

Az objektív mérés lehetősége: a Rasch-modell

Az objektív mérés lehetősége a pedagógiában kulcsfontosságú kérdés, amely néhány évtizede foglalkoztatja már a pedagógiai kutatókat, de a végső megoldás, az objektív, adaptív skálák megteremtése még várat magára.

Az objektív mérést a természettudósok már rutinszerűen alkalmazzák. Gondoljunk csak a súly, a hosszúság, a mennyiség, a tömeg, az idő mérésére. Ez azonban a természettudományokban sem volt mindig így, hosszú folyamat eredménye, amíg kialakultak ezek a mérőeszközök, hiteles, egységesített skálák, beosztások.

Például az idő standardizálásnak első lépcsőfoka a naplementéhez, illetve napfelkeltéhez való igazodást jelentett. A 17. században *Galilei* már vízórával hasonlította össze a különböző mozgások időtartamát. „Felakasztottunk egy tekintélyes vödröt vízzel tele, jó magasra, amelynek aljából, egy nyíláson keresztül, a víz vékony fonál alakjában folydogált, ezt a vizet fogtuk fel egy kis edényben, míg a golyó a lejtőt vagy annak egy részét befutotta. Időről időre megmértük ezen kis vízmennyiségeket, melyeket így gyűjtöttünk, egy igen pontos mérlegen. Ezek súlyának különbségét és viszonyát adta; és ezt olyan pontossággal, hogy – bármennyiszor is ismételtük meg a kísérletet, soha nem tértek el egymástól.” (*Simonyi*, 1986, 192.) A pontosabb időméréshez az ingák mozgásának vizsgálata vezetett el. 1657-ben *Christiaan Huygens* szabadalmaztatta az első ingaórát, amelynek mozgása a Föld mozgásával függ össze, azonban a Föld mozgása nem egyenletes, állandó, aminek következtében az ingaóra sem pontos, így nem lehet a pontos időmérés alapja sem. 1967-ben szabadalmaztatták az atomórát, amely az abszolút nulla fok közvetlen közelébe hűtött céziumlabda (cézium 133) periódusidejét méri meg rendkívüli pontossággal, és ettől fogva ehhez kötötték a másodperc meghatározását. Azonban még ebben sem bíztak maradéktalanul a kutatók, ezért a világon felállított 200 atomóra küld információt egy párizsi obszervatóriumba, ahol a beérkezett adatokat átlagolják és így kapjuk meg a pontos időt. Az időmérés történetében még kiemelhetnénk a pulzusunkkal összefüggő időmérést, a gyertyaórát, a homokórát, a kerekés órát, és még sorolhatnánk a különböző elven nyugvó időmérők sorát. (*Simonyi*, 1986; *Greguss*, 1985)

A fizikai mérések közül kitérhetnénk például a súly-, a tömeg-, a hosszúság és mennyiségmérésre, amelyek egy-egy hasznos absztrakcióval a különböző méretű tárgyak problémáját oldják meg egyforma egységekre osztott skálával. Ezt kellene a társadalomtudományok terén is tenni, nagyminták alapján azonos, reprodukálható egységek absztrakcióját megalkotni, kalibrálni, hogy biztosak lehessünk a használhatóságban.

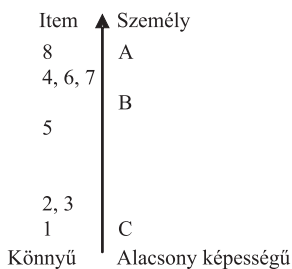
Az objektivitás mellett a természettudományok fejlődéséhez hasonlóan felmerül a determinisztikusság és valószínűség kérdése is. Magyarországon a klasszikus tesztelméleti módszerekkel történő elemzéseknek jelentős múltja van, de az utóbbi évek nemzetközi vizsgálatainak elemzései rávilágítanak egy alapjaiban más módszerekkel, más alapokon nyugvó tesztelmélet fontosságára. Ez a más módszer a tesztelméletek újabb generációját képező modern (probabilisztikus, valószínűségi) tesztelmélet (*Item Response Theory [IRT]*), amely az itemek tulajdonságait valószínűségelméleti eszközökkel jellemzi. A

modern tesztelmélet nem a klasszikus tesztelmélet egy továbbfejlesztett vagy „jobb” változata, hanem alapvetően más matematikai eszközökre támaszkodó, statisztikai eljárásokat használó, modelleket felállító és függvényekkel dolgozó tesztelmélet. A valószínűségi alapokon nyugvó megközelítésre a természettudományok területén is várni kellett, hiszen *Arisztotelész*, *Aquinói Tamás*, Galilei és *Descartes* törvényei, megközelítései még determinisztikus szemléletűek voltak, amelyeket csak a 17. századtól kezdve követte *Newton*, *Maxwell*, *Planck*, *Einstein*, *Heisenberg* valószínűségi megközelítése. (Simonyi, 1986)

