számot az X halmaz egy tetszőleges eleméhez, és ugyanígy az Y halmaz valamely eleméhez! Ennek megmutatására egy analógiát fogok használni (ahogyan ez rendszerint az ezzel foglalkozó szakirodalomban is történik, ld. pl. Bouyssou és Pirlot 2005).

Nézzük meg, hogyan történik a tömeg mérése! Az eljárást a hossz mérésével kapcsolatban már korábban leírtam, amikor a reprezentációs méréselméletnek megfelelő eljárásokra mutattam példát. Nem árt most feleleveníteni ezt a gondolatsort, illetve egy másik példát is megismerni. A tömeg méréséhez – ahogy erről már korábban is volt szó –

az „egyenlő tömegűek”, „az egyik test nagyobb tömegű, mint a másik” relációk értelmezésére és empirikus vizsgálhatóságára van szükségünk. Ezt az egyenlő karú mérleg alkalmazása teszi elérhetővé, a jól ismert módon. Bármely két testről meg tudjuk állapítani (legalábbis a könnyebben kezelhető testek esetében), hogy egyenlő-e a tömegük, s ha nem, akkor melyiké a nagyobb. Kiválaszthatunk egy testet, amelynek egységnyinek tekintjük a tömegét (a Párizs mellett, egy levéltárban őrzött őskilogramm tömege), és az is világos, hogy elő tudunk állítani akárhány olyan testet, amelynek a hibahatáron belül szintén egységnyi a tömege (megegyezik az őskilogramm tömegével). Ha most van egy ismeretlen tömegű (de a mérlegre ráhelyezhető) testünk, akkor megnézhetjük, hogy véletlenül nem igaz-e, hogy valahány egységnyi tömegű testet helyezve a mérleg másik serpenyőjébe, egyensúly alakul ki. Itt a tömegmérés esetén kihasználható összegződésről van szó, vagyis arról, hogy ha egyesítünk több testet (egyszerre tesszük őket a mérleg egyik serpenyőjébe), akkor az így keletkezett egyesített test tömege az egyes testek tömegének összege lesz. Ha a mérendő tömegű test éppen kiegyensúlyozható néhány, mondjuk *n* db. 1 kg tömegű testtel, akkor az összegződés miatt a mért tömeg *n* kg lesz. De ilyen helyzet csak nagyon ritkán fordulhat elő. Egy tetszőleges test tömegét mindig valamilyen pontossággal tudjuk megmérni, egy becslést, közelítést kapunk.

Ennek módszere a kétkarú mérleg gyakorlati használata esetén (pl. piacon, vagy egy kémiai laboratóriumban), hogy rendelkezésre állnak megfelelő mennyiségben olyan tömegű testek, az ún. súlysorozat elemei, amelyeknek a tömegei lehetővé teszik a pontosság figyelembevételével és bizonyos határok közt mindenféle tömegérték megmérését. Nekünk azonban egy elvi definícióra van szükségünk, nem feltételezhetjük egyelőre, hogy vannak 10 dkg-os, 5 dkg-os, 1 dkg-os „súlyok”. A mérés így is lehetséges. A módszer az, hogy készítünk a mérendő tömegű testtel egyező tömegű testeket, méghozzá *k* db-ot (*k* lehet akármilyen pozitív egész szám). Ha a mérendő tömeg (amit egyelőre nem ismerünk) *m* kg, akkor az összegzési szabály miatt a *k* db. vele egyező tömegű test egyesítésével (az egyik serpenyőre való helyezésével) elő tudunk állítani egy *km* tömegű testet. Most rakjunk sorban egymás után 1 kg tömegű testeket a mérleg másik serpenyőjébe, s figyeljük, hogy lebillen-e a mérleg. Előáll egy olyan helyzet, hogy a *K*-adik 1 kg tömegű test odarakásánál még nem billent le a mérleg, azonban amikor még egy ilyen testet rárakunk a serpenyőre, a mérleg lebillen, jelezve, hogy most már az 1 kg-os testek serpenyőjén egy a másik test tömegénél nagyobb tömegű test van. Ez azt jelenti a tömegek vonatkozásában, hogy

$$K \times 1\ \text{kg} \leq km\ \text{kg} < (K+1) \times 1\ \text{kg}.$$

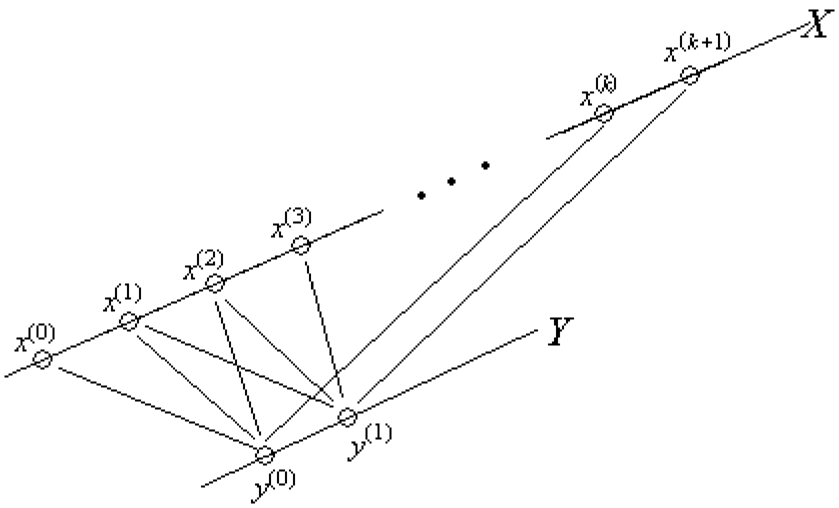
Ha elhagyjuk a mértékegységet, az 1-gyel való szorzást, és elosztjuk az egyenlőtlenségek minden oldalát *k*-val, akkor ezt kapjuk:

$$\frac{K}{k} \leq m < \frac{K+1}{k} = \frac{K}{k} + \frac{1}{k}$$

Vagyis ha a mérendő tömeget (pontosabban annak mértékét), az *m*-et a *K*/*k* hányadossal becsüljük, akkor biztos, hogy 1/*k* kg-nál kevesebbet tévedünk (hiszen sikerült egy ilyen hosszúságú intervallumba „beszorítani” a meghatározandó értéket). Mivel *k*-t magunk választhatjuk meg, akármilyen előírt pontossággal meghatározhatjuk a tömeget (legalábbis a mérleg pontossága felett, hiszen ha túl nagy a *k*, akkor lehet a *K* már olyan nagy, hogy a mérlegnek le kellene billennie, ezt azonban nem teszi, mert ilyen kicsi különbségeknél a saját pontossági korlátaiból következően még nem tud lebillenni).

Valami ilyesmit kellene az együttes mérésnél is véghezvinnünk. Legyen x a mérendő entitás az X halmazban. $f_x(x)$-et kellene meghatároznunk, úgy, hogy ne kerüljünk ellentmondásba a már eddig ismert mértékekkel. Természetesen könnyű a dolgunk, ha igaz valamilyen már ismert $x_i \in X, y_j \in Y$ (i, j egész számok) elemekre, hogy $(x, y_j)E(x_i, y_j)$, hiszen ez csak úgy lehetséges, ha $f_x(x) = f_x(x_i) = i$. De mint a tömegmérésnél, ilyen „szerencsénk" csak ritkán lehet.

Előállítunk viszont egy sorozatot az X elemeiből, amelynek a nulladik eleme maga a zéruspont, tehát x_0 lesz, ez itt $x^{(0)}$–val lesz jelölve. A sorozat első eleme az, amelynek a mértékét akarjuk megállapítani, tehát x, de a sorozatban az $x^{(1)}$ jelölést kapja, és a mértékét jelöljük most a-val, vagyis $f_x(x) = f_x(x^{(1)}) = a$. A sorozat második eleme egy olyan $x^{(2)}$ elem lesz X-ből, amelyre viszont $f_x(x^{(2)}) = 2a$, és így tovább. A sorozat i. eleme egy olyan $x^{(i)}$ elem lesz X-ből, amelyre $f_x(x^{(i)}) = ia$. Hogy ez valóban megtehető, és tényleg léteznek ezek az elemek, az a következő eljárással látható be (a megértést a 6. ábra segíti):



6. ábra: az X *és* Y *halmazok szemléltetése egy-egy egyenessel*

A már korábban megismert y_0 itt $y^{(0)}$ jelölést kap. Legyen továbbá $y^{(1)}$ egy olyan elem Y-ból, hogy $(x^{(0)}, y^{(0)})E(x^{(1)}, y^{(1)})$ legyen. Ilyen elem létezik a megoldhatósági feltétel teljesülése esetén. Írjuk fel, mit jelent az ekvivalencia reláció az elemekhez rendelt számokkal:

$$f_x(x^{(0)}) + f_y(y^{(0)}) = f_x(x^{(1)}) + f_y(y^{(1)})$$

$$0 + 0 = a + f_y(y^{(1)})$$

$$f_y(y^{(1)}) = -a.$$

Most legyen $x^{(2)}$ egy olyan elem X-ből, hogy $(x^{(1)}, y^{(0)})E(x^{(2)}, y^{(1)})$ legyen. Ilyen elem létezik a megoldhatósági feltétel miatt. Ismét felírva a számok közti egyenlőséget:

$$f_x(x^{(1)}) + f_y(y^{(0)}) = f_x(x^{(2)}) + f_y(y^{(1)})$$

$$a + 0 = f_x(x^{(2)}) - a$$

$$f_x(x^{(2)}) = 2a.$$

Vagyis tényleg sikerült megtalálnunk a fent előre jelzett sorozat második elemét. Még egy lépést megmutatok: legyen $x^{(3)}$ egy olyan elem X-ből, hogy $(x^{(2)}, y^{(0)})E(x^{(3)}, y^{(1)})$ legyen. Ilyen elem létezik a megoldhatósági feltétel miatt. A számok közti egyenlőség:

$$f_x(x^{(2)}) + f_y(y^{(0)}) = f_x(x^{(3)}) + f_y(y^{(1)})$$

$$2a + 0 = f_x(x^{(3)}) - a$$

$$f_x(x^{(3)}) = 3a.$$

Ezen a módon tehát tényleg előállítható a fent jelzett sorozat. Eljuthatunk egy előre megadott k pozitív szám esetén az $x^{(k)}$ elemig, amelyre $f_x(x^{(k)}) = ka$. De miért jó ez a sorozat? Azért, amiért jó volt a tömegmérés esetén az m, $2m$, …, km sorozat. Ott megpróbáltuk az $1, 2, \ldots, K, K + 1$ kg tömegű súlyokkal kiegyensúlyozni a km tömegű testet, a K. lépésnél még nem nyomták le a testek a mérleg serpenyőjét, a $K + 1$-edik esetben viszont már igen. Itt is valami nagyon hasonlót kell mondanunk: az x_0, x_1, …, x_K, elemek (ezeket definiáltuk korábban, mértékeik rendre 0, 1, …, K voltak) mind megelőzik $x^{(k)}$–t, x_{K+1} viszont már követi azt. De vajon mit jelent ez a „megelőzés” és „követés”?

Itt szükségünk van egy nagyon fontos feltétel érvényesülésére. Az elemeknek egy monoton növekvő sorozatot kell alkotniuk, abban az értelemben, hogy az X halmazon kell léteznie valamilyen rendezési relációnak. Az együttes mérés elméletében az X és az Y halmazok rendezését a Z halmaz rendezéséből eredeztetjük. Vagyis léteznie kell a Z halmazon egy rendezésnek, legyen ez a reláció – használtuk már ezt a jelölést – a $\gtrsim$ jellel jelölve, úgy, hogy a bal oldalára írt Z-beli elem nem követi a jobb oldalára írt elemet. Ennek a relációnak tehát eleve adottnak kell lenni, nyilván meg kell majd vizsgálnunk, hogy a képességfejlettség mérése során számíthatunk-e a meglétére, van-e értelmes definíciója a helyes feladatmegoldások valószínűségei közti egyenlőtlenségnek.

Tegyük fel, hogy olyan Z halmazzal van dolgunk, amelyen értelmezett a $\gtrsim$ rendezési reláció. Most az a kérdés, hogy az X halmaz tetszőleges két elemét, az x_1, x_2 elemeket tudjuk-e rendezni, találunk-e értelmes definíciót arra, hogy az egyik nem követi a másikat. Az együttes mérés elméletében pontosan bizonyítható (ld. Luce és Tukey 1964), hogy matematikailag következetesen ez az egyenlőtlenség csak a következő módon definiálható: jelöljük az X halmazon értelmezendő egyenlőtlenségi relációt a $\gtrsim_X$ jellel. Akkor mondjuk, hogy $x_1 \gtrsim_X x_2$, ha minden $y \in Y$ elemre igaz, hogy $(x_1, y) \gtrsim (x_2, y)$. Ez nagyon fontos pont. **Az együttes mérés elmélete csak akkor alkalmazható, ha magán az X halmazon (és persze ugyanígy az Y halmazon is) az itt leírt értelemben értelmezhető egy rendezési reláció**. Ismét fogalmazzunk a képességfejlettség mérés során használt terminusokkal: akkor mondhatjuk, hogy az egyik ember képességfejlettsége (x_1) magasabb szintű, mint a másik ember képességfejlettsége (x_2), ha bármelyik feladat kiválasztása esetén igaz, hogy az első ember nagyobb valószínűséggel tudja azt megoldani, mint a másik ember. Csak olyan esetekben szabad alkalmazni az együttes mérés elméletét, amelyekben ez a feltétel érvényesül. Itt ragadhatjuk majd meg azt a problémát, amely az együttes mérés alkalmazhatóságával (és vigyázat, nem magával az együttes méréssel) kapcsolatos, illetve látjuk, hogy valójában a Rasch-modell alkalmazhatóságával kapcsolatban is alapvető jelentőséggel bíró monotonicitásról van szó.

De mielőtt a kritikára térnénk rá, vigyük végig a gondolatmenetet, mutassuk be, végül is miképpen tudjuk hozzárendelni a megfelelő számot az X halmaz x eleméhez! Ahogy az előbb már megmutattam, az x_0, x_1, …, x_K, elemek mind megelőzik $x^{(k)}$–t, x_{K+1} viszont már követi azt, vagyis $f_x(x_K) < f_x(x^{(k)}) < f_x(x_{K+1})$, vagy a konkrét értékekkel: $K < ka < K + 1$. És ez a tömegméréshez hasonlóan azt jelenti, hogy a-t becsülhetjük, közelítőleg meghatározhatjuk a K/k törttel, s ha ennél az $1/k$ pontosságnál jobbat akarunk, akkor növelhetjük a k értékét.

Most az X halmaz elemeivel kapcsolatos mérésre mutattam meg a mérési folyamatot. Természetesen az Y halmaz elemeivel kapcsolatban a mérés ugyanilyen lenne, a két halmaz szerepe abszolút szimmetrikus. Ha már tudjuk mérni az X és az Y halmazok elemein vizsgált tulajdonságok mértékét, akkor bármelyik Z-beli elem vizsgált tulajdonságának mértéke is adott, hiszen csak meg kell nézni, mely X-beli és mely Y-beli elemek párjaként áll elő az adott Z-beli elem, és a már azokra meghatározott mértékeket össze kell adni.