A Rasch-modell alapelvei

A Rasch-modell azon az elképzelésen alapul, hogy az adatokban egyféle logikus hierarchiának kell lennie (kevesebb, mint – több, mint). Például, ha valaki a diákok problémamegoldó képességét szeretné megmérni, akkor az eredményben lesznek jobb, illetve kevésbé jó problémamegoldó képességgel rendelkező diákok. Bár minden egyes diák számos lényeges képességgel rendelkezik, egyszerre értelmesen csak egy tulajdonság jellemezhető. Ezáltal az eredmény modellezhető egy egyenes mentén, ahol a kevesebb felől haladunk a több felé, mint a számegyenesen.

Ezt az elképzelést egy egyszerű eljárással átültették egy matematikai modellbe. A Rasch-modell kiindulópontként a diákok teszten elért összpontszámát számolja ki a helyes, illetve helytelen válaszok valószínűségének megadásához. Ezek után arra az egyszerű gondolatra alapoz, hogy az emberek sokkal nagyobb valószínűséggel teljesítenek jól a könnyű, mint a nehéz itemeken, valamint a magasabb képességszintű emberek nagyobb valószínűséggel oldják meg jól a feladatokat, mint az alacsonyabb képességszintűek. Hasonlóképpen azokat az itemeket veszi nehéznek a modell, amelyeken kevesebben teljesítenek jól és azokat sorolja a könnyűek közé, amelyeket sokan jól megoldanak. Ezt ábrázolja közös skálán a személy és itemtérkép. Leegyszerűsítetten mutatja ezt az 1. ábra, amelyen három diákkal (A, B, C) és 7 itemmel modelleztük a személy-/itemtérképet. A modell egyértelműen jelzi a tesztfejlesztőnek, hogy (a) melyik item nehezebb és melyik item könnyebb, melyik személy magasabb, melyik alacsonyabb képességű, (b) milyen nehézségű itemek hiányoznak a tesztből, (c) mennyire felel meg a teszt nehézsége a diákok képességszintjének. A továbbiakban egy sétálóutca analógiájával modellezzük a személy- és itemtérkép továbbfejlesztett változatát.



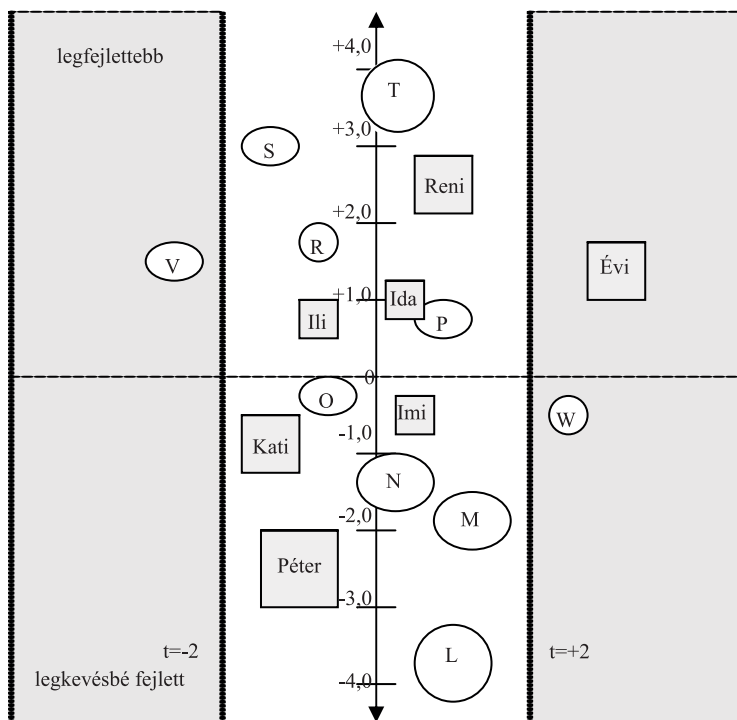
1. ábra. 7 item nehézségi szintje és 3 diák képességszintje közös skálán