Intervallumskálát alkottunk

A fentiekben azt láttuk, hogy ha két halmazon akarjuk mérni a halmazok elemei adott tulajdonságainak mértékét, és megbízhatunk abban a modellben, hogy e mértékek összege egyenlő egy harmadik változó értékével (most kissé pontatlanul fogalmaztam a rövidség kedvéért), akkor skálák jönnek létre. De milyen skálák ezek? A folyamat során voltak önkényes elemek, méghozzá az X halmazból a 0 és 1 mértékkel, valamint az Y halmazból a 0 mértékkel rendelkező elemek kiválasztása[39]. Vagyis az X és az Y halmazokon a skála kezdőpontját és az egységet önkényesen határoztuk meg, utána azonban a többi már (ha feltételezzük valóban az additivitást) determinált, az elemekhez meghatározott számokat kell rendelnünk. Az X halmaz elemeihez ξ-ket rendelünk, és ha már megtörtént ez a hozzárendelés, akkor az összes értékre alkalmazhatunk bármilyen lineáris transzformációt, s ez a művelet egy újabb legitim skálát eredményez. Vegyünk egy tetszőleges, nem 0 valós számot, mondjuk s-t, és egy másik tetszőleges valós számot, mondjuk r-t, és az x-ek mértékeit, a ξ-ket transzformáljuk ekképpen: $\xi' = s\,\xi + r$. A ξ'-k is az x-ekhez hozzárendelt mértékek, csak a skála „össze van húzva” vagy „meg van nyújtva”, és valamilyen irányban el van tolva. Éppen azokat a skálákat nevezzük intervallum-skáláknak, amelyeknél ez a transzformáció viszi egymásba az egyes skálákat, s ezen a körön kívül nincs más transzformáció közöttük.

De vajon csak a lineáris transzformációk ilyen szerepe miatt tekinthetjük intervallum-skáláknak a létrejött hozzárendeléseket? Ez azért fontos kérdés, mert önmagában az, hogy a skálák halmazának elemeit lineáris transzformációkkal és csak azokkal lehet egymásba vinni, bizony még nem sokat ér. Képzeljük el, hogy van egy halmazunk, amelynek elemeivel kapcsolatban nincs semmilyen információnk. Tegyük fel, hogy készítünk egy teljes mértékben tetszőleges, egy-egyértelmű hozzárendelést a halmaz elemei és a valós számok között (legyen mondjuk kontinnum számosságú, tehát a valós számok halmazával megegyező számosságú a halmazunk). Most állítsunk elő ebből az egyből további skálákat, úgy, hogy az összes lehetséges módon lineáris transzformációkat alkalmazunk erre az előre létrehozott skálára. Könnyű bebizonyítani,

---

39 De emlékeztetek arra, hogy valójában a 0 mértékű feladat kiválasztása nem teljesen önkényes, a feladathalmaz egy olyan részhalmazából történhet csak, amely feladatokat a 0 képességfejlettség mértékűnek választott vizsgált személy 0,5 valószínűséggel tud megoldani. A megoldhatósági feltétel érvényesülése esetén azonban ennek a választásnak nincs akadálya.

hogy ekkor bármely két skála közt egy lineáris transzformáció, és csak az teremti meg a kapcsolatot. Akkor most intervallum-skálát hoztunk létre? Sőt, rögtön végtelen sokat? Szó sincs róla. Az, hogy a létrehozott skálák közt lineáris transzformáció működjék, egy szükséges, de nem elégséges feltétele annak, hogy a mondott skálák intervallumskálák legyenek. Ahogy már láttuk, és a Rash modell esetében részletesen is megvizsgáltuk korábban: kell, hogy legyen egy a mért elemek közti kapcsolatot leíró, kvalitatív módon értelmezett „távolság” fogalom. E fogalom segítségével értelmezettnek kell lenni egy a halmaz elempárjai közti rendezési relációnak, valamilyen meghatározásnak kell lennie arra, mikor nagyobb két elem „távolsága” másik két elem „távolságánál”. Soha nem „úszhatjuk meg” a megfelelő empirikus relációk megtalálását. Mesterségesen létrehozhatunk a transzformációs feltételt kielégítő skálákat, ha viszont nincs semmi a tapasztalati világunk leképezésére posztulált, matematikai struktúrában (ERR), aminek megfelelnek a számok közti különbségek, illetve nincs semmi, aminek e különbségek nagyság szerinti rendezése megfelel, akkor csak játszadozásról van szó.

Minek felelnek meg az együttes mérés eljárása esetén a létrejövő intervallumskálán mért értékek különbségei? A fent definiált $x^{(0)}, x^{(1)}, \ldots, x^{(n)}, \ldots$ sorozatban bármely két, egymással szomszédos elemre igaz a konstrukció módja miatt a következő: $(x^{(i)}, y^{(0)})E(x^{(i+1)}, y^{(1)})$, vagyis

$$f_x(x^{(i)}) + f_y(y^{(0)}) = f_x(x^{(i+1)}) + f_y(y^{(1)})$$

$$f_x(x^{(i)}) + 0 = f_x(x^{(i+1)}) - a$$

$$f_x(x^{(i+1)}) - f_x(x^{(i)}) = a.$$

Vagyis a sorozat bármely két szomszédos eleméhez hozzárendelt számok közti különbség a. Vagyis itt megleljük az „intervallumhosszak” egyenlőségét, s azt is látjuk, hogy akkor lesznek ezek az „intervallumok” azonos hosszúságúak, ha mindegyik megfeleltethető az $[y^{(0)}, y^{(1)}]$ „intervallumnak” abban az értelemben, hogy bármely i-re, ahogy fentebb láttuk: $(x^{(i)}, y^{(0)})E(x^{(i+1)}, y^{(1)})$.

Egy talán még inkább érthető, és egyben az előbbinél általánosabban érvényes fogalmazásban: ha az x_a és az x_b elemekhez rendelt számok közti különbség ugyanannyi, mint az $x_{a'}$ és az $x_{b'}$ elemekhez rendelt számok közti különbség, vagyis

$$f_x(x_a) - f_x(x_b) = f_x(x_{a'}) - f_x(x_{b'}),$$

a mögött az a tény húzódik meg, hogy a következő két reláció egyszerre áll fenn:

$$(x_a, y_c)E(x_b, y_d) \text{ és}$$

$$(x_{a'}, y_c)E(x_{b'}, y_d)$$

alkalmasan megválasztott y_c és y_d elemekkel. De minden további nélkül mondhatjuk azt is, hogy ha az x_a és az x_b elemekhez rendelt számok közti különbség nagyobb, mint az $x_{a'}$ és az $x_{b'}$ elemekhez rendelt számok közti különbség, vagy egyenlők, vagyis ha

$$f_x(x_a) - f_x(x_b) \geq f_x(x_{a'}) - f_x(x_{b'}),$$

a mögött az a tény húzódik meg, hogy a következő két reláció egyszerre áll fenn:

$$(x_a, y_c) \gtrsim (x_b, y_d) \text{ és}$$

$$(y_c, x_{a'}) \gtrsim (y_d, x_{b'})$$

Vagyis az egyik változó értékei mintegy referenciát jelentenek a másik változó „intervallumainak" „azonos hosszával" illetve „nagyobb vagy egyenlő hosszával" kapcsolatban. Az, hogy a megfelelő relációk fennállnak-e, empirikus kérdés, a feladatokkal kapcsolatban például annak a kérdése, hogy a megfelelő valószínűségek megegyeznek-e egymással, illetve, hogy fennállnak-e a megfelelő rendezési relációk.

Nagyon fontos a kapott eredmény. Gyönyörű ez a konstrukció, melyet *Luce* és *Tukey* megalkottak. Olyan mennyiségekből indulunk ki, amelyekkel kapcsolatban még csak elképzelésünk sincs kezdetben, hogyan alkotunk meg rájuk intervallumskálákat. Néhány feltételezéssel kell csak élni, és az intervallumskálák létrejönnek. Mérhetővé válnak valamik, amik a folyamat kezdetén nem tűntek mérhetőknek, legalábbis nehéz volt elképzelni, hogyan jöhetnek létre intervallumskálák csak annyi információ alapján, ami rendelkezésünkre áll.

Az alkalmazhatóság

A modern tesztelmélet (annak legegyszerűbb esete, a Rasch-modell) a vizsgált személyek képességfejlettsége, mint egyik változó, a feladatok könnyűsége, mint másik változó, és a helyes feladatmegoldás valószínűségéből képezett logit között éppen az együttes mérés tárgyalása során használt összefüggést (összegzés) feltételezi. Ez azt jelenti, hogy ha a képességfejlettség mérése megfelel a folyamat közben felhasznált feltételeknek, akkor a képességfejlettség mérésével kapcsolatban valóban ki tudunk alakítani intervallumskálákat, méghozzá a képességfejlettségre és a feladatkönnyűségre vonatkozóan ugyanazon skálán helyezhetők el az adatok. Az intervallumskála létezését – mint láttuk – az együttes mérés elméletéhez kötött eljárás biztosítja. De nézzük végig, hogy e konstrukció során milyen feltételezésekkel kellett élnünk!

1. Mindhárom halmaz elemein léteznie kell egy-egy az adott halmazon vizsgált tulajdonsággal kapcsolatos ekvivalencia-relációnak.
2. A Z halmaz elemein, konkrétan a helyes feladatmegoldások eseményein léteznie kell egy a vizsgált tulajdonsággal (ez az események valószínűsége) kapcsolatos rendezési relációnak.
3. Az X és az Y halmaz elemein igaznak kell lenni, hogy ha $x_1, x_2 \in X$ tetszőleges elemek, akkor minden $y \in Y$-ra $(x_1, y) \gtrsim (x_2, y)$, vagy $(x_2, y) \gtrsim (x_1, y)$, de minden y elem esetén ugyanaz, és ha $y_1, y_2 \in Y$ tetszőleges elemek, akkor minden $x \in X$-re $(x, y_1) \gtrsim (x, y_2)$, vagy $(x, y_2) \gtrsim (x, y_1)$, de minden x elem esetén ugyanaz.
4. Érvényesnek kell lenni a megoldhatósági kritériumnak, ahogy azt fentebb bemutattam (ha $x_a, x_b \in X$, és $y_c, y_d \in Y$, és e négy elem közül három adott, akkor mindig létezik olyan negyedik, hogy $(x_a, y_c)E(x_b, y_d)$).

Itt egy az értelmezés szempontjából nagyon fontos megjegyzést kell tennem. Felmerül ugyanis a kérdés, hogy e feltételek vajon milyen értelemben feltételek. Nézzük például az 1. követelményt! Ha a teljes egészében absztrakt, matematikai struktúraépítésben gondolkodunk, akkor e feltétel érvényesítésének semmilyen akadálya nincs, absztrakt, közelebbről nem meghatározott halmazokon nincs akadálya sok-sok ekvivalencia-reláció értelmezésének. És a másik három feltétel is teljesíthető absztrakt módon, nagyon könnyen. De itt nem arról van szó, hogy létezhet-e ilyen matematikai struktúra, amely e három halmazból és a rajta definiált relációkból áll. Hanem arról, hogy a konkrét alkalmazás esetén, esetünkben a képességmérés során tudunk-e ajánlani olyan empirikus eljárásokat, amelyek reális összefüggéseket, viszonyokat, műveleteket, stb. feleltetnek meg az absztrakt struktúra elemeinek, és ezzel azt, amit az elvont struktúrára kijelentünk, le tudjuk fordítani az empirikus rendszerre, és a kijelentésben szereplő jelenség bekövetkezését, vagy egy összefüggést, egy viszonyt empirikusan ellenőrizni tudunk. Hogy a valószínűleg mindenki számára jobban átlátható hosszúságméréssel szemléltessünk: nem az az érdekes, hogy szakaszok közt lehet-e definiálni egyenlőségi (ekvivalencia) relációt, hanem az, hogy van-e olyan empirikus eljárás, amellyel ellenőrizni tudjuk, hogy két szakasz, két rúd, stb. egyenlő hosszúak-e egy kvalitatív értelemben.

Könnyen megállapíthatjuk, hogy elvileg a 4. feltétel kielégítése, illetve az előbbi megjegyzés szerint a feltételnek megfelelő empirikus műveletek megtalálása nem jelenthet gondot. Ha adott két feladat, és adott egy vizsgált személy, akkor nyugodtan feltételezhetjük, hogy találunk egy másik személyt, hogy az első személy az első feladatot ugyanakkora valószínűséggel legyen képes megoldani, mint a talált személy a másik feladatot. Magát a műveletet, tehát a konkrét személy megkeresését empirikusan nem kell elvégeznünk, itt csak az elvi lehetőségnek kell rendelkezésre állnia. A modern tesztelmélet az itt kifejtett háttérrel egy másik műveletsort ad meg a képességfejlettség és a feladatnehézség értékeinek becslésére. Az is igaz nyilván elvileg, hogy ha van két vizsgált személy, egy feladat, akkor találunk biztosan egy másik feladatot, hogy az első személy az első feladatot ugyanakkora valószínűséggel legyen képes megoldani, mint a második személy a talált feladatot. A megoldhatósági kritérium teljesítése körül tehát nincs különösebb elvi probléma. Annál inkább a másik három kritérium esetében.

Az 1-3. követelmények teljesítése azon múlik, hogy képesek vagyunk-e empirikus eljárásokat javasolni arra, hogy a jó feladatmegoldással kapcsolatos események valószínűségei közti egyenlőséget és kisebb-nagyobb viszonyokat definiáljuk. Péter megoldott helyesen egy matematika feladatot, Klári megoldott helyesen egy szövegértési feladatot, vajon miképpen értelmezzük kvalitatív módon, hogy e két esemény valószínűsége megegyezik, vagy azt, hogy az egyik nagyobb, mint a másik? E kérdésre valójában már válaszoltam a Rasch-modellel kapcsolatos elemzés során. Ha sok feladatot oldatunk meg sok személlyel, akkor a megoldások elemzése alapján kijelölhetők azok a feladatok, amelyeket ugyanannyian oldottak meg, vagyis úgy tűnik, ezek nehézsége azonos lehet. Így azt is feltételezhetjük, hogy e feladatokat egy adott vizsgált személy ugyanolyan valószínűséggel oldotta meg helyesen. Meg tudjuk állapítani, hogy két esemény, esetünkben két helyes feladatmegoldás közül melyiknek a bekövetkezési valószínűsége nagyobb (vagy esetleg egyenlők a valószínűségek):

- Sok feladatot kell megoldatni sok vizsgált személlyel.
- Ki kell választani mindkét feladat esetén azokat a hozzájuk tartozó további feladatokat, amelyeket ugyanannyian oldottak meg, vagyis meghatározzuk a két kiválasztott feladattal megegyező nehézségű feladatok két halmazát.
- Majd megnézzük, hogy az első kiválasztott személy az általa megoldott feladathoz tartozó feladathalmazból mennyit oldott meg helyesen.
- Ugyanezt elvégezzük a másik személyre is.
- E két szám egymáshoz való viszonya alapján döntünk a feltett kérdésben.

Minden mérés lehet hibás, ebben az eljárásban is megvan a hibázás lehetősége, ez azonban természetes. Esetleges tényezők hathatnak a feladatok megoldásának konkrét sikerességére. Világos, hogy az eljárás magában hordozza a hibák elkövetésének lehetőségét. De elvileg a feladatszám és a minta növelésével a tévedések valószínűsége csökken, javítható a biztonság. Elfogadhatjuk tehát, hogy a jó feladatmegoldásoknak, mint eseményeknek a valószínűségével kapcsolatban rendelkezünk kvalitatív eljárással az egyenlőség és a kisebb-nagyobb viszonyok megállapítása tekintetében. A gondjaink akkor kezdődnek, amikor a vizsgált személyek, valamint a feladatok halmazán kívánjuk értelmezni az ekvivalencia- és a rendezési relációkat.

A fenti felsorolásból a 3. feltétel érvényesülése kritikus szerepet játszik. Miről is van szó? Két ember között rendezést tudunk megállapítani a vizsgált képességük fejlettségének tekintetében, ha igaz, hogy egyikőjük bármely feladatot nagyobb valószínűséggel tud megoldani, mint a másik. Márpedig – és ezt az érvet már többször használtam – ez éppen a pedagógiát érdeklő komplex képességek esetén nem teljesül. Vagyis az együttes mérés, valamint a modern tesztelmélet alkalmazhatóságának egyik feltétele most egyazon kérdéssé válik: lehetséges-e az X (és az Y) halmazon úgy értelmezni a rendezési relációt, hogy akkor mondjuk, hogy egy vizsgált személy képességfejlettsége nagyobb, mint egy másik vizsgált személy képességfejlettsége, ha bármely feladat esetén igaz, hogy az első személy nagyobb valószínűséggel tudja azt megoldani, mint a másik. Ugyanez volt a Rasch-modell érvényességének is az egyik feltétele. *Benjamin Wright* is megfogalmazza ezt a feltételt (pl. Wright és Stone 1979), de idézhetünk akár *Georg Raschtól* is:

> Egy személy, akinek a képessége [képességének fejlettségi szintje, képességfejlettsége (NI)] nagyobb, mint egy másik személyé, minden más, a kérdéses típusba tartozó feladatot nagyobb valószínűséggel old meg, és hasonlóan, egy item, amely nehezebb egy másiknál, minden személy számára kisebb valószínűséggel oldható meg, mint a könnyebb item (Rasch 1960, 117. o.).

S amikor az együttes mérés elmélete „segítene” a modern tesztelmélet alapján megformált eljárásoknak, akkor is az derül ki, hogy az X halmazon egy olyan rendezési relációnak kell értelmezettnek lenni, amely szerint két vizsgált személy esetén, bármilyen feladatot is oldjanak meg, az egyik személy általi helyes feladatmegoldásokhoz tartozó valószínűségnek nagyobbnak, vagy egyenlőnek kell lennie, mint a másik személy általi jó feladatmegoldás valószínűsége.

Láttuk, hogy itt az egydimenziósság kérdéséről van szó. Megjegyzendő, hogy a 20. század első harmadában a pszichometriai mérésekkel kapcsolatban ez a követelmény már ismert volt. *Louis Leon Thurstone* a képességek mérésének alapvető elvei, a vele szemben támasztható követelmények közt fontos szerepet szánt az egydimenziósságnak. Az ő megfogalmazásában: a mért objektumnak csak egyetlen tulajdonsága, egyetlen sajátossága határozhatja meg a mért értéket (Thurstone 1938, 257. o.). *Louis Guttmann* 1950-ben megjelent művében az egydimenziósságot kifejezetten úgy fogalmazza meg, ahogy itt is szerepelt. *Guttmann* szociológus volt, ezért elsősorban az állításokkal való egyetértés kérdése érdekelte:

> Ha egy személy egyetért egy szélsőséges állítással, akkor egyet kell, hogy értsen minden kevésbé szélsőséges állítással, ha az állításoknak egy skálán kell elhelyezkedniük. Egy közös tartalomhoz köthető itemek egy halmazát skálának nevezzük, ha egy másik személynél magasabb pontszámmal rendelkező személy minden item esetén magasabb pontszámot produkál, mint a másik (Guttmann 1950, 62. o.).

Benjamin Wright több írásában is használja azt a nagyon szemléletes példát, amely szerint, ha egy tesztben kétféle feladatot keverünk össze, mondjuk részben olvasási, szövegértési feladatok vannak, részben pedig matematikai műveletekkel kapcsolatosak, akkor a teszt nyilván nem lesz egydimenziós. Ez abban is megmutatkozik, hogy rendkívül sok vizsgált személy esetében mondanának ellent a feladatmegoldások konkrét eredményei az egydimenziósság követelményének, ami itt úgy jelentkezne, hogy nehéz szövegértési feladatokat megoldani tudó vizsgált személyek közül sokan nem tudnának megoldani viszonylag könnyű matematikai feladatokat, és lennének olyanok, akik elboldogulnának nehéz matematikai feladatokkal, de beletörne a bicskájuk könnyebbnek bizonyuló szövegértési feladatokba. De egyáltalán nem lehetünk biztosak benne, hogy ha tisztán csak szövegértési feladatokat használnánk, érvényesülne az egydimenziósság (magam persze biztos vagyok benne, hogy nem). Elég annyit elképzelnünk, hogy a vizsgált személyek egy nagyobb csoportja valamilyen területen „otthon van", vagyis meghatározó, nagy tömegű, jól szervezett tudással rendelkezik, egy másik csoportról meg ugyanez mondható el egy másik tudásterület tekintetében. Ha olyan tesztet készítünk, amely tartalmaz egyrészt az egyik tudásterülethez tartozó ismeretekre épülő szövegeket, és ezektől elkülönülve a másik tudásterülethez tartozókat is, akkor könnyen láthatjuk, hogy a szövegértés feladatrendszerének körében sem biztosítható az egydimenziósság.

Vagyis a modern tesztelmélet, valamint az együttes mérés elmélete igencsak tisztességesen megmondják, hogy mi a feltétele annak, hogy működjenek az általuk javasolt eljárások, legalábbis egy ezek közül az egydimenziósság. Ez azonban nem érvényesül az összetettebb emberi teljesítmények mérése esetén. És éppen ez a pedagógia problémája. Amiket a mi tudományunk mérni szeretne, az mind összetett, egydimenziósnak egyáltalán nem tekinthető „képesség". Nem véletlen, hogy a képességmérés hazánkban leginkább elismert szakemberei is többször leírják, hogy a mérésekben törekedni kell tartalomfüggetlen feladatok kialakítására. Nyilván azért fontos ez a törekvés, mert a tesztelméletek alkalmazásának alapvető feltétele, hogy „ne zavarja" más tényező az adott képesség működését. Én azonban úgy vélem, hogy ez a feltétel nem teljesíthető. A gyerekeknek a pedagógiai mérést szolgáló tesztekben adott feladatok – én csak ilyen példákat ismerek – sokféle tudás közbenjárását igénylik, a mért képességek nem egydimenziósak. A matematika persze szenvtelen, a néha nagy munkát igénylő tesztfejlesztéssel létrehozhatunk egydimenziós teszteket, vagy a fokozatos kizárásokkal kialakíthatunk homogén mintákat a vizsgált személyekből, azonban a sokféle feladat száműzésével már nem fogjuk tudni, valójában milyen képességet is mér a tesztünk, a személyek nagy számának kizárásával már nem tudjuk, hogy milyen speciális sajátossággal homogén módon rendelkező populációt vizsgálunk.

Idézzük itt kicsit hosszabban *Duncan Luce* és *Patrick Suppes* kemény, bár e két szerző esetében „látatlanban" megfontoltnak tekinthető szavait:

> Az általánosan „pszichológiai mérésnek" nevezett ténykedéseknek egy nagy része a különböző képesség- és teljesítménytesztek kidolgozását és használatát jelenti, és rendszerint a „pszichometria" megnevezés alá soroltatik. E területtel nem foglalkozunk, mert ezek az eljárások határozottan nem a reprezentációs mérések valamely ágát jelentik, és nem is előfutárai azoknak. Itt egyfajta összeszámlálás zajlik, egy tesztben a helyesen megválaszolt itemek számának a meghatározása, amelyet aztán az életkor vagy valamely más sajátosság szerint normálnak statisztikai módszerekkel, aminek eredményeként az összeszámlálás adatai normál eloszlású pontszámokká transzformálódnak. Ismét leszögezzük: nem léteztek és nem léteznek axiómák. A pszichometriai megközelítéseket tekintve mi csak Thurstone érzékelési mérésekkel

> kapcsolatos munkáival foglalkozunk. Nemrég Doignon és Falmagne dolgoztak ki a képességek mérésére vonatkozóan egy eljárást, a tudástér-elméletet, amely magán viseli a reprezentációs méréselmélet hatását (Luce és Suppes 2002, 12-13. o.).

Az túlzás, hogy a pszichometrikus-, illetve a képességfejlettség mérésekre vonatkozóan nincsenek axiomatikus rendszerek. Mint láttuk, a modern tesztelmélet önmagában is alkalmas intervallumskálák létrehozásának hátterét adni, továbbá a Rasch-modell, valamint az együttes mérés „házasságából" is megformálható egy következetes gondolati rendszer, azonban olyan kemény feltételeknek kell teljesülniük, hogy emiatt félő, a képességfejlettség méréseknek egy igen kis része, a legegyszerűbb emberi teljesítményekre vonatkozók halmaza lehet csak méréselméleti értelemben korrekt. De érdemes felfigyelnünk *Luce* és *Suppes* megjegyzésére a tudástér-elméletre vonatkozóan. Később magam is tárgyalom azokat a lehetőségeket, amelyek az itt felvetett problémák megoldása felé mutatnak, ezek közt a tudástér-elmélet valóban előkelő helyen kell, hogy szerepeljen (Falmagne és mts. 2003; Tóth 2005; Doignon és Falmagne 1999; Taagepera és mts. 1997; Falmagne és mtsai. 1990; Doignon és Falmagne 1985).

Eddigi eredményeink vegyesen pozitívak és negatívak. Láttuk, hogy a klasszikus tesztelmélet nem felel meg a reprezentációs méréselméletben szereplő követelményeknek. A modern tesztelmélet (legalábbis az itt tárgyalt részkérdésre korlátozva) már alkalmas intervallumskálák konstrukciójának hátteréül szolgálni, és az együttes mérés is alkalmas erre. Ezen elméletek esetében is van azonban negatívum: csakis az egydimenziós képességek esetén használhatjuk a kínált eljárásokat, és így a pedagógia számára a módszerek sajnos elég kevéssé elérhetők. Magam nem vagyok pszichológus, ezért nem szívesen élnék feltételezésekkel, hogy milyen egyszerű emberi teljesítmények esetén beszélhetünk egydimenziósságról. Mint láttuk, a modern tesztelmélet kínál statisztikai módszereket az egydimenziósságot sértő feladatok kiszelektálására, de egyrészt az eljárások bizonytalanok, másrészt a tesztcsonkítás veszélyes folyamat. Csak annyit szerettem volna jelezni, hogy az egész – talán néhol túl szigorúnak is tekinthető – elemzés egyáltalán nem zárja ki a mérés lehetőségét, csak rendkívül erősen korlátozza azt, és félő, hogy elsősorban a pedagógiai mérések lehetőségét zárja ki.

Azt is láttuk korábban, hogy ha kilépünk az egydimenziósság „birodalmából", és több dimenzió jelenléte esetén alkalmazható módszereket fejlesztünk ki, az sem jelent generális megoldást, hiszen a pedagógia számára fontos „végtelenül" sokdimenziós képességek esetén ezekkel sem megyünk semmire.

A megoldás keresése

A cím, „A megoldás keresése” annyiban pontatlan, hogy az igazán izgalmas próbálkozások nem az eredeti probléma megoldását célozzák. Az eredeti probléma az, hogy tudunk-e az emberek képességeinek fejlettségéhez úgy számokat hozzárendelni, hogy az megfeleljen a reprezentációs méréselmélet követelményeinek, és ennek eredményeként legalább intervallumskálák jöjjenek létre. Számos képesség, különösen a pedagógiát igazán érdeklők esetén – ez az én véleményem – nem csak az igaz, hogy ez eddig nem sikerült, hanem az is, hogy *ez nem is lehetséges következetes módon*. A többdimenziós vizsgálatok esetén persze nem egy számot rendelnénk hozzá a személyekhez képességfejlettségként, hanem az adott képesség dimenzióinak megfelelő számok összességét, másképpen egy-egy vektort, de az igazán fontos képességek esetén ez sem lehetséges – ahogy ezt igyekeztem alátámasztani. Vagyis a számok (legyen szó egyetlen számról, vagy számok által alkotott vektorokról) hozzárendelése úgy, hogy vizsgálódásunk igényeinek az megfeleljen (legalább intervallumskálák jöjjenek létre), nem lehetséges. Így az eredeti probléma valójában nem megoldható (a sokszor leírt kivételekkel). Amikor tehát itt megoldásokat keresünk, valójában egészen új megközelítéseket próbálunk konstruálni. Ezek az új megközelítések is azt a célt szolgálják, hogy valamiképpen jellemezni tudjuk az emberek adott képességeit, többek között azért, hogy azok fejlesztésével, fejleszthetőségével kapcsolatban jussunk használható információkhoz.

A most következő, befejező részben egyrészt megadom azokat a „szűk sávban létező” lehetőségeket, amelyek esetében még az eredeti probléma megoldásáról van szó (számok hozzárendelése), de sokkal fontosabb lesz a legutolsó részfejezet, amelyben a tudás struktúrájának vizsgálati lehetőségeire mutatok rá.

Ordinális skálák

Ha „nincsenek nagy igényeink”, akkor megelégedhetünk az ordinális skála létezésével is. Sőt, az ordinális skálán elhelyezkedő adatokat egyáltalán nem szabad lebecsülnünk, a modern matematikai statisztika ezekkel összefüggésben is számtalan magas szinten fejlett módszert kínál az analízisre (ld. pl. Cliff és Keats 2003; Csíkos 1999). A baj az, hogy a komplex képességek esetében – ezt láttuk a fenti elemzés során is, állandóan vissza-visszatértem e helyzet bemutatására – még az ordinális skála kialakítása is problematikus. Azt mondtuk, hogy az intervallumskála kialakítását megakadályozza már az is, hogy nem érvényesül, hogy bármely A személy ha egy B személynél egy feladatot nagyobb valószínűséggel tud megoldani, akkor ugyanez a helyzet az összes feladattal. De ha ez a feltétel nem érvényesül – márpedig a komplex képességek esetén biztosan nem teljesül – akkor az ordinális skála létezése is kizárt.

Gondolhatnánk azt, hogy elképzelhetők olyan egyszerűbb képességek, amelyek esetén intervallumskála nem hozható létre, de ordinális jellegű igen. Sajnos ez sem igaz. Ugyanis ha teljesülne a monotonitás, akkor már intervallumskála is létrehozható lenne. Ezt láttuk az együttes mérés bemutatása során. Ha értelmezhető rendezési reláció az X és Y halmazokon, akkor az együttes mérés elmélete már biztosítja, hogy intervallumskála is felépíthető. Vagyis az a helyzet, hogy a képességmérés speciálisabb feladata esetén (de az attitűdvizsgálatok is ide sorolhatók) *nincs olyan eset, hogy ordinális skála van, de nincs intervallumskála*. Az intervallumskála létezését kizáró többdimenziósság, a monotonicitás, az egydimenziósság hiánya kizárja egyben az ordinális skála létezését is.

Ez „szomorú következmény”, de nem mondhatok mást, ha tartom magam az eddigi elemzés logikájához.