A sétálóutca analógiája

Nagyon leegyszerűsítve képzeljünk egy utcát, ahol az utca elején a gyengébb, a végén a magasabb képességszintű gyerekek haladnak. Az utca különböző nehézségű, különböző fejlődési szinteket reprezentáló, egyre nehezedő feladatokkal van kikövezve, amelyeket meg kell oldaniuk a diákoknak. Az egyes feladatokon mutatott eredmény fényében

haladnak tovább, jobbra vagy balra, egészen odáig, amíg a fejlettségük viszi, azaz amíg elérik az utca azon pontját, ahol a hozzájuk azonos képességszintű diákok állnak. Ezáltal az utca minden egyes pontja megfeleltethető egy képességszintnek, az adott személy, illetve az item 50 százalékos valószínűséggel történő megoldásához szükséges képességszint reprezentációjának. Ennek megfelelően minden tanulónál megvannak a képességének megfelelő itemek, és minél inkább ebből a nehézségi tartományból kerülnek ki a teszt itemei, annál nagyobb valószínűséggel oldja meg azokat (zone of success), valamint minél inkább a hibazónában van egy item, annál nagyobb valószínűséggel ront az itemen (zone of failure). (Bond és Fox, 2001)

A 2. ábra egy ilyen sétálóutcat modellez. Az ellipszis, illetve kör alakú kövek a teszt egyes itemeit reprezentálják (L, M, N, O, P ...), a négyzetekbe írt nevek pedig a feladatokot megoldó diákokat. Minél közelebb van egy itemet reprezentáló kő a sétálóutca elejéhez, annál könnyebb az adott item és minél feljebb van, annál nehezebb. Az itemek reprezentálásához hasonlóan a sétálóutca legalsó részén az alacsonyabb (Péter, Kati), majd felfelé haladva az egyre magasabb képességszintű diákok (Ili, Reni) állnak. E párhuzamosság, egymásra vetítettség magyarázza azt, hogy miért lényeges, hogy a lehető legjobban lefedjük a diákok által közrefogott képességskála teljes intervallumát.



2. ábra. A fejlődési pálya sétálóutca analógiája (Bond és Fox, 2001 ötlete alapján)

Az itemeket reprezentáló kövek távolsága a sétálóutca elejétől meghatározza, hogy a másik itemhez képest mennyivel nehezebb az adott item. A Rasch-moddal elemző szoftverek logaritmikus transzformációt hajtanak végre az item és személyadatokon, azaz az ordinális skálán lévő adatokat átranszformálják intervallumskálára, ezért a térkép alapján nem csak az mondható el, hogy az egyik feladat nehezebb, mint a másik vagy az egyik diák jobb képességű, mint a másik, hanem azt is meg tudjuk mondani, hogy meny-

nyível könnyebb-nehezebb az adott item, illetve mennyivel jobb-rosszabb képességű az érintett diák. Az itemekre vonatkoztatva a vertikális skálán ezt a mutatót nevezzük item-nehézségi indexnek, míg az emberekre nézve a személy képességparaméterének, a skálát pedig logit skálának.

Ez a féle ábrázolási mód, ahol az item- és személytérkép kapcsolatát is leolvashatjuk, számos, a Rasch mérésben központi szerepet játszó információval szolgál. Felsorolunk néhány kérdést, amelyekre a válasz a 2. ábráról leolvasható.

- Melyik item a nehezebb, az L, az N vagy az T?
- Melyik itemet oldják meg legnagyobb valószínűséggel helyesen a diákok?
- Melyik itemen rontanak legnagyobb valószínűséggel a diákok?
- Vajon Imi magasabb képességű-e, mint Ili?
- Melyik diák teljesített legrosszabbul ezen a teszten?
- Vajon Reni helyesen oldotta-e meg a T itemet?
- Melyiket várhatjuk el inkább, hogy Reni az R vagy az M itemet oldja meg helyesen?
- Péter 1 pontot ért el a teszten, melyik itemet oldhatta meg helyesen?
- Ki az, aki nem ugyanolyan módon járta be a sétálóutcat, mint a többiek? (például

puskázott, csalt)

- Imi képességszintjét vajon jól mérte-e ez a teszt?

A tesztfejlesztésre vonatkozó néhány kérdés:

- Milyen nehézségű itemek hiányoznak még a tesztből?
- Az itemek nehézsége mennyire felel meg a minta képességszintjének?
- Melyik itemek azok, amelyek nem hasznosak számunkra és jelenlegi formájukban

törölhetőek?