Amikor mégis van intervallumskála

Két olyan szituációt, mérési helyzetet tudunk mondani, amikor nem érvényes a negatív eredményt hozó elemzés. Az egyik a rendkívül egyszerű képességek fejlettségének mérése, a másik az az eset (korábban már szerepelt), amikor a használt feladatrendszer nem tartozik jól megjelölhető képességhez, pusztán egy számunkra gyakorlati okokból fontos halmazt alkotnak a feladatok. Az első esetben tisztán érvényesíthető a modern tesztelmélet valamely modellje, az eredmények megbízhatók, a második esetben is valószínűleg ez a helyzet, de még szükség lenne kutatásokra, amelyek eredményei igazán meggyőzők lennének.

Miért ne fordulhatna elő, hogy bizonyos képességek mégiscsak egydimenziósak? Ilyenek lehetségesek, de nyilván hiába keressük őket a komplex, a pedagógiát sokkal inkább érdeklő képességek között.

Wright és Stone 1979-ben megjelent nagy jelentőségű könyvükben, amelyben a Rasch-modell alapjait mutatták be egy szélesebb közönség számára, végig egyetlen mérés eredményeit elemzik (Wright és Stone 1979). Az Egyesült Államok hadseregében is alkalmazott egyik tesztről van szó. Egy panelen sorban kis lámpák vannak elhelyezve, sorszámmal ellátva. A lámpákat a kísérletvezetők valamilyen sorrendben felkapcsolják (minden egyes feladat esetén valamilyen, a feladathoz tartozó számú lámpáról van szó), és arra kérik a résztvevőket, hogy reprodukálják a lámpák felvillanásának sorrendjét. Egy, két vagy három lámpavillanással nem érdemes foglalkozni, ezeket mindenki jól reprodukálja. Négy lámpa felvillanása során már lehetnek tévesztések, bár nagyon ritkán. Ahogy nő az egy item esetén felvillanó lámpák száma, úgy szaporodnak a hibás megoldások. 8-9 lámpa felvillanása esetén már lényegében nincs jó megoldás.

Ebben a tesztben egy feladatnak, egy itemnek tekinthetünk egy lámpa felvillanási sorozatot, amelyet a vizsgált személynek reprodukálnia kell. 0-t rendelünk egy személy esetében az item megoldásához, ha rosszul reprodukálja a felvillanások sorrendjét, és 1-et, ha a felidézés helyes. Wright és Stone az egész könyvön keresztül ezen az egy mérésen, 35 fővel, és 14 item felhasználásával lefolytatott mérés segítségével mutatják be a Rasch-modell keretében folytatható vizsgálatokat. Elképzelhető, hogy a lámpák felvillanási sorrendjének reprodukálása, mint képesség egydimenziós. Ezt sosem tudhatjuk biztosan, hiszen nem ismertek számunkra az „objektív" valószínűségek, a helyes reprodukálás valószínűségei, amelyek az egyes személyekhez az egyes felvillanássorozatok esetén rendelhetők. Ha ezeket ismernénk, ahogy azt már bemutattam a Rasch-modell tárgyalásánál, ellenőrizni tudnánk, hogy léteznek-e megfelelő képességfejlettségek, és feladatnehézségek, amelyek minden itt szóba került valószínűség kiszámítását lehetővé tennék a Rasch képletbe való behelyettesítéssel. Wright és Stone paraméteres fit-elemzéseket végeztek el, és még ennél az igen egyszerűnek vélhető képességnél is az adódott, hogy volt „rosszul viselkedő item", és volt „rosszul viselkedő vizsgált személy", vagyis mind az itemek, mind a vizsgált személyek halmazát szűkíteni kellett. Ebből következően bennem bizonyos kételyek merülnek fel azzal kapcsolatban, hogy a szerzőpáros által elemzett képesség vajon tényleg egydimenziós-e.

Hogy ez a „lámpafelvillanós" feladat, az ebben nyújtott teljesítmények megismerése hasznos-e a pedagógia számára, azt nem tudnám megmondani. Lehet, hogy igen. Az ilyen észlelési, emlékezeti feladatokban megnyilvánuló sikeresség lehet jelzője annak, hogy az agy működésének biokémiai feltételei miképpen alakultak az egyén esetében elsősorban genetikai hatásra, bár itt esetleg sok más tényezőnek is lehet szerepe. A biokémiai feltételek (pl. az idegsejtek ingerületátvivő anyagaival összefüggő paraméterek) alakulása globálisan is befolyásolhatja az agy működését számos, ha nem minden területen, így általában a legkülönbözőbb feladatokban elérhető sikerességre kapunk információkat. Az azonban nem igaz, hiszen éppen ezt hangsúlyoztam az egész elemzés során, hogy ha sikeres vagyok az egyik területen, akkor – csak azért, mert bizonyos, az agy működését meghatározó biokémiai paraméterek szerencsésen alakultak – sikeres leszek minden más területen is. Igaz, az intelligenciakutatások során nagyon sok kognitív képességünkben nyújtott teljesítmény korrelált egymással erősen, és valójában ez vezetett a *g* faktor (az általános intelligenciaösszetevő) fogalmának megalkotásához. Az intelligenciakutatás eredményei, és különösen azok interpretációi azonban erősen vitatottak, és a józan eszünk is azt mondatja velünk, hogy az eltérő feladattípusokban nyújtott eltérő mintázatú teljesítmények megjenése nem kivétel, hanem sokkal inkább szabály.

Megfogalmazható a kérdés, hogy találhatunk-e akár az egyes tantárgyak, műveltségi területek tanításával összefüggésben olyan egyszerű képességeket, amelyek egydimenziósak. Én ezt egyáltalán nem szeretném kizárni. Lehetnek olyan fejlődési szakaszok, amelyekben például bizonyos matematikai feladatokkal leírható képességek esetén ez komolyan felmerül. Általános iskola 4. osztályában a gyerekek már tudnak írásban összeadni háromjegyű számokat (tanulták). Elképzelhető, hogy ez a képesség (közel egymillió feladat) egydimenziós. De miért fontos az, hogy bizonyos életkorokról beszéljünk? Azért, mert minél idősebb gyerekekről, majd fiatalokról, felnőttekről van szó, ha még tanulnak olyasmit, amiben szükségük van ezeknek a feladatoknak a megoldására, vagy ha olyan munkát végeznek, amelyben szerepet kapnak az ilyen feladatok, akkor a képesség nagyon fejlett lehet, a képesség fejlettsége okán már nem nagyon tévesztenének, viszont ehhez viszonyítva megnő a véletlenül előadódó, vagy a legkülönbözőbb külső okok miatt bekövetkező tévesztések aránya. Ez utóbbi hatások viszont – ha volt is – „elrontják" az egydimenziósságot, hiszen hogy melyik feladat esetén hibázunk, az esetleges tényezőktől függ. Ez utóbbi jelenség egyébként a teljesítmény stabilitását is megkérdőjelezi. Jó kutatási feladat egy ilyen képesség vizsgálata az itt leírt szempontokból.

Jelöltek lehetnek egydimenziós képességekre: a számok normál alakjának felírása; valamilyen korlátozással számok szorzása; egyszerű, nem hiányos mondatokban az alany felismerése; angolul megfogalmazott kérdőmondatok esetén annak megadása, hogy nyelvtanilag helyes, vagy helytelen a fogalmazás (bár ez utóbbi két példa esetén már erősen bizonytalan vagyok, hogy azok valóban egydimenziós képességekkel kapcsolatosak-e). Én tehát nem tartom elképzelhetetlennek, hogy akár a pedagógiát is érdekelheti néhány igen egyszerű egydimenziós képesség, de ezek valószínűleg nem túl terjedelmes halmazának elemei összetettség és fontosság tekintetében igen messze állnak a gondolkodástól, a problémamegoldástól, a tanulástól, a kommunikációtól, a szövegesfeladat megoldástól, a döntéstől, stb.

A másik lehetőséget korábban már tárgyaltam: a „nagy mérések" esetén, ha a képességet (a „mesterséges képességet") azonosítjuk éppen azokkal a feladatokkal, amelyeket kiválasztunk a számunkra fontos feladatok közül, hogy így egy egydimenziós itemhalmaz jöjjön létre, vagyis a fit-elemzés során jó tulajdonságokat mutat a teszt, akkor szigorúan ezekre a feladatokra korlátozva érvényes az egydimenziósság, és így a számítások korrektek. Ez rendkívül megnyugtató eredmény, hiszen a PISA, a TIMSS, az IALS, a hazai kompetenciamérések rendkívül fontos eredményeket produkálnak.

Az egyik kérdés, ami felmerül, hogy vajon mi haszna van egy ilyen mérésnek, ha nem általában a matematikai eszköztudás, nem általában a szövegértés, nem általában a természettudományos problémamegoldás az a képesség, amit vizsgálnak, hanem egy abszolút „mesterségesen" előállított „képesség". A válasz erre a kérdésre az, hogy mondjuk a matematika teszt eredményét ne tekintsük (nem is tekinthetjük) a matematikai eszköztudás fejlettsége mértékének, hanem egy olyan aktuális mértéknek, amely arról tájékoztatja az oktatáspolitikusokat, a kutatókat, a pedagógusokat és más szakembereket, hogy mire számíthatnak a tanulók majdani matematikai problémamegoldásaival kapcsolatban, hogyan hasonlíthatók össze egy fontosnak tartott teljesítmény szempontjából a legkülönbözőbb csoportok (például az egyes országokhoz tartozók csoportjai), hogyan változnak az egymást követő mérésekben a csoportok értékei. Abban biztos vagyok, hogy mondjuk, a PISA szövegértés tesztjének eredményeit nem tekinthetjük általában valamifajta szövegértés képességfejlettség becsléseinek. A végtelen sok szövegértés feladatnak van sok olyan részhalmaza, amely esetén a nem paraméteres fit-elemzések nem jeleznek problémát. Mi ezek közül egyet használunk. Az viszont nagyon érdekes kérdés, hogy vajon az „éppen használt dimenziót" miképpen lehet jellemezni, mi is az a bizonyos dimenzió.

Ezzel összefüggésben szeretném felhívni a figyelmet egy fontos összefüggésre. Sokak számára ismertek az adatok: Magyarország teljesítménye – különböző mértékben – a 2012. évi PISA vizsgálatokban a korábbi eredményekkel összehasonlítva nem kis mértékben csökkent. Érdekes módon sem a szövegértésben, sem matematikában nem tapasztaltunk hasonlót az Országos kompetenciamérésben (a 8. és a 10. évfolyamon tanulók esetében kellett volna jelentkeznie a hatásnak 2012-t megelőzően, illetve azt követően). Ez ellentmondásnak tűnik, ha azt képzeljük, hogy a PISA és az OKM egyaránt a szövegértés képességének és a matematikai eszköztudásnak a fejlettségét mérik. Az ellentmondás feloldása az, hogy valójában a tesztek nem ezeket a konstrukciókat mérik, hanem azoknak csak egy-egy dimenzióját. Hogy melyiket, milyen jellegűt, azt nem tudjuk, de valószínűleg két olyat, amelyben nem azonos módon alakulnak a változások. Ezt akár nyugtalanító jelenségnek is tekinthetjük. Fontos lenne, hogy elemzések szülessenek az egyes tesztekben alkalmazott feladatok jellegéről, mennyire fontosak azok valóban a számunkra. És az is fontos lenne, hogy más értékelési módszerek is szerepet kapjanak az iskolarendszer és részei fejlettségének vizsgálatában, hogy komplexebb és árnyaltabb képet tudjunk kialakítani a pedagógiai munka eredményeiről.

A „nagy mérések" „megmentése" azért is fontos, mert amióta léteznek e nemzetközi vizsgálatok, azóta – éppen ezen adatok felhasználásával – olyan ismeretekhez jutottunk az oktatási rendszerek működésével kapcsolatban, amelyek megkérdőjeleződése óriási veszteség lenne. Gondoljunk arra, hogy az adatok segítségével az esélyegyenlőtlenséggel kapcsolatos kritikus összefüggésekre tudtunk rámutatni, sokkal többet tudunk arról, hogy mitől függ egy ország oktatási rendszerének színvonala, biztosabb ismeretekkel rendelkezünk a legkülönbözőbb társadalmi csoportokhoz tartozók iskolai sikerességének összehasonlítása tekintetében, stb.

Ugyanakkor a neveléstudományi kutatások számára, amiről írok „nem jó hír". A kutatások nem egy aktuálisan összeválogatott itemrendszer tekintetében nyújtott teljesítményekről kellene, hogy szóljanak, hanem átfogó, végtelen sok feladatot tartalmazó itemhalmazokkal reprezentálható, komplex képességekről. Márpedig azok esetében – ahogy az egész írásban ezt igyekeztem alátámasztani – a legmodernebb eljárások sem alkalmazhatók.

És ugyanez a helyzet az egyes tanulók értékelésével is, vagyis a módszereknek a pedagógiai gyakorlatban való alkalmazásával. Itt még az is gondot jelent, hogy ha minden rendben lenne egyébként, tehát a mért képesség egydimenziós lenne, és jól alkalmaznánk valamelyik IRT modellt, akkor is igaz lenne, hogy a tanuló teljesítményének önmagában is van egy szórása, amint láttuk a klasszikus tesztelméletnél a legrészletesebben: számolnunk kell egy hibataggal. Az országos kompetenciamérés esetén például 1500-1600 pontos évfolyamátlagok alakulnak ki, és egy tanuló teljesítményének szórása eléri, meghaladhatja a 60 pontot. Ha tehát valaki például 1560 pontot ért el az egyik tesztben (matematika vagy szövegértés), akkor 1/3 valószínűséggel fordulhat elő az az eset, hogy tényleges képességpontszáma kisebb, mint 1500 pont, vagy nagyobb, mint 1620 pont. Ha nagyobb csoportot vizsgálunk (legalább egy intézményt), akkor a valódi képességfejlettség pontszámoktól való eltérések „kiátlagolódnak”, pontosabban egy nagyobb csoport pontszámainak átlaga nagyon jól megegyezhet a valódi képességpontszámok átlagával. De egy tanuló esetében a mérés eredménye nagyon megbízhatatlan.

E negatív hatás kiküszöbölésére van állítólag megoldás, ez az *adaptív mérés*. Az adaptív mérés (ld. pl. Molnár 2013) számítógéppel történik. Minden feladat megoldása után megtörténik az addigi teljesítmények kiértékelése, és annak eredménye függvényében adja a szoftver a következő feladatot. Ugyanis kimutatható (én ezzel részletesen nem foglalkoztam a korábbiakban), hogy mindenkit az adott képessége tekintetében a fejlettségszintjéhez közel álló nehézségű feladatokkal lehet a leginkább jól felmérni. Egy rögzített tesztben szükségképpen nagyon különböző nehézségű feladatok vannak (minél szélesebb képességfejlettség-tartományban tudjon a teszt mérni), az adaptív tesztelés esetén azonban flexibilis módon alakítható, hogy milyen nehézségű feladatot kap a tanuló. Egy ponton az eddigi feladatmegoldásokból kiszámítható képességfejlettséggel azonos, vagy ahhoz közeli nehézségű feladatot ad a gép, ennek megoldásával kapcsolatos eredménnyel bővítve az eddigieket, még pontosabb becslést kaphatunk, de ez lehetővé teszi, hogy még inkább e szinthez közeli nehézségű feladatot adjunk, és így tovább. A tapasztalat azt mutatja, hogy ez a megoldás egyrészt csökkenti az egyéni szórásból adódó kellemetlen következményeket, másrészt kevesebb feladat elég a felméréshez. Hátrány, hogy egész csoportok felmérése esetén minden felmért tanuló számára biztosítanunk kell egy számítógépet. Az adaptív tesztelés természetesen azt is feltételezi, hogy rendelkezésünkre áll egy nagy, és alapos munkával kifejlesztett feladatbank. Be kell vallanom, hogy ismereteim túl szűkek ezen a területen, ám az a kevés, amivel rendelkezem, engem nem győzött meg arról, hogy az adaptív mérés számottevő mértékben csökkenti a mérési hibákat.

És végül: lépjünk túl a skálákon

A leginkább ígéretes megoldások radikálisan szakítanak az eddigi paradigmával. No persze akkor, ha az eddigi paradigma alatt azt értjük, hogy mindenáron számokkal akarjuk jellemezni a tanulók valamilyen képességének fejlettségét (attitűdjeik erősségét, személyiségjellemzőiket, stb.). Ezek a megoldások azt a kérdést teszik fel, hogy vajon *milyen a tudás szerkezete*.

Hogy mi is az emberi tudás, annak milyen a szerkezete, milyen funkciói vannak, és „ténylegesen” hogyan „működik”, az valószínűleg amióta öntudatára ébredt, érdekli az embert. A filozófia, a pszichológia, a mesterséges intelligencia tudománya, a kibernetika és még számos más tudomány érdeklődik erősen e kérdések iránt. Az a gyanúm, hogy ebben a kérdésben alakult ki a legtöbb tudományos magyarázó modell. Reménytelen vállalkozás áttekinteni minden e témával kapcsolatos eredményt. A pedagógiai szakember itt ismét bajban van (ahogy bajban van a méréselméletek, a képességekkel kapcsolatos elméletek és empirikus kutatások áttekintése során is), ha átfogó képet szeretne nyújtani. Ahhoz, hogy ezt biztonsággal, egy vállalható színvonalon megtegye, e szakterület szakemberének kellene lennie. Ennek hiányában viszont kénytelen egy elengedhetetlen mélységig beleásni magát a szakterületbe, állandó félelemmel, hogy nem szakemberként elvéti az irányt, nem megfelelően használja a terület nyelvét, kis területet fog át az „érzékeivel”, miközben figyelmének óriási területre kellene kiterjednie, és nem képes a legújabb eredményeket, fejleményeket figyelembe venni. Mi, pedagógiával, annak elméletével is foglalkozók állandóan szembe kerülünk ezzel a problémával. Amit mi vizsgálunk, vagyis a nevelés folyamata, annyira komplex, hogy számtalan felületen érintkezik más tudományok által vizsgált jelenségekkel, folyamatokkal, objektumokkal. Szerencsétlen neveléstudományi szakembernek sok ilyen diszciplína jó értőjének kellene lenni, és ez bizony nagyon nehéz feladat. De ha már ezt választottuk …

A tudás az emberben lévő „valami”, ami lehetővé teszi, hogy értelmezze, megértse környezetét, előrejelezze annak történéseit, és cselekedni legyen képes a maga számára adaptív módon ebben a környezetben. Fizikailag – legalábbis a legszélesebb körben elfogadott paradigmák szerint – az agysejtek közötti kapcsolatok adják a tudás birtoklásának materiális alapját. Nem tudjuk hogyan, de a tudást az idegsejtek közötti kapcsolatok, vagyis az elemekből, összeköttetéseikből, és ezek erősségének sajátos mintázataiból kialakuló hálózatok kódolják.

A tudás szónak nincs még az érintett szaktudományokban sem egységes jelentése. Számtalan tudományos munkát említhetnénk, amelyben a birtokolt, elraktározott ismereteket és csak az ismereteket jelenti. Más szakemberek a tudás szót ennél jóval kiterjedtebb értelemben használják: minden a világ reprezentálását, a történések előrejelzését és a feladataink megoldását szolgáló, belső kognitív feltételt tudásnak tekintenek. Vagyis az is egy legitim szemlélet, amelyben a képességek (bármik is legyenek azok) ugyanúgy tudások, mint az például, hogy tudjuk, mely országok Magyarország szomszédai. De arra az értelmezésre is találunk példát, hogy bármely az agyban működtetésre használt, működő struktúra tudásnak tekinthető, vagyis még akár az összetett attitűdök, az érzelmek sem hagyhatók ki ebből a fogalomból.

A tudásrendszerek struktúrájának vizsgálata során nehéz lett volna kiindulni egy ilyen körvonalazatlan fogalomból. Amikor *Jean-Paul Doignon* és *Jean-Claude Falmagne* a múlt század 80-as éveiben megalkották a tudástérelmélet alapjait (Doignon és Falmagne 1985), úgy döntöttek, hogy felhasználják az akkor már megalapozottnak tekinthető tesztelméleti paradigmák kiindulópontjait. Ez azt jelentette, hogy nem közvetlenül a tudást, a világ agyban megvalósuló reprezentációját kívánták vizsgálat tárgyává tenni, hanem megmaradtak a feladatok megoldásában megnyilvánuló teljesítmény elemzésénél, értékelésénél. A tudástérelmélet kidolgozása is arra épül, hogy legyenek bármilyenek is a belső struktúrák, azokat közvetlenül nem tudjuk megragadni, ám regisztrálni tudjuk, hogy valaki jól vagy rosszul old meg bizonyos feladatokat, és ebből igyekszünk továbbra is levonni következtetéseket (Doignon 1994, 4. o.). Ezek a következtetések azonban már nem azonosak az egy területen kialakult tudáshoz egy szám, esetleg számokkal kifejezhető vektor hozzárendelésével. *Doignon* és *Falmagne* a tudáshoz tartozó feladatok esetén a jó és a rossz megoldások struktúráját, megoldásmintázatát rendelték hozzá.

Vagyis mielőtt még valaki azt hinné, hogy a tudástérelmélet képes megragadni magának a tudásnak a szerkezetét, sietek leszögezni, hogy erről nincs szó, itt „csak” azt ragadjuk meg, hogy feladatok egy előre adott halmazából ki melyeket tudja megoldani. Ám ez a „csak” is nagyon sokat jelent. A huszadik század nyolcvanas éveitől kezdve egy terebélyes matematikai elmélet épült ki *Doignon* és *Falmagne* alapvetéseiből (ennek lényegét már ők leírták).

A tesztelméletek, és különösen komplexebb, a pedagógia szempontjából alapvető képességek fejlettségének mérésére való használatuk esetén elsősorban abba a problémába ütközünk, hogy csak bizonyos megoldásstruktúrákat vagyunk képesek az elmélet hatálya alá vonni. Azokat, amelyekben nem túl sűrűn fordul elő, hogy egy kisebb képességfejlettséggel rendelkező valaki meg tud oldani egy feladatot, miközben azt egy nagyobb képességfejlettséggel rendelkező valaki nem tudja megoldani. És az sem fordul elő túl sűrűn, hogy egy nehezebbnek számító feladatot meg tud oldani valaki, míg egy könnyebbnek bizonyulót nem. Azt mondtuk, hogy ha az ilyen esetek túl sűrűn fordulnak elő egy feladathalmazt tekintve, akkor a feladathalmaz által reprezentált tudásrendszerre nem illik a modern tesztelmélet semmilyen modellje, még a többdimenziós modellek sem jelentenek megoldást, mert a bennünket igazán érdeklő képességek esetén megragadhatatlanok a dimenziók. A tudástérelmélet nem indult ki ilyen korlátozó feltételezésekből. Itt akár nagyon sűrűn is előfordulhat, hogy A ember meg tudja oldani az a feladatot, ezzel szemben b-t nem, míg a B ember ezzel éppen fordítva van (nem tudja megoldani a-t, de megoldja b-t). Más a tudásuk szerkezete, és ezek a megoldásmintázatok erről árulkodnak.

De van egy másik nagyon fontos különbség a két szemléletmód (a tesztelméleti és a tudástérelméleti) között. A tesztelméletek esetén az egy konkrét alkalmazásban szereplő, szerepeltethető feladatok halmazát egy adott képességhez tartozóan igyekeztünk körülhatárolni. Ezt ugyan megtehetjük a tudástérelmélet esetében is, ám jellemzőbb az az alkalmazás, amelyben a feladatok valamilyen jól körülhatárolható tanulási területhez kapcsolódnak, mondjuk egy iskolai témához, egy szakma, vagy (inkább) egy részterülete eljárásainak megtanulásához, stb. Vagyis a tudástérelmélet által kezelt tudásrendszerek inkább felelnek meg a kompetenciafogalomnak, konkrétabbak, szemben a tesztelméletek által vizsgálni kívánt nagyon általános képességekkel. Jellegzetes példa az iskolai alapozó algebrai ismeretek köre, vagy egy másik a statisztikai hipotézisvizsgálatok tudásterülete. Viszonylag jól modellezhető, hogy melyek azok a típusfeladatok, amelyek e témákhoz (és számos hasonlóhoz) kapcsolódnak, szakértők bevonásával e tudásterületek feltérképezhetők.

A tudástérelmélet kiindulópontja tehát az, hogy egy adott (körülírt) tudásterületen van véges számú feladatunk, amelyeket a vizsgált személyek vagy meg tudnak oldani, vagy nem. A vizsgált területen megfogalmazott feladatok halmazát rendszerint Q-val jelölik, és sokszor hívják *problématérnek*, vagy *feladattérnek*. Ha megfigyelünk egy vizsgált személyt, akkor ő az e halmazba tartozó feladatok közül bizonyosakat tud megoldani, vagyis Q egy részhalmazával, mondjuk K-ba tartozó feladatok esetén lenne sikeres, ha szembe kerülne velük. K-t *tudásállapotnak* nevezzük. Már itt megragadható a tesztelméletek és a tudástérelmélet egy lényeges eltérése. A tesztelméletek (a klasszikus és a modern egyaránt) használatakor a tudás jellemzésére egy-egy vizsgált személy esetén egy-egy számot használunk pusztán. E szám „mögött” már nincs ott, hogy pontosan mely feladatokat tudta megoldani a személy, információt vesztettünk. Eltérő mintázatú megoldások vezethetnek ugyanahhoz a pontszámhoz[40]. A tudásállapot még őrzi az információt, akár kvalitatív elemzésnek vethetjük alá, vizsgálva, hogy éppen milyen jellegű feladatokat sikerült, és milyeneket nem sikerült megoldani.

Most jöjjön két először technikainak tűnő fogalom, a tudásstruktúra és a tudástér fogalma. Mindkettő tudásállapotok részhalmaza. Ha egy problématérben van – ez egy példa – 3 feladat, akkor összesen 8 tudásállapot van. Legyenek a feladatok a, b és c! Nyilván tudásállapot az üres halmaz (az a személy jellemezhető vele, aki a három közül egyetlen feladatot sem tud megoldani). Tudásállapotok továbbá a következő halmazok is (a kapcsos zárójelek között azt jelezzük, hogy az adott halmzba mely elemek tartoznak): $\{a\}$, $\{b\}$, $\{c\}$, $\{a, b\}$, $\{a, c\}$,$\{b, c\}$,$\{a, b, c\}$. Így az üres halmazzal (szokásosan a jele: $\emptyset$) együtt

[40] Természetesen, mint láttuk, ha jelentős különbségek alakulnak ki a személyek megoldásmintázatai között, akkor a teszt valószínűleg nem bizonyul egydimenziósnak, vagyis egyik tesztelmélet alkalmazása sem szolgáltathat valid eredményt. Ilyen esetben tehát nem pusztán az a probléma, hogy a tesztelméletek alkalmazása során információk vesznek el, hanem ennél súlyosabb gond adódik, valójában az adott képesség tesztelméleti eszközökkel nem vizsgálható.

összesen 8 tudásállapot van, köztük különleges szerepű az üres halmaz, és a felsorolásban az utolsó, amely mindhárom feladatot tartalmazza. Ez utóbbi nyilván maga a Q. A tudásállapotokat általában a K betűvel szoktuk jelölni. A felsorolt tudásállapotok ezzel a jelöléssel: $K_1 = \emptyset$, $K_2 = \{a\}$, $K_3 = \{b\}$, $K_4 = \{c\}$, $K_5 = \{a, b\}$, $K_6 = \{a, c\}$, $K_7 = \{b, c\}$, $K_8 = \{a, b, c\}$.

Tudásstruktúrának nevezzük egy adott problématérben tudásállapotoknak egy olyan halmzát, amelynek eleme az üres tudásállapot és a teljes tudásállapot is (vagyis $\emptyset$ és $Q = \{a, b, c\}$ is a példában). *Tudástér* viszont az olyan tudásstruktúra, amely az únióra, mint halmazműveletre nézve zárt. Mit jelent ez? Azt jelenti, hogy egy tudástérben lévő bármely két tudásállapot feladatainak egyesítésével egy olyan tudásállapot keletkezik, amely szintén benne van az adott tudástérben. Tudástér-e például a $\{K_1, K_2, K_3\}$? Nem, mert nem is tudásstruktúra, hiszen nincs ott a felsorolt tudásállapotok között a teljes tudásállapot, a Q, vagyis K_8. Tudástér-e $\{K_1, K_2, K_3, K_8\}$? Az biztos, hogy tudásstruktúra, hiszen a felsorolt tudásállapotok között ott van a $\emptyset$ is (K_1), és Q is (K_8). Az a kérdés, hogy tartalmazza-e bármennyi elemének únióját, egyesítését is. Ha végig próbálgatjuk, meghatározva két, három, illetve mind a négy elem únióját meghatározva, azt tapasztaljuk, hogy van egy olyan egyesítés, amelynek eredménye nincs felsorolva a tudásstruktúra elemei között. Ugyanis $K_2 \cup K_3 = K_5$, ami nincs felsorolva az elemek között, tehát a megadott tudásstruktúra nem tudástér. Tudásteret kapunk viszont akkor, ha K_5-öt felvesszük a szereplő tudásállapotok közé: $\{K_1, K_2, K_3, K_5, K_8\}$. Hogy a tudástér fogalmát miért konstruáltuk meg, az mindjárt kiderül, de ehhez be kell vezetnünk még egy fogalmat, az előfeltétel reláció fogalmát.