A felvetett kérdésekre röviden válaszolva: az item-személy térkép alapján az O item nehezebb, mint az N, a teszt legnehezebb iteme pedig a T, amit az ábrán reprezentált diákok legnagyobb része nem old meg helyesen. (Az ábrán a legjobb képességű diák Reni, aki 25 százalékos valószínűséggel teljesít jól ezen az itemen, a többiek ennél jóval kisebb valószínűséggel oldják meg helyesen ezt az itemet. Általánosságban megfogalmazható, hogy aki jól oldja meg ezt az itemet, magasabb képességszintű, mint Reni.) Ezzel szemben minden egyes diák több mint 75 százalékos valószínűséggel teljesít jól az L itemen, bár az sem kizárt, hogy épp Reni, a modellen ábrázolt legjobb képességű diák ront ezen az itemen. Reni képességszintjéhez legközelebb az R és az S item áll. A térkép alapján Ili nagy valószínűséggel jobban teljesít ezen a teszten, mint Imi, jókora különbség van kettőjük képességszintje között. A legrosszabbul teljesítő diák pedig Péter, aki nagy valószínűséggel egy pontját az L item helyes megoldásával érte el. Évi a sétálóutca határain kívül van, ami arra utal, hogy más módon használta a tesztet, mint a többiek. Ezzel az ábrán szürkével sátozott résszel, a későbbiekben még külön foglalkozunk. Évivel szemben Imi a sétálóutca területén helyezkedik el, az ő képességszintjét jól mérte a teszt.

A teszt esetleges továbbfejlesztésére vonatkozott az itemek nehézségének homogenitását érintő kérdés, azaz, hogy az itemnehézségi indexek lefedik-e a diákok képességszintjei által meghatározott képességszála-intervallumot. A modell alapján még ki lehetne egészíteni néhány itemmel a tesztet, például a túl könnyű L itemet egy kicsit nehezebbre cserélve – aminek nehézségi indexe közelíti Péter képességszintjét – vagy a 4 logit-os nehézséget közelítő T itemet egy könnyebbre cserélve, aminek nehézségi indexe az S és az R item nehézségi indexe között van. A már érintett szürke sávban találhatóak a V és a W itemek, ami azt jelzi, hogy ez a két item mást mért, mint a többi item, ezért egy esetleges tesztfejlesztés során kicserélendőek. A végső cél, a képességszála teljes lefedése, elegendő sok lépéskö letétele, aminek megvalósításához, az itemek nehézségi indexének meghatározásához egy elég nagy mintától begyűjtött adatra van szükség.

Az eddig feltett kérdésekre adott válaszok egy része a klasszikus tesztelemzési módszerekkel is megadható, azonban ezen a ponton kiemelnék egy példát, amit a klasszikus

módszerek nem tudnak kezelni. Tegyük fel, hogy Évi 6 pontot ért el a teszten, Ili pedig 5-öt. Évi összpontszámát az L, V, P, R, S, T itemek, Ili pedig az L, M, N, O, P itemek helyes megoldásával érte el. A klasszikus elemzések csak azt mutatják, hogy Évi több pontot ért el, mint Ili, azaz jobb képességű, de nem vizsgálják azt, hogy melyik itemeken érte el azt a 6 pontot. A modell megmutatja, hogy Évi a nehezebb itemeket oldotta meg helyesen, a könnyebb itemeken rontott. Ennek több oka lehet, például a koncentrációzavar vagy a könnyebb itemekben szereplő ismeretek hiánya (információ a tanárnak, hogy mit kell pótolni Évinél), vagy esetleg találgatott a nehezebb itemek megoldásánál, vagy puszkázott. Konkrét választ nem tudunk adni a személy-item térkép alapján erre a kérdésre, mindenesetre teljesítménye nem illik a modellbe, a 6 pont által reprezentált képességszint semmi esetre sem tükrözi valódi képességszintjét.

Egy másik példával élve, amit szintén nem tudnak kezelni a klasszikus módszerek, előfordulhat az is, hogy az azonos vagy kevesebb pontszámot elérő diák mutat magasabb képességszintet. Például Ida, aki az M, N, O, P, R itemek helyes megoldásával Ilihez hasonlóan 5 pontot ért el, de mivel magasabb képességszintet igénylő feladatot is megoldott (R), ezért képességszintje is magasabb lett. Ida esete különbözik Éviétől, hiszen Idánál nem tapasztalható olyan nagy ugrás a jól megoldott feladatok között, mint Évinél (L és V item között közel 5 logit távolság van), ezért az ő képességszintjét a teszt jól mérte.

Itemilleszkedés (itemfit)

Az itemilleszkedés problémáját már a korábbi fejezetekben is érintettük, amikor a szaggatott vonalon kívül eső, a szürke területen lévő itemekről és diákokról beszéltünk. Az item modell-illeszkedése a modell által elvárt, előre jelzett és a valós teljesítmény közötti különbséget mutatja.

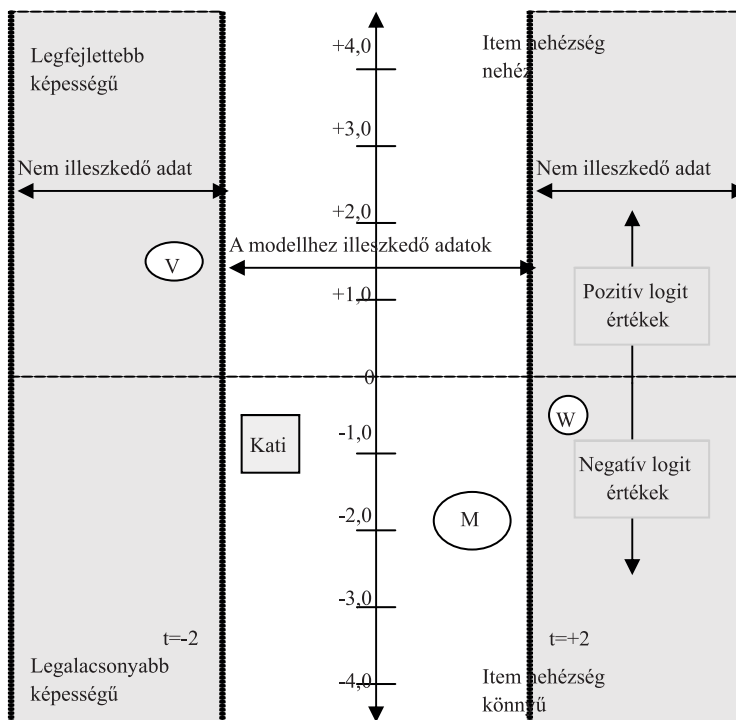
A képességszint horizontális mozgásával szemben az illeszkedés esetében vertikális mozgásról beszélünk. Egy item annál jobban illeszkedik a vizsgált képességterületre vonatkozó adatok által meghatározott modellbe, minél közelebb van az itemet reprezentáló kő a sétálóutca képzeletbeli középvonalához. (Az M, N, O itemek nem pontosan a sétálóutca közepén meghúzott vonalon fekszenek, mégis jól illeszkednek a modellhez, azonos képességterületet mérnek.) Ezzel szemben, ha egy item a sétálóutcán kívülre esik, akkor nem illeszkedik a modellbe, ezért célszerű a diákok képességszintjének meghatározásakor ezeket az itemeket elhagyni és esetleg egy más tesztben alkalmazni. (3. ábra) Ezek az itemek (V, W) más képességterületet (is) mérnek, mint a tesztben előforduló többi item.

Hogy könnyebben el lehessen dönteni, melyek a modellbe nem illeszkedő, illetve illeszkedő itemek, egy-egy szaggatott vonalat húztunk a sétálóutca két oldalára, jelezve az illeszkedés határát – hasonlóan a 95 százalékos konfidencia-intervallumhoz. Néhány Rasch-modellel dolgozó szoftver ezt meg is teszi és a 4. ábrán látható formában ábrázolja az itemek modellilleszkedését. Természetesen minden Rasch-modellel dolgozó szoftver kiszámolja az illeszkedési paramétereket, csak külön táblázatos formában közli. (Az infit paraméterek – az illeszkedést mutatják – nem táblázatos, hanem grafikus prezentációjának bemutatásához a Rasch-modellel dolgozó Quest programot használtuk. A Quest program néhány paraméterében eltér a ConQuest program beállításaitól, ugyanis a Quest az infit paraméterek átlagát automatikusan 1-nek veszi és nem 0-nak, ahogy azt a ConQuestnél láthattuk. Ebből adódóan a 0,70 és a 1,30 közötti értékek fogadhatóak el, az 1,30 feletti és a 0,70 alattiak és a (-2, +2) intervallumba tartozó értékek nem.)

Nehézség-, képességbecslés és a hiba

A Rasch-modellel dolgozó szoftverek alapelvei:

– A magasabb képességűek nagyobb valószínűséggel oldják meg a teszt itemeit helyesen (pl.: 2. ábra: Reni válaszai nagyobb valószínűséggel jók, mint Péter válaszai).



3. ábra. A sétálóútcán modellezett modell-illeszkedés

INFIT	0,63	0,71	0,83	1,00	1,20	1,40
MNSQ						
1 item 10					*	
2 item 13		*				
3 item 14			*			
4 item 15		*				
5 item 17			*			
6 item 18					*	
7 item 19		*				
8 item 20				*		
9 item 21				*		
10 item 09					*	
11 item 22			*			
12 item 23					*	
13 item 24					*	
14 item 25				*		
15 item 33				*		
16 item 34				*		
17 item 35				*		
18 item 36					*	
19 item 37		*				
20 item 38					*	
21 item 39					*	
22 item 40					*	
23 item 30					*	
24 item 31					*	
25 item 32					*	
26 item 29					*	
27 item 28			*			
28 item 26			*			
29 item 27		*				

4. ábra. A Quest program infit paraméterekre vonatkozó grafikus outputja

– A könnyebb itemeket nagyobb valószínűséggel oldja meg jól mindenki, mint a nehezebbeket (például: Ili és Ida nagyobb valószínűséggel teljesítenek jól az L itemen, mint a P itemen és nagyobb valószínűséggel teljesítenek jól a P itemen, mint a T itemen).