A tudás struktúrájának elemzésekor – ahogyan ez minden rendszer vizsgálatában alapvető – az elemek közötti kapcsolatok, matematikai nyelven szólva: a relációk játszanak alapvető szerepet. A tudástérelmélet a tudás szerkezetének elemzésében az *előfeltétel relációt* állítja a középpontba. Maga a reláció nagyon egyszerű, és mindenki számára jól érthető is, aki már járt iskolába. Ugyanis két tudáselem között értelmezhetjük az előfeltétel relációt, mégpedig akkor mondjuk, hogy az fennáll, ha az egyik tudáselem birtoklása szükséges a másik tudáselem megtanulásához. Nagyon könnyen látható, hogy ha egy a feladat megoldani tudása feltétele a b feladat jó megoldásának, és a b feladatban a sikeresség feltétele annak, hogy megbirkózzunk a c feladattal, akkor az a feladat megoldani tudása feltétele a c feladat helyes megoldásának. Ahogy a matematikusok mondják: az előfeltétel reláció tranzitív. Érdemes úgy értelmezni ezt a relációt, hogy minden feladat jó megoldásának előfeltétele, hogy azt a feladatot jól tudjuk megoldani (pestiesen erre mondanánk azt, hogy naná), vagyis matematikai nyelven ez egy reflexív reláció. A reflexivitás és tranzitivitás tulajdonságával rendelkező relációkat részben rendezési relációknak hívják. Beszédes a név: az elemek részben rendezettek, vagyis mint az előbbi példában, rendezés van az a, b, c feladatok között, mert az előbb álló az utána következők jó megoldásának előfeltétele, de ha van egy tudásállapotban egy d elem, amelynek a b elem előfeltétele, de a és c nem, akkor a felsorolt négy elem között már nem állapíthatunk meg rendezést. A konkrét példa is mutatja, hogy a tudástérelmélet mennyire más alapstruktúrákkal dolgozik, mint a tesztelméletek: a klasszikus és a modern tesztelmélet is csak olyan feladatrendszerekkel tud dolgozni, amelyekben a feladatok helyes megoldásának valószínűségei teljes mértékben, egyetlen sorrendbe állíthatók az összes feladatot tekintve. Ha az x feladat megoldani tudása előfeltétele az y feladat megoldásának, akkor minden tanulóra nyilván igaz, hogy az x feladatot legalább akkora valószínűséggel tudja jól megoldani, mint az y feladatot. A tudástérelméleti tárgyalás esetén azonban a példában szereplő d feladatot ebből a szempontból csak a b feladattal tudjuk összehasonlítani, az ugyanis biztos, hogy a b feladat helyes megoldásának valószínűsége nem kisebb, mint a d feladat helyes megoldásának valószínűsége. Még az is igaz, hogy ugyanez nem csak a *b*, hanem az *a* feladatra is igaz, hiszen van a példában szereplő feladatokkal kapcsolatban egy *a*, *b*, *d* sorozat is, ahol az előbb álló előfeltétele a hátrébb álló feladat jó megoldásának. De a *c* és a *d* feladatokat így már nem tudjuk összehasonlítani.

Ahhoz, hogy egy lineáris egyenletet, amelyben az ismeretlen csak első fokon szerepel, meg tudjak oldani, birtokolnom kell a számok összeadását, vagyis, meg kell tudnom oldani ilyen típusú feladatokat. Mondhatja erre valaki azt, hogy nem szükséges elemi aritmetikai műveletek megoldása akkor, ha egy számítógépes szoftver oldja meg a feladatot magát, nekem csak az adatokat kell megadnom. Ez igaz, és arra figyelmeztet, hogy adott feladatok megoldása több úton is lehetséges. Éppen ezért az egyismeretlenes, lineáris egyenlet megoldásának előfeltétele lehet a számok összeadása, de – ettől teljesen függetlenül – előfeltétele lehet a feladatot megoldani képes szoftver jó kezelése is. Attól függ, milyen feladatokat veszünk fel abba a halmazba, amelynek a struktúráját vizsgálni kívánjuk. Ha olyan széles módon definiáljuk ezt az alaphalmazt, hogy az egyenletmegoldás egészen más, nem csak hagyományos módszerei is beleférnek, akkor az egyenletmegoldáshoz legalább két úton lehet eljutni (az iskolában tanított „hagyományos" módszer, valamint a megfelelő szoftver használata). Az egyenletmegoldásnak, mint feladatnak előfeltétele vagy a számok összeadása, vagy a szoftverhasználat, ám ez utóbbiak, nincsenek egymással előfeltétel relációban.

Ez a leírás azonban még túl tág. *Doignon* és *Falmagne* olyan struktúrák vizsgálatába fogott bele annakidején, amelyek esetében egy adott tudáselem vagy elemi abban az értelemben, hogy az adott feladattérben nincs másik olyan elem, amelynek megléte a feltétele lenne, vagy pedig egyértelműen megmondható, hogy a kijelölt tudáselem meglétének egy másik határozottan a feltétele, nem lehet olyan szituáció, hogy a kijelölt feladatot a személy meg tudja oldani, miközben a feltételeként megjelölt feladatot nem. Vagyis Doignon és Falmagne nem kívántak vizsgálni olyan tudásstruktúrákat, amelyekben van olyan feladat, amelynek jó megoldásához többféle tudás meglétével is el lehet jutni. Ez nyilván korlát, és ezért nem véletlen, hogy a későbbiekben elkezdődött olyan struktúrák vizsgálata is, amelyekben az egyes tudáselemekhez alternatív utak is vezethetnek.

Tudásállapotnak neveztük el (ld. fent) a feladatok halmazának egy részhalmzát, ha valaki pontosan e kiválasztott feladatokat tudja helyesen megoldani, míg a többit nem. Világos, hogy ha egy tudásállapot tartalmaz egy feladatot, akkor annak a feltételét jelentő feladatokat is tartalmaznia kell (hiszen azokat is meg tudja oldani a kiválasztott személy). Doignon és Falmagne bebizonyították, hogy az ebben a most leírt értelemben lehetséges tudásállapotok halmaza tudástér (az eredeti, halmazelméleti bizonyítás még *George David Birkhofftól*, amerikai matematikustól származik). A tétel „fordítva is igaz": bármelyik tudástér esetén találunk olyan relációt, amely reflexív és tranzitív, vagyis lehet előfeltétel reláció. Itt valójában az emberi tudásként megvalósuló tudásállapotok leírásáról van szó. Ha valaki meg tud oldani egy feladatot, akkor meg tudja oldani annak összes előfeltételét is. Olyan tudásállapot nem valósulhat meg ténylegesen, amelyből egy benne lévő feladat előfeltétele, mint feladat hiányzik. Tegyük hozzá mindehhez, hogy az előfeltétel reláció alapján kialakuló tudásterek esetén igaz, hogy egy ilyen tudástér a halmazelméleti metszet műveletre nézve is zárt. Ez azt jelenti, hogy ha van egy tudásterünk, és vesszük annak bármely két elemét, akkor az azokban lévő feladathalmazok közös feladatai (amik tehát mindkét tudásállapotban megtalálhatók), szintén egy ugyanabban a tudástérben szereplő tudásállapotot eredményeznek.

Amiket a tudástérelmélet kapcsán eddig leírtam, lényegében csak illusztrációnak szántam. Szerettem volna érzékeltetni, hogy ez a modern teória mennyire más megfontolásokat tartalmaz a tudás vizsgálatával kapcsolatban. Nem folytatom az elmélet bemutatását, mert ahhoz egy újabb könyv lenne szükséges, és meg is bontaná talán azt az egységet, amit az írásban eddig a mérés- és tesztelméletek bemutatásával kialakítottam. A tudástérelmélet „már nagyon más". Újabb fejezet az emberi tudás tanulmányozásában, ezért sem szeretném egyfajta függelékként megjeleníteni ebben a könyvben. Sőt, talán már másoknak, nálam sokkal fiatalabbaknak kellene ezt egyébként is megtenni.

Irodalom

Aamodt, A. és Plaza, E. 1994. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications, 7*(1), 39-59. Az Interneten 2016. január 13-án: http://www.idi.ntnu.no/emner/tdt4173/papers/Aamodt_1994_Case.pdf

van Abswoude, A.A.H., van der Ark, L.A. és Sijtsma, K. 2004. A Comparative Study of Test Data Dimensionality Assessment Procedures Under Nonparametric IRT Models. *Applied Psychological Measurement, 28*(1), 3–24. Az Interneten 2016. szeptember 9-én: http://hbanaszak.mjr.uw.edu.pl/TempTxt/AbswoudeEtAl_2004_ComparativeStudyOfDimensionalityProceduresUnderNonparametricIRTModels.pdf

Ackerman, T.A. 1996. Graphical representation of multidimensional item response theory analyses. *Applied Psychological Measurement, 20,* 311–329. Az Interneten 2016. szeptember 9-én: http://conservancy.umn.edu/bitstream/handle/11299/119465/1/v20n4p311.pdf

Alagumalai, S., Curtis, D.D. és Hungi, N. (Szerk.) 2005. *Applied Rasch Measurement: A Book of Exemplars. Papers in Honor of John P. Keeves.* Springer, Norwell. (Google-könyv)

Alagumalai, S. és Curtis, D.D. 2005. Classical Test Theory. In Alagumalai, S., Curtis, D.D. és Hungi, N. (Szerk.) *Applied Rasch Measurement: A Book of Exemplars. Papers in Honor of John P. Keeves.* Springer, Norwell. 1-14. (Google könyv)

Anderson, J.R. 1976. *Language, Memory and Thought.* Lawrence Erlbaum Associates, Hillsdale. (Google könyv)

Anderson, J.R. 1993. Problem Solving and Learning. *American Psychologist, 48*(1), 35-44. Az Interneten 2016. szeptember 9-én: https://www.ida.liu.se/~729G15/res/kompendium/ACT_R_learning.pdf

Anderson, J. R., Fincham, J.M., Qin, Y. és Stocco, A. 2008. A central circuit of the mind. *Trends in Cognitive Science, 12*(4), 136-143. Az Interneten 2016. szeptember 9-én: http://act-r.psy.cmu.edu/wordpress/wp-content/uploads/2012/12/800TiCS%202008.pdf

Anderson, M.J. 1992. *Intelligence and development: A cognitive theory*. Blackwell, Oxford. Magyarul: 1998. *Intelligencia és fejlődés.* Kulturtrade Kiadó, Budapest.

Antal, J. 2003. *Fit Indices for the Rasch Model*. Dissertation. Ohio State University, Columbus. Az Interneten 2016. szeptember 9-én: https://etd.ohiolink.edu/!etd.send_file?accession=osu1054222470&disposition=attachment

Baker, F.B. 2001. *The Basics of Item Response Theory.* ERIC Clearinghouse on Assessment and Evaluation, Madison. Az Interneten 2016. szeptember 9-én: http://files.eric.ed.gov/fulltext/ED458219.pdf

Barrett, H.C. és Kurzban, R. 2006. Modularity in Cognition: Framing the Debate. *Psychological Review, 113*(3), 628-647. Az Interneten 2016. szeptember 9-én: http://www.sas.upenn.edu/psych/PLEEP/pdfs/2006%20Barrett%20and%20Kurzban%20Psy%20Rev.pdf

Barrett, P. 2000. Intelligence, Psychometrics, Iq, G, and Mental Abilities: Quantitative Methodology Dressed as Science. *Psycoloquy*: 11(4). Az Interneten 2016. szeptember 9-én: http://www.cogsci.ecs.soton.ac.uk/cgi/psyc/newpsy?11.046

Barrett, P. 2003. Beyond psychometrics: Measurement, non-quantitative structure, and applied numerics. *Journal of Managerial Psychology, 18*(5), 421-439. Az Interneten 2016. szeptember 9-én: http://www.psyconsult.de/bosnjak/cpr2005/cd/3_Materials_from_Participants/Barrett/Barrett_Beyond_Psychometrics_2003.pdf

Barrett, P. 2008. The Consequence of Sustaining a Pathology: Scientific Stagnation – a Commentary on the Target Article "Is Psychometrics a Pathological Science?" by Joel Michell. *Measurement, 6,* 78-83. Az Interneten 2016. szeptember 9-én: http://www.pbarrett.net/publications/Measurement_Scientific_Stagnation_commentary_Barrett_2008.pdf

Batchelder, W.H. és Narens, L. 1977. A Critical Examination of the Analysis of Dichotomous Data. *Philosophy of Science, 44*, 113-135. Interneten 2012.12.28-án: Az Interneten 2016. szeptember 9-én: http://aris.ss.uci.edu/~lnarens/1977/Batchelder%26Narens_PhilosophyOfScience_1977.pdf

Baxter, P. és Browne, W. 2010. Memory as the substrate of cognition: a developmental cognitive robotics perspective. In: Johansson, B., Sahin, E. és Balkenius, C. (Szerk.) *Proceedings of the Tenth Conference of Epigenetic Robotics: Modelling Cognitive Development in Robotic Systems*. Lund University Cognitive Studies, Lund. 19-26. Az Interneten 2016. szeptember 9-én: http://www.lucs.lu.se/LUCS/149/baxter.pdf

Bogard, T., Liu, M. és Chiang, Y-h.V. 2013. Thresholds of knowledge development in complex problem solving: a multiple-case study of advanced learners' cognitive processes. *Educational Technology Research and Development.* Az Interneten 2016. szeptember 9-én: http://alienrescue.edb.utexas.edu/researchpapers/Thresholds.pdf

Boring, E.G. 1923. Intelligence as the Tests Test It. *New Republic, 36*, 35-37.

Boring, E.G. 1945. The use of operational definitions in science. *Psychological Review*, 52, 243-245.

Borsboom, D. 2005. *Measuring the Mind. Conceptual Issues in Contemporary Psychometrics*. Cambridge University Press, Cambridge. (Google könyv)

Borsboom, D. 2006a. The Attack of the Psychometricians. *Psychometrika, 71*(3), 425-440. Az Interneten 2016. szeptember 9-én: http://www.springerlink.com/content/t650617564233751/fulltext.pdf

Borsboom, D. 2006b. Can We Bring About a Velvet Revolution in Psychological Measurement? A Rejoinder to Commentaries. *Psychometrika, 71*(3), 463-467. Az Interneten 2016. szeptember 9-én: http://www.springerlink.com/content/w53773hq40133458/fulltext.pdf

Borsboom, D., Mellenbergh, G.J. és van Heerden, J. 2003. The Theoretical Status of Latent Variables. *Psychological Review, 110*(2), 203-219. Az Interneten 2016. szeptember 9-én: http://rhowell.ba.ttu.edu/BorsboomLatentvars2003.pdf

Bouyssou, D. és Pirlot, M 2005. Conjoint Measurement Tools for MCDM. In: Figueira, J., Greco, S. és Ehrgott, M., (Szerk.) *Multiple Criteria Decision Analysis: State of the Art Surveys.* Kluwer Academic Publishers, Boston, Dordrecht, London. 73-132. Az Interneten 2016. szeptember 9-én: http://hal.archives-ouvertes.fr/docs/00/02/89/21/PDF/cahierLamsade210.pdf

Bridgman, P. W. 1927. *The logic of modern physics*. Macmillan, New York.

British Association for the Advancement of Science 1933. Interim Report of the Committee Appointed to Consider and Report upon the Possibility of Quantitative Estimates of Sensory Events. *Report of the Annual Meeting*, 277–334.

Buller, D. és Hardcastle, V.G. 2000. Evolutionary psychology, meet developmental neurobiology: Against promiscuous modularity. *Brain and Mind*, *1*, 302–325. Az Interneten 2016. szeptember 9-én: http://commons.lib.niu.edu/bitstream/handle/10843/13181/evolutionarypsych.pdf?sequence=1

Campbell, N. R. 1920. *Physics, the elements*. Cambridge University Press. Cambridge. (Google könyv)

Cervone, D. és Caldwell, T.L. 2008. From Measurement Theory to Psychological Theory, in Reverse. *Measurement, 6,* 84-88.

Chadderdon, G.L. és Sporns, O. 2006. A large-scale neurocomputational model of task-oriented behavior selection and working memory in prefrontal cortex. *Journal of Cognitive Neuroscience, 18*, 242–257. Az Interneten 2016. szeptember 9-én: http://www.indiana.edu/~cortex/JOCN.pdf

Chi, M.T.H., Feltovich, P.J. és Glaser, R. 1981. Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science, 5*(2) 121-152. Az Interneten 2016. szeptember 9-én: http://onlinelibrary.wiley.com/doi/10.1207/s15516709cog0502_2/pdf

Cliff, N. 1992. Abstract measurement theory and the revolution that never happened. *Psychological Science*, *3*, 186–190.