Az adatok logaritmikus transzformációval logitskálára konvertálását felhasználva alapbeállításban úgy határozzák meg egy személy és item képesség-, illetve nehézségparaméterét, hogy a közös logit skálán azokat az itemeket és személyeket teszik azonos szintre, ahol az adott személy képességparamétere alapján 50 százalékos valószínűséggel oldja meg helyesen az adott itemet. (Ili képességparaméterének becslése megegyezik a P item nehézségi indexének becslésével, azaz Ili 50 százalékos valószínűséggel oldja meg jól a P itemet.) Ez a valószínűség 75 százalékra nő azon itemek esetében, amelyek 1 logittal könnyebbek (pl.: item O) és 25 százalékra csökken azon itemeknél, amelyek 1 logittal nehezebbek (pl.: item R).

A horizontális és vertikális mozgáson kívül még egy változóval találkozhatunk a sétálóutca analógiája során. (2. ábra) A kövek vízszintes és függőleges helyzetén kívül még eltérhetnek nagyságukban is (Lásd L és R itemet). A kövek nagysága modellezi az elkövetett hiba nagyságát, egyféle pontatlanság zónát („error”, „zone of imprecision”). Minél kisebb a kő, annál kisebb az elkövetett hiba nagysága, annál pontosabban meg tudjuk mondani az adott item pontos helyzetét. Minél nagyobb a kő, annál nagyobb az elkövethető hiba nagysága, az item megadott helye kevésbé pontosan reprezentál egy pontot.

A 2. ábrán megfigyelhető, hogy azon kövek nagysága (például: O, P, R) kisebb, amelyek közelében képességszint alapján több diák található (Imi, Ili, Ida, Kati, Reni). Válaszaik elegendő információval szolgálnak az adott item nehézségének pontosabb becsléséhez. Ezzel szemben az L, M, N és T itemek nehézségi indexét relatív nagy hibával tudtuk megadni, mivel a szimulált modell mintájában nagyon kevés diák képességszintje közelíti ezen itemek nehézségi szintjét, ezért ezen itemek elhelyezésénél nagyobb szerepet játszott a találgatás.

Az itemekhez hasonlóan minden egyes diák képességszintjének meghatározása is magában hordoz bizonyos méretű hibafaktort. Például a Pétert reprezentáló kő nagyobb, mint az Imi reprezentáló kő. Péter képességszintjének meghatározása több bizonytalanságot hordoz magában, a teszt kevés olyan nehézségű itemet tartalmaz, ami megegyezne vagy közel állna az ő képességszintjéhez. Ezzel szemben Imi képességszintjének közelében több item található, amelyek részletesebb információval szolgálnak képességszintjének pontosabb megadásához.

A mérés pontossága függ a tesztet kitöltő együttműködő-készségétől is. Ha valaki találgat, lemásolja a szomszédjáról, puskázik, emlékezetből próbál például problémákat megoldani, olvasási nehézséggel küzd vagy koncentrációproblémája van stb. eredménye alapján becsült képességszintje nem a valós képességszintjének megfelelő szintet mutatja. E jól ismert problémák ellenére is, amelyek egy részére a korábbiakban említett módon fényt lehet deríteni, törekednünk kell a legpontosabb becslésre. (Bond és Fox, 2001)

Reliabilitás

Tegyük fel, hogy a teszt készítői nem tettek le elegendő „követ” a sétálóutcára. Ennek első következménye az, hogy sem az itemparaméterek, sem a személyparaméterek megadott helye nem eléggé pontos. Nincs elegendő item a fejlettségi szintek minél sokrétűbb elkülönítésére, aminek hatására a diákok csoportokban helyezkednek el a képességskálán.

A Rasch-modell segít a tesztfejlesztőnek, hogy eldöntse, elegendő és megfelelő itemet tartalmaz-e a tesztje, valamint, hogy a minta képességeloszlása elég nagy-e. A személy reliabilitása (person reliability index) a személy képességparaméterének megismételhetőségét jellemzi: ha ugyanazon mintának más, ugyanazon képességet mérő tesztet adunk, bizonyos hibahatáron belül ugyanaz lesz-e a diákok képességparamétere.

(Bond és Fox, 2001). Ha az egyik teszten Reni képességparamétere magasabb, mint Idáé, akkor a másik teszten is fennáll-e ez a kapcsolat. A minta reliabilitásának pontos becslése nem csak egy gazdag item-poolt feltételez, hanem egy széles képességskálán szóródó nagy mintát is.