Cliff, N. és Keats, J.A. 2003. *Ordinal Measurement in the Behavioral Sciences*. Lawrence Erlbaum Associates, Mahwah. (Google könyv)

Corbetta, M., Patel, G. és Shulman, G.L. 2008. The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron, 58*(3), 306-324. Az Interneten 2016. szeptember 9-én: http://www.cell.com/neuron/pdf/S0896-6273(08)00369-3.pdf

Cortina, J.M. 1993. What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104. Az Interneten 2016. szeptember 9-én: http://www.psycholosphere.com/what%20is%20coefficient%20alpha%20by%20Cortina.pdf

Cronbach, L.J. 1947. Test „Reliability": Its Meaning and Determination. *Psychometrika, 12*(1), 1-16.

Csapó Benő 2000. Tudásszintmérő tesztek. In: Falus Iván. (Szerk.) Bevezetés a pedagógiai kutatás módszereibe. Műszaki Könyvkiadó, Budapest. 277-315.

Csapó Benő 2001. Az induktív gondolkodás fejlődésének elemzése országos reprezentatív felmérés alapján. *Magyar Pedagógia,* 3. 373-391. Az Interneten 2016. szeptember 9-én:

http://www.edu.u-szeged.hu/~csapo/publ/indmp.pdf
Csapó Benő 2003. *A képességek fejlődése és iskolai fejlesztése.* Akadémiai Kiadó, Budapest.
Csíkos Csaba 1999. Nem-paraméteres statisztikai módszerek alkalmazási lehetőségei a pedagógiai kutatásban. *Iskolakultúra, 2.*: 113-119. Az Interneten 2016. szeptember 9-én: http://epa.oszk.hu/00000/00011/00024/pdf/iskolakultura_EPA00011_1999_02_113-119.pdf
Díez, J.A. 1997. A Hundred Years of Numbers. An Historical Introduction to Measurement Theory 1887-1990, Part I: The Formation Period. Two Lines of Research: Axiomatics and Real Morphisms, Csales and Invariance. *Studies in History amd Philosophy of Science, 28*(1), 167-185. Part II: Suppes and the Mature Theory. Representation and Uniqueness. *Studies in History amd Philosophy of Science, 28*(2), 237-265. Az Interneten 2016. szeptember 9-én: http://www.lps.uci.edu/~johnsonk/CLASSES/FoundationsOfMeasurement/Diez.AnHistoricalIntroductionToMeasurementTheoryPartOne.pdf
Domotor, Z. és Batitsky, V. 2008. The Analytic Versus Representational Theory of Measurement: A philosophy of Science Perspective. *Measurement Science Review, 8, Section 1,* (6)129-146. Az Interneten 2016. szeptember 9-én: http://www.measurement.sk/2008/S1/Domotor.pdf
Dingle, H. 1950. A theory of measurement. *British Journal of the Philosophy of Science*, 1, 5-26.
Doignon, J-P. 1994. Probabilistic assessment of knowledge. In: Albert, D. (Szerk.) *Knoeledge Structures.* Springer Verlag, New York. 1-56. Az Interneten 2016. szeptember 9-én: http://css-kti.tugraz.at/research/cssarchive/publicdocs/publications/albert1994.pdf#page=14
Doignon, J-P., és Falmagne, J-C. 1985. Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, 23, 175–196. Az Interneten 2016. szeptember 9-én: https://www.researchgate.net/profile/Jean-Paul_Doignon/publication/223630803_Spaces_for_the_assessment_of_knowledge/links/56bb73da08ae2481ab6abf72.pdf
Doignon, J-P. és Falmagne, J-C. 1999. *Knowledge Spaces.* Springer Verlag; Berlin, Heidelberg
Dunn, T.J., Baguley, T. és Brunsden, V. 2014. From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399–412. Az Interneten 2016. szeptember 9-én: http://s3.amazonaws.com/academia.edu.documents/31685358/Dunn_et_al._From_alpha_to_omega.pdf?AWSAccessKeyId=AKIAJ56TQJRTWSMTNPEA&Expires=1473410944&Signature=aeEsDvB5Idm3Y3zBoW186pQ8DPA%3D&response-content-disposition=inline%3B%20filename%3DFrom_alpha_to_omega_A_practical_solution.pdf
Edin, F., Klingberg, T., Johansson, P., McNab, F., Tegnér, J. és Compte, A. 2009. Mechanism for top-down control of working memory capacity. *Proceedings of the National Academy of Sciences of the United States of America, 106*(16), 6802-6807. Az Interneten 2016. szeptember 9-én: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2672558/pdf/zpq6802.pdf
Embretson, S.E. és Reise, S.P. 2000. *Item Response Theory for Psychologists.* Lawrence Erlbaum Associated Inc., Publishers, Mahwah.
Emons, W.H.M, Meijer, R.R. és Sijtsma, K. 2002. Comparing Simulated and Theoretical Sampling Distributions of the U3 Person-Fit Statistic. *Applied Psychological Measurement, 26*(1), 88-108.
Falmagne, J-C., Doignon, J-P., Koppen, M., Villano, M. és Johannesen, L. 1990. Introduction to knowledge spaces: How to build, test, and search them. *Psychological Review, 2*, 201–224.
Falmagne, J-C., Doignon, J-P., Cosyn, E. és Thiery, N. 2003. *The Assessment of Knowledge in Theory and in Practice*. Paper 26. Institute for Mathematical Behavioral Sciences, University of California, Irvine. Az Interneten 2016. szeptember 9-én: http://www.aleks.com/about_aleks/Science_Behind_ALEKS.pdf
Fechner, G. T. 1860. *Elemente der psychophysik.* Breitkopf & Hartel, Leipzig
Fodor, J. 1983. *The Modularity of Mind.* MIT Press/A Bradford Book, Cambridge
Fuster, J.M. 1997. Network Memory. *Trend sin Neuroscience, 20*(10), 451-459. Az Interneten 2016. szeptember 9-én: http://www.ufrgs.br/ppgneuro/artigos/Fuster_NetworkMMR_TINS1997.pdf
Fuster, J.M. 2002. Frontal lobe and cognitive development. *Journal of Neurocytology*, *31*, 373–385. Az Interneten 2016. szeptember 9-én: http://www.hss.caltech.edu/~steve/fuster.pdf
Glaser, R.J. 1983. *Education and Thinking: The Role of Knowledge*. University of Pitsburg, Pitsburg. Az Interneten 2016. szeptember 9-én: http://www.dtic.mil/dtic/tr/fulltext/u2/a130532.pdf
Glasersfeld, E. v. 1995. *Radical Constructivism. A Way of Knowing and Learning*. The Palmer Press; London, Washington D. C.
Goldstein, H. 1980. Dimensionality, bias, independence and measurement scale problems in latent trait test score models. *British Journal of Mathematical and Statistical Psychology, 33*, 234-246

Goldstein, H. 2004. International comparisons of student attainment: some issues arising from the PISA study. *Assessment in Education, 11*(3), 319-330.
Gould, S.J. 1981. *The Mismeasure of Man.* W.W. Norton & Company, New York. Magyarul: Gould, S.J. 1999. *Az elméricskélt ember.* Typotex, Budapest.
Green, S.B., Lissitz, R.W.és Mulaik, S.A. 1977. Limitations of Coefficient Alpha as an Index of Test Unidimensionality. *Educational and Psychological Measurement, 37*, 827-838.
Green, S.B. és Yang, Y. 2009. Commentary on coefficient alpha. *Psychometrika, 74*(1), 121-135.
Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., és Csapó, B. 2013. Complex Problem Solving in Educational Settings – something beyond g: Concept, Assessment, Measurement Invariance, and Construct Validity. *Journal of Educational Psychology*, *105*(2), 364-379. Az Interneten 2016. szeptember 9-én: http://archiv.ub.uni-heidelberg.de/volltextserver/14844/1/Dissertation%20Sascha%20W%C3%BCstenberg%20-%20Nature%20and%20Validity%20of%20Complex%20Problem%20Solving.pdf#page=59
Guttman, L. 1950. The problem of attitude and opinion measurement. In: Stouffer, S.A., Guttman, L., Suchman, E.A., Lazarsfeld, P.F., Star, S.A. és Clausen, J.A. (Szerk.) *Measurement and prediction.* Wiley, New York, 46-59.
Hambleton, R.K., Swaminathan, H. és Rogers, H.J. 1991. *Fundamental of Item Response Theory.* Sage Publications; Newbury Park, London, Delhi.
Helmholtz, H. V. 1887. Zahlen und Messen, erkenntnistheoretish betrachtet. In: Helmholtz, *Schriften zur Erkenntnistheorie*. 70-108. (Angol fordítás: Numbering and Measuring from an Epistemological Viewpoint. In: Helmholtz, *Epistemological Writings,* 72-l 14.)
Hirschfeld, L.A. és Gelman, S.A. 1994. Toward a topography of mind: An introduction to domain specificity. In Hirschfeld, L.A. és Gelman, S.A. (Szerk.) *Mapping the Mind: Domain Specificity in Cognition and Culture.* Cambridge University Press, new York. Az Interneten 2016. szeptember 9-én: http://is.muni.cz/el/1423/jaro2009/SAN217/um/7455540/Zaklad_k_27_03.pdf
Holland, P.W. és Rosenbaum, P.R. 1986. Conditional association and unidimensionality in monotone latent trait models. *Annals of Statistics, 14,* 1523–1543. Az Interneten 2016. szeptember 9-én: http://projecteuclid.org/download/pdf_1/euclid.aos/1176350174
Horváth György 1991. *Az értelem mérése.* Tankönyvkiadó, Budapest.
Horváth György 1993. *Bevezetés a tesztelméletbe.* Keraban Könyvkiadó, Budapest.
Hölder, O. 1901. Die Axiome der Quantität und die Lehre vom Mass. *Berichte über die Verhandlungen der Königlich Sachsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematische-Physicke Klasse,* ***53****,* 1–64.
Jöreskog, K.G. 1971. Statistical Analysis of Sets of Congeneric Tests. *Psychometrica, 36*(2), 109-133. Az Interneten 2016. szeptember 9-én: http://www.helsinki.fi/~ranne/thesis/Joreskog-1971a/Joreskog-1971a.pdf
Józsa Krisztián 2007. A számlálási készség kritériumorientált fejlesztése. In: Nagy József (szerk.): *Kompetencia alapú kritériumorientált pedagógia*. Mozaik Kiadó, Szeged. 291–298.
Karabatsos, G. 2000. A Critique of Rasch Residual Fit Statistics. *Journal of Applied Measurement, 1*(2), 152-176.
Karabatsos, G, 2003. Comparing the Aberrant Response Detection Performance of Thirty-Six Person-Fit Statistics. *Applied Measurement in Education, 16*(4), 277-298.
Karmiloff-Smith, A. 1992. *Beyond Modularity. A Developmental Perpective on Cognitive Science.* MIT Press, Cambridge.
Krantz, D. Luce, R.D., Suppes, P. és Tversky, A. 1971. *Foundations of Measurement.* Vol. 1. Academic Press, New York.
Kreiner, S. és Christensen, K.B. 2014. Analyses of Model Fit and Robustness. A New Look at the PISA Scaling Model Underlying Ranking of Countries According to Reading Literacy. *Psychometrika, 79*(2), 210-231.
Kuhn, T. 1962. *The Structure of Scientific Revolutions.* University of Chicago Press; Chicago, London. Magyarul: 1984 *A tudományos forradalmak szerkezete.* Gondolat Kiadó, Budapest.
Langley, P. és Rogers, S. 2005. An extended theory of human problem solving. *Proceedings of the Twenty-seventh Annual Meeting of the Cognitive Science Society*. Cognitive Science Society, Stresa. Az Interneten 2016. szeptember 9-én: http://ingenieriasimple.com/problemas/ExtendedTheory.pdf
Lannert Judit 2015. A PISA-adatok használata és értelmezése a módszertani kritikák tükrében. *Educatio, 24*(2), 18-29.
Loevinger, J. 1957. Objective Tests as Instruments of Psychological Theory. *Psychological Reports,* 3. 635-694.

Lord, F.M. 1980. *Applications of Item Response Theory to Practical Testing Problems*. Lawrence Erlbaum Associates, Hillsdale.
Lord, F. M. és Novick, M. R. 1968. *Statistical theories of mental test scores.* Addison-Wesley, Reading.
Luce, R.D. 1987. Measurement Scales on the Continuum. *Science, 236,* 1527-1532. Az Interneten 2016. szeptember 9-én: https://www.researchgate.net/profile/Louis_Narens/publication/6004350_Measurement_Scales_on_the_Continuum/links/0fcfd510fe34677a56000000.pdf
Luce, R.D. 1992. A Path Taken: Aspects of Modern Measurement theory. In Healy, A.F., Kosslyn, S. és Shiffrin, R. (Szerk.) *From Learning Theory to Connectionist Theory: Essays in Honor of William K. Estes, Vol. 1*. Erlbaum, Hillsdale. 45-64. Az Interneten 2016. szeptember 9-én: http://www.imbs.uci.edu/files/personnel/luce/1992/Luce_Book%20Chapter_1992.pdf
Luce, R.D. 1996. The Ongoing Dialog between Empirical Science and Measurement Theory. *Journal of Mathematical Psychology, 40,* 78-98. Az Interneten 2016. szeptember 9-én: http://www.socsci.uci.edu/~rdluce/bio/1996/Luce_JMP_1996a.pdf
Luce, R.D. és Narens, L. 1981. Axiomatic Measurement Theory. *SIAM-AMS Proceedings, Volume 12,* 213-235. Az Interneten 2016. szeptember 9-én: http://www.socsci.uci.edu/~rdluce/bio/pre1990/1981/LuceNarens_SIAM-AMS-Proceedings_1981.pdf
Luce, R.D. és Narens, L. 1983. Symmetry, Scale Types, and Generalizations of Classical Physical Measurement. *Journal of Mathematical Psychology, 27,* 44-85. Az Interneten 2016. szeptember 9-én: http://imbs-dev.ss.uci.edu/files/personnel/luce/pre1990/1983/LuceNarens_JMP_1983.pdf
Luce, R.D. és Narens, L. 1984. Classification of real measurement representations by scale type. *Measurement, 2*(1), 39-44. Az Interneten 2016. szeptember 9-én: https://www.researchgate.net/profile/Louis_Narens/publication/222466028_Classification_of_real_measurement_representations_by_scale_type/links/02e7e51cc502fe6454000000.pdf
Luce, R.D. és Narens, L. 1985. Classification of Concatenation Measurement Structures according to Scale Type. *Journal of Mathematical Psychology, 29*, 1-72. Az Interneten 2016. szeptember 9-én: http://www.imbs.uci.edu/~lnarens/1985/LuceNarens_JMP_1985.pdf
Luce, R.D. és Narens, L. 1994. Fifteen problems concerning the representational theory of measurement. In Humphreys, P. (Szerk) *Patrick Suppes: Scientific Philosopher, Vol. 2: Philosophy of Physics, Theory Structure, Measurement Theory, Philosophy of Language, and Logic*. Kluwer Academic Publishers, Dordrecht. 219–245. Az Interneten 2016. szeptember 9-én: http://www.imbs.uci.edu/~lnarens/1994/Luce&Narens_Book%20Chapter_1994.pdf
Luce, R.D. és Suppes, P. 2002. Representational Measurement Theory. In Pashler, H. és Wixted, J. (Szerk.) *Stevens' Handbook of Experimental Psychology, 3rd Edition, Vol 4*. Wiley, New York. 1-41. Az Interneten 2016. szeptember 9-én: http://suppes-corpus.stanford.edu/articles/mpm/382.pdf
Luce, R.D. és Tukey, J.W. 1964. Simultaneous conjoint measurement: a new type of fundamental measurement. *Journal of Mathematical Psychology, 1*, 1-27. Az Interneten 2016. szeptember 9-én: http://www.imbs.uci.edu/files/personnel/luce/pre1990/1964/LuceTukey_JMP_1964.pdf
Mari, L. 1997. The role of determination and assignment in measurement. *Measurement, 21*(3), 79-90.
Marr, D.C. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman, New York.
Mayer, R. 1998. Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science, 26.,* 49-63. Az Interneten 2016. szeptember 9-én: http://www.cimm.ucr.ac.cr/ojs/index.php/eudoxus/article/viewFile/169/303
McDonald, R.P. 2000. A Basis for Multidimensional Item Response Theory. *Applied Psychological Measurement, 24*(2), 99–114.
Meehl, P.E. 1991. Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. In: Anderson, C.A. és Gunderson, K. (Szerk.) *Selected philosophical and methodological papers – Paul E. Meehl.* University of Minnesota Press, Minneapolis. Eredetileg: Mehl, P.E. 1978. Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consukting and Clinical Psychology, 46*, 806-834. Az Interneten 2016. szeptember 9-én: http://www.psy.plymouth.ac.uk/research/mtucker/Psy245/Theoretical%20Risks%20and%20tabular%20asterisks.pdf
Meijer, R.R. és Sijtsma, K. 1995. Detection of Aberrant Item Score Patterns: A Review of Recent Developments. *Applied Measurement in Education, 8*(3), 261-272. Az Interneten 2016. szeptember 9-én: https://pure.uvt.nl/portal/files/1029526/detection.pdf