Az item reliabilitása (item reliability index) pedig az item nehézségi paraméterének megismételhetősége: ugyanazokat az itemeket megírattatjuk egy másik, a mintánkkal összehasonlítható képességű mintával. (Bond és Fox, 2001) Vajon ha P item az eredeti mintában nehezebb, mint az L item, akkor ez az állítás fennáll-e az újabb minta esetében is? Ennek ellenőrzésére nagy minta szükséges.

Egy- és többdimenzionalitás

A képességek fejlődésének fent említett mérése egydimenziós, azaz hasonlít a méret, súly, hőmérséklet méréséhez, amikor egyszerre csak egy sajátosságot mérünk. De ha a tárgyak, emberek, vagy akár az időjárás fizikai tulajdonságait is mérni szeretnénk, a legtöbb esetben nem elegendő az egy dimenzió. (Bond és Fox, 2001)

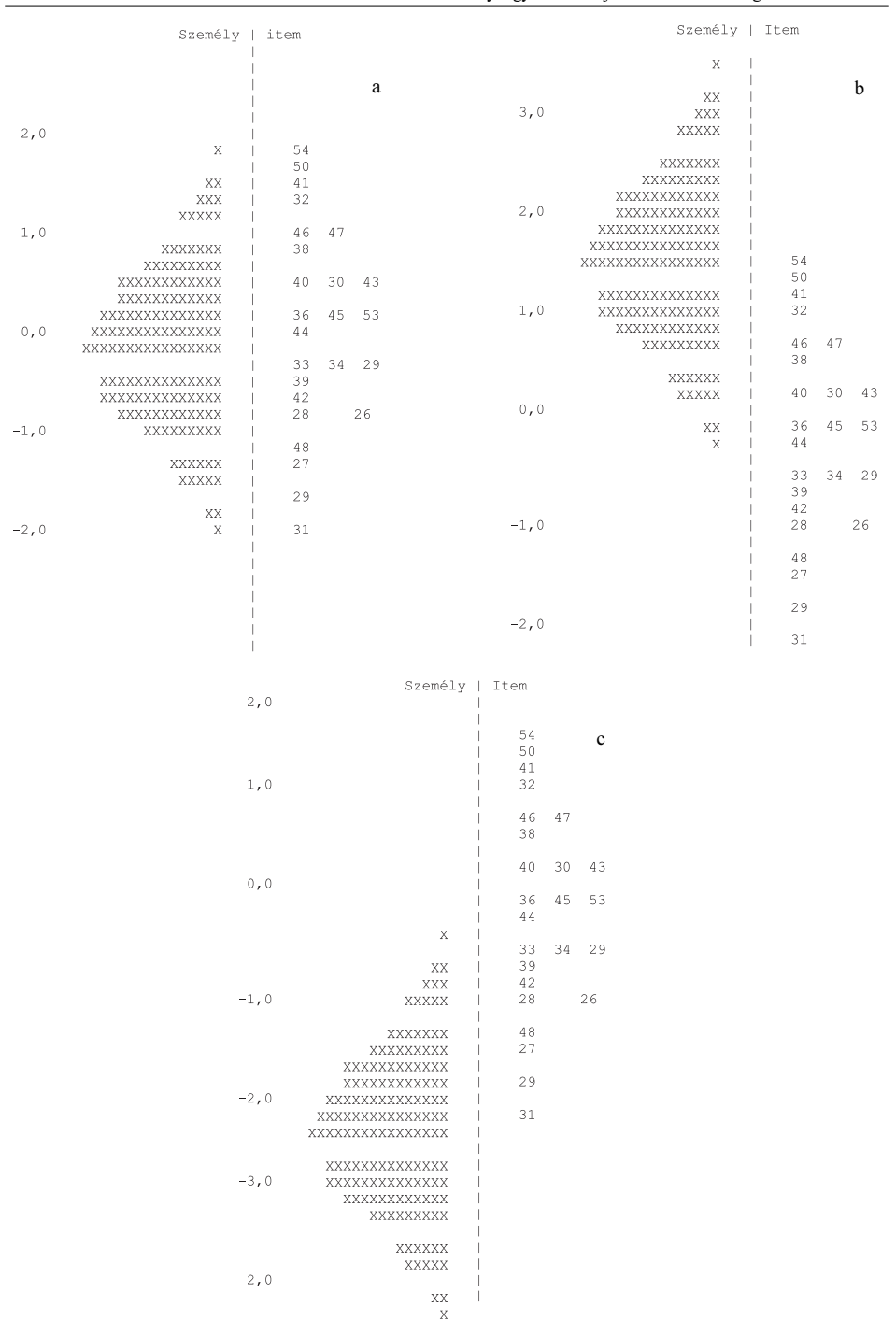
Egydimenziós mérést végzünk, ha megmérjük az emberek súlyát, magasságát, derék-bősségét, mellbősségét, vérnyomását stb., de már új, kétdimenziós skálát hozunk létre, ha a skála létrehozása során figyelembe vesszük a magasságot és a súlyt is. A mindennapi életben is találkozunk ehhez hasonló egy-, illetve kétdimenziós skálákkal. Például egydimenziós skálával találkozunk cipővásárlásnál, ahol megadják a cipő hosszának méretét (36, 37). Néhány országban a cipőméret nem csak egydimenziós, hanem kettő (9A, 9B..), mivel a hosszúsága mellett a szélességét is tartalmazza a cipő méretét jellemző skála. Magyarországon is találkozhatunk az üzletekben kétdimenziós skálákkal, például néhány farmermárkánál a méret a derék-bősséget és a szár hosszúságát is tartalmazza (W27, L28) vagy a melltartóméret (75A, 75B...75F) is két méretet foglal magába (mellbőség, kosárméret).

Ennek ellenére tapasztalatból tudhatjuk, hogy nem elegendő a ruha vagy cipő megadott méretére hagyatkozni, érdemes felpróbálni azt, mivel előfordulhat, hogy a megfelelő hosszúságú cipő szűk, alacsony a rüsztye, magas a sarka... Már egy cipő vásárlásakor szembesülünk azzal a problémával, mint a képességek mérése során. Az ember sokdimenziós, összetett lény. Tudjuk, hogy az ember komplexitását soha nem fogjuk kielégítően leírni egy teszt eredményével, de kifejleszthetünk néhány használható, az emberi tulajdonságokra, képességekre vonatkozó skálát. A skálák kialakítása során szem előtt kell tartanunk, hogy egyszerre csak egy tulajdonságot, illetve képességet mérhetünk megfelelő hatékonysággal, pontossággal.

A ConQuest program képesség- és nehézségi indexre vonatkozó grafikus outputjának értelmezése

A Rasch-moddal dolgozó szoftverek nem képesek arra, hogy az itemek és a minta különböző tulajdonságait közös ábrán, a sétálóutca analógiához hasonlóan bemutassák. A 4. ábrán bemutattuk, hogyan ábrázolja a program az egyes itemek modell-illeszkedését, a továbbiakban kitérünk a képességszintek és nehézségi indexek ábrázolási és értelmezési módjára.

Az 5.a, b, c ábra a személy képességparaméterek és az item nehézségi mutatók közötti lehetséges háromféle relációt mutatja. Egy, a mintához jól illesztett teszt során a személyparaméterek átlaga (mintaátlag) közelíti a nullát. (5.a ábra) Ebben az esetben a személyeket reprezentáló X-ek és az itemeket reprezentáló számok a „fa” két oldalán párhuzamosan futnak. Ha a teszt a mintának túl nehéz, akkor az átlagos képességparaméter egy nullától távolabb eső negatív szám (5.c ábra), ha a teszt túl könnyű, akkor egy nagyobb pozitív szám. (5.b ábra)



5a, b, c ábra. Személy-ítem térkép egy a) a mintához jól illeszkedő, b) a mintának túl könnyű és c) a mintának túl nehéz teszt esetén

Ezekről a „fa” ábráról továbbá leolvasható mindazon információ, amelyeket a sétáló-utca analógiánál a képesség és nehézségi indexekkel kapcsolatosan érintettünk. Hol van ugrás az egyes itemek nehézségi indexei között? Honnan hiányzik még item és esetleg milyen nehézségű itemből tartalmaz többet a teszt? A diákok képességszintjének megfelelő-e a teszt? Milyen mintában lehetne még alkalmazni a tesztet (jobb-rosszabb képességűeknél)? E kérdések megválaszolására konkrét empirikus adatok elemzése révén láthatunk példát *Molnár* (2003) tanulmányában.

Irodalom

Bond, T. – Fox, C. M. (2001): *Applying The Rasch Model. Fundamental Measurement in the Human Sciences*. Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey.

Greguss Ferenc (1985): *Élhetetlen feltalálók, halhatatlan találmányok*. Móra Ferenc Ifjúsági Könyvkiadó, Budapest.

Molnár Gyöngyvér (2003): Az ismeretek alkalmazásának vizsgálata modern tesztelméleti eszközökkel. *Magyar Pedagógia*, 103. 4.

Simonyi Károly (1986): *A fizika kultúrtörténete*. Gondolat Kiadó, Budapest.

Wu, M. – Adams, R. J. – Wilson, M. R. (1998): *ACER ConQuest. Generalised Item Response Modelling Software*. ACER Press, Australia.



Az OKI könyveiből