Meijer, R.R. és Sijtsma, K. 2001. Methodology review: Evaluating person fit. *Applied Psychological Measurement, 25,* 107–135.
Menestrel, M. Le és Lemaire, B. 2006. Biased extensive measurement: The general case. *Journal of Mathematical Psychology 50*, 570–581. Az Interneten 2016. szeptember 9-én: http://www.econ.upf.edu/~lemenestrel/IMG/pdf/jmp2006.pdf
Michell, J. 1990. *An Introduction to the Logic of Psychological Measurement*. Lawrence Erlbaum Associates, Hillsdale.
Michell, J. 1997. Bertrand Russell's Critique of the Traditional Theory of Measurement. *Synthese, 110*. 257-276.
Michell, J. 1999. Measurement in Psychology: Critical History of a Methodological Concept. Cambridge University Press, New York.
Michell, J. 2000. Normal Science, Pathological Science and Psychometrics. *Theory and Psychology, 10*(5), 639-667. Az Interneten 2016. szeptember 9-én: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.202.2742&rep=rep1&type=pdf
Michell, J. 2005. The logic of measurement: A realist overview. *Measurement, 38*, 285-294.
Molnár Gyöngyvér 2003. A komplex problémamegoldó képesség fejlettségét jelző tényezők. *Magyar Pedagógia, 103*(1), 81-103. Az Interneten 2016. szeptember 9-én: http://www.magyarpedagogia.hu/document/Molnar_MP1031.pdf
Molnár Gyöngyvér 2006. A Rasch-modell alkalmazása a társadalomtudományi kutatásokban. *Iskolakultúra*, 12, 99-113. Az Interneten 2016. szeptember 9-én: http://www.staff.u-szeged.hu/~gymolnar/MGy2006-12.pdf
Molnár Gyöngyvér 2013. *A Rash-modell alkalmazási lehetőségei az empirikus kutatások gyakorlatában.* Gondolat, Budapest.
Nagy József 2000a. A kritikus kognitív készségek és képességek kritériumorientált fejlesztése. *Új Pedagógiai Szemle, L*(7) 255-269. Az Interneten 2016. szeptember 9-én: http://www.edu.u-szeged.hu/difer/download/nagy_kkkkkf.pdf
Nagy József 2000b. *XXI. század és nevelés.* Osiris Kiadó, Budapest.
Nahalka I. 2002. *Hogyan alakul ki a tudás a gyerekekben? Konstruktivizmus és pedagógia.* Nemzeti Tankönyvkiadó, Budapest.
Narens, L. 1981a. A general theory of ratio scalability with remarks about the measurement-theoretic concept of meaningfulness. *Theory and Decision*, *13*, 1-70. Az Interneten 2016. szeptember 9-én: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.390.3658&rep=rep1&type=pdf
Narens, L. 1981b. On the scales of measurement. *Journal of MathematicalPsychology, 24*, 249-275. Az Interneten 2016. szeptember 9-én: http://aris.ss.uci.edu/~lnarens/1981/Narens_JMP_1981.pdf
Narens, L. 1985. *Abstract Measurement Theory.* MIT Press, Massachusetts.
Narens, L. 2002a. A Meaningful Justification for the Representational Theory of Measurement. *Journal of Mathematical Psychology, 46,* 746-768. Az Interneten 2016. szeptember 9-én: http://aris.ss.uci.edu/~lnarens/2002/Narens_JMP_2002a.pdf
Narens, L. 2002b. *Theories of Meaningfulness.* Lawrence Erlbaum Associates, Mahwah. (Google-könyv)
Newell, A., Shaw, J.C. és Simon, H.A. 1958. *Report on a General Problem-Solving Program.* The RAND Corporation, Santa Monica. Az Interneten 2016. szeptember 9-én: http://129.69.211.95/pdf/rand/ipl/P-1584_Report_On_A_General_Problem-Solving_Program_Feb59.pdf
Panayides, P. 2013. Coefficient Alpha: Interpret with Caution. *Europe's Journal of Psychology, 9*(4), 687-696. Az Interneten 2016. szeptember 9-én: http://ejop.psychopen.eu/article/viewFile/653/pdf
Perline, R. és Wright, B.D. 1979. The Rasch Model as Additive Conjoint Measurement. *Applied Psychological Measurement, 3*(2), 237-255. Az Interneten 2016. szeptember 9-én: http://www.rasch.org/memo24.htm
Postle, B.R. 2006. Working Memory as an Emergent Property of the Mind and Brain. *Neuroscience, 139*(1), 23-38. Az Interneten 2016. szeptember 9-én: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1428794/
Prinz, J.J. 2006. Is the mind really modular? In Stainton, R. (Szerk.) *Contemporary Debates in Cognitive Science* Blackwell, Oxford. 22–36. Az Interneten 2016. szeptember 9-én: http://subcortex.com/PrinzModularity.pdf
Rasch, G. 1960/1980. *Probabilistic models for some intelligence and attainment tests.* (Danish Institute for Educational Research, Copenhagen). The University of Chicago Press, Chicago
Roberts, F.S. 1979. *Measurement Theory with Applications to Decisionmaking, Utility and the Social Sciences.* Addison-Wesley, Maqssachusetts. Az Interneten 2016. szeptember 9-én: http://fitelson.org/roberts_measurement_theory.pdf

Rogers, H.J. és Hattie, J.A. 1987. A Monte Carlo Investigation of Several Person and Item Fit Statistics for Item Response Models. *Applies Psychological Measurement, 11*(1), 47-57. Az Interneten 2016. szeptember 9-én: http://conservancy.umn.edu/bitstream/handle/11299/103384/v11n1p047.pdf?sequence=1&isAllowed=y

Rossi, G. B. 2007. Measurabbility. *Measurement, 40,* 545-562.

Scott, D., és Suppes, P. 1958. Foundational aspects of theories of measurement. *Journal of Symbolic Logic*, *23*(2), 113–128. Az Interneten 2016. szeptember 9-én: http://suppes-corpus.stanford.edu/article.html?id=16

Shevlin, M., Miles, J.N.V., Davies, M.N.O. és Walker, S. 2000. Coefficient alpha: a useful indicator of reliability? *Personality and Individual Differences 28,* 229-237. Az Interneten 2016. szeptember 9-én: https://www.researchgate.net/profile/Mark_Shevlin/publication/222524414_Coefficient_alpha_a_useful_indicator_of_reliability/links/0deec5289f3570bc06000000.pdf

Sijtsma, K. 2009a. Correcting Fallacies in Validity, Reliability, and Classification. *International Journal of Testing*, 9(3), 167-194. Az Interneten 2016. szeptember 9-én: https://pure.uvt.nl/ws/files/1109845/Correcting_Fallacies_in_Validity.pdf

Sijtsma, K. 2009b. On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. Psychometrica, 74(1), 107-120. Az Interneten 2016. szeptember 9-én: http://link.springer.com/article/10.1007/s11336-008-9101-0/fulltext.html

Smith, R.M., Schumacker, R.E. és Bush, M.J. 1995. *Using Item Mean Squares to Evaluate Fit to the Rasch Model*. A paper presented at the 1995 Annual Meeting of the American Educational Research Assosiation in San Francisco. ERIC Documentation reproduction Service. Az Interneten 2016. szeptember 9-én: http://files.eric.ed.gov/fulltext/ED384617.pdf

Stanfill, C. 1987. Memory-Based Reasoning Applied to English Prononciation. *AAAI-87 Proceedings*. 577-581. Az Interneten 2016. szeptember 9-én: http://www.aaai.org/Papers/AAAI/1987/AAAI87-103.pdf

Stanfill, C. és Waltz, D. 1986. Toward Memory-Based Reasoning. *Communications of the ACM, 29*(12), 1213-1228. Az Interneten 2016. szeptember 9-én: http://128.59.11.206/~waltz/Papers/Torward%20Memory-Based%20Reasoning-ACM%201986.pdf

Steegen, S., és De Neys, W. 2012. Belief inhibition in children's reasoning: Memory-based evidence. *Journal of Experimental Child Psychology*, *112*, 231-242. Az Interneten 2016. szeptember 9-én: https://www.researchgate.net/profile/Wim_De_Neys/publication/221688391_Belief_inhibition_in_childrens_reasoning_Memory-based_evidence/links/0046353106006f05d0000000.pdf

Stevens, S. S. 1946. On the theory of scales of measurement. *Science*, 103, 667-680. Az Interneten 2016. szeptember 9-én: http://www.mpopa.ro/statistica_licenta/Stevens_Measurement.pdf

Steyer, R. 1989. Models of Classical Psychometrical Test Theory as Stochastic measurement Models: Representation, Uniqueness, Meaningfulness, Identifiability, and Testability. *Methodika, 111,* 25-60. Az Interneten 2016. szeptember 9-én: http://www.metheval.uni-jena.de/materialien/publikationen/steyer1989_models_of_classical_psychometric_test_theory.pdf

Steyer, R. 2001. Classical (psychometric) test theory. In Cook, T. és Ragin, C. (Szerk.) *International encyclopedia of the social and behavioral sciences. Logic of inquiry and research design*. Pergamon, Oxford. 1955 – 1962. Az Interneten 2016. szeptember 9-én: http://www.metheval.uni-jena.de/materialien/publikationen/ctt.pdf

Suppes, P. 1951. A Set of Independent Axioms for Extensive Quantities. *Portugaliae Mathematica, 10,* 163-172. Az Interneten 2016. szeptember 9-én: http://suppes-corpus.stanford.edu/article.html?id=1

Taagepera, M., Potter, F., Miller, E.G. és Lakshminarayan, K. 1997. Mapping students' thinking patterns by the use of the knowledge space theory. *International Journal of Science Education, 3*, 283–302.

Taatgen, N.A. 1997. A rational analysis of alternating search and reflection strategies in problem solving. Proceedings of the 19th Annual Conference of the Cognitive Science Society. Erlbaum, Hillsdale. Az Interneten 2016. szeptember 9-én: http://cogprints.org/684/3/strat.pdf

Thorsen, S.V. és Bjorner, J.B. 2000. Reliability of the Copenhagen Psychosocial Questionnaire. *Scandinavian Journal of Public Health, 38*(3), 25–32. Az Interneten 2016. szeptember 9-én: https://www.researchgate.net/profile/Sannie_Vester_Thorsen/publication/49697246_Reliability_of_the_Copenhagen_Psychosocial_Questionnaire/links/0046352a02f3aa663a000000.pdf

Thurstone, L.L. 1938. *Primary mental abilities*. Psychometric Monographs, No. 1. University of Chicago Press, Chicago.

Tooby, J. és Cosmides, L. 1992. The psychological foundations of culture. In: Barkow, J.H. és mts. (Szerk.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press, New York. 19–136. Az Interneten 2016. szeptember 9-én: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.6260&rep=rep1&type=pdf

Tóth Zoltán 2005. A tudásszerkezet és a tudás szerveződésének vizsgálata a tudástér-elmélet alapján. *Magyar Pedagógia, 105*(1), 59-82. Az Interneten 2016. szeptember 9-én: http://www.magyarpedagogia.hu/document/Toth_MP1051.pdf

Trendler, G. 2009. Measurement Theory, Psychology and the Revolution That Cannot Happen. *Theory and Psychology, 19*(5), 579-599. Az Interneten 2016. szeptember 9-én: http://psychology.okstate.edu/faculty/jgrice/psyc4333/Trendler2009.pdf

Ursin, J. 2015. A nemzetközi összehasonlító mérés, mint politikai eszköz – Kritikai megközelítés. *Educatio,* 2. 64–70.

Varga József, Józsa Krisztián és Pap-Szigeti Róbert 2007. Az arányosságszámítási készség kritériumorientált fejlesztése 7. osztályban. *Magyar Pedagógia, 107*(1) 5-27. Az Interneten 2016. szeptember 9-én: http://www.magyarpedagogia.hu/document/Varga_MP1071.pdf

Velleman, P.F. és Wilkinson, L. 1993. Nominal, Ordinal, Interval, and Ratio Typologies are Misleading. *The American Statistician, 47*(1), 65-72. Az Interneten 2016. szeptember 9-én: http://gaius.fpce.uc.pt/niips/novoplano/mip1/mip1_201314/scales/Velleman.pdf

Walker, C.M. és Beretvas, S.N. 2000. *Using Multidimensional vesrus Unidimensional Ability Estimates to Determine Student Proficiency in Mathematics.* Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, April 24-29, 2000 (ED 442 821) Az Interneten 2016. szeptember 9-én: http://www.eric.ed.gov/PDFS/ED442821.pdf

Wright, B.D. 1983. *Fundamental Measurement in Social Science and Education*. Research Memorandum No. 33a. MESA Psychometric Laboratory.

Wright, B.D. és Douglas, G.A. 1977. Best Procedures For Sample-Free Item Analysis. *Applied Psychological Measurement, 1*(2), 281-295. Az Interneten 2016. szeptember 9-én: http://conservancy.umn.edu/bitstream/handle/11299/98553/1/v01n2p281.pdf

Wright, B.D. és Stone, M.H. 1979. *Best test design. Rasch Measurement.* MESA Press, Chicago.

Wuttke, J. 2007. Uncertainties and Bias in PISA. In: Hopmann S.T., Brinek, G. és Retzl, M. (szerk.) PISA zufolge PISA. PISA According to PISA. Hält PISA, was es verspricht? Does PISA Keep What It Promises? Lit-Verlag, Wien. 241-264.

Zhong, N., Wang, Y. és Chiew, V. 2010. On the cognitive process of human problem solving. *Cognitive System Research, 11.*, 81-92. Az Interneten 2016. szeptember 9-én: http://www.ucalgary.ca/icic/files/icic/61-Elsevier-CogSys-ProblemSolving.pdf