

A Rasch-modell kiterjesztése nem dichotóm adatok elemzésére: a rangskálás és a parciális kredit modell

A tesztelméletek újabb, a nemzetközi mérésekben is egyre gyakrabban alkalmazott generációját adják az objektív mérést is lehetővé tevő valószínűségi tesztelméletek. Az objektív mérés megvalósításának lehetőségével, módszereivel régóta foglalkoznak a társadalomtudósok, mivel az lehetőséget biztosítana olyan egyetemes, mindenki által elfogadott skálák megalkotására, mint a természettudományokban például a hőmérsékleti skálák vagy éppen az idő beosztása (Molnár, 2005).

Hazánkban jelentős múlttal rendelkeznek a klasszikus tesztelméleti módszerekkel történő elemzések, azonban ezek nem alkalmasak az objektív mérés, az objektív skálák megalkotására, továbbá módszereik segítségével bizonyos kérdéseket nem tudunk megválaszolni. (Az objektív mérés megvalósításának lehetőségéről lásd: Molnár, 2005, 2006, a valószínűségi és a klasszikus tesztelmélet összevetéséhez: Molnár és Józsa, 2006, konkrét elemzésekhez: Molnár, 2003, 2004.)

A valószínűségi tesztelmélet egyik, talán legfontosabb és legismertebb modellje, a Rasch-modell csak dichotóm adatok elemzésére alkalmas, ezért a kutatók továbbfejlesztették a modellt, hogy más, nem dichotóm adatokból álló adatbázisok elemzését is lehetővé tegyék. A Rasch-modell főbb tulajdonságait, matematikai hátterét egy korábbi tanulmányban foglaltuk össze (Molnár, 2006). E tanulmány célja olyan valószínűségi modellek és valószínűségi függvényeken nyugvó elemzések bemutatása, amelyek alapját rangskálán lévő adatok képezik.

A tanulmány elején a konzisztencia végett röviden bemutatjuk a Rasch-modell matematikai formalizálását, majd áttekintjük a parciális kredit modell és a rangskálás modell tulajdonságait. Bemutatjuk e modellek matematikai hátterét, levezetését a Rasch-modellből, továbbá a karakterisztikus görbék, nehézségi indexek értelmezési módját, tulajdonságait az egyes modellekben. Kitérünk e modellek megkülönböztető tulajdonságaira is.

A parciális kredit modell – felépítése, levezetése következtében, mint korábban utaltunk rá – egy speciális esete a Rasch-modell. Ennek következtében azok a szoftverek, amelyek kezelni tudják a parciális kredit modellt, Rasch-modellel történő elemzéseket is el tudnak végezni, sőt egy modellben variálni is tudják a dichotóm és nem dichotóm itemek elemzését. A tanulmányban bemutatott elemzések, ábrák a ConQuest (Wu, Adams és Wilson, 1998) szoftverrel készültek.

A Rasch-modell

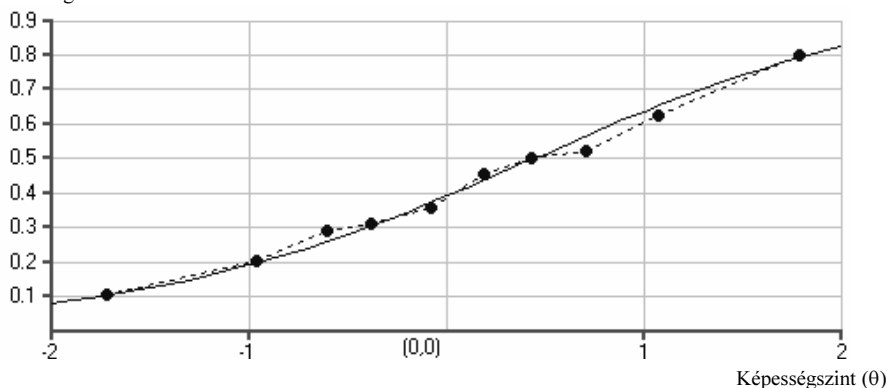
Egy itemre adott legegyszerűbb válaszmintázat az, amikor két válaszlehetőség közül választunk: igen-nem, jó-rossz, minden-semmi. Az ily módon kódolt itemekre tekinthe-

tünk akár mint „egy-lépcsős” ítemekre, ahol, ha valaki megtette azt az egy lépcsőt, akkor 1 pontot kap, ha nem, akkor 0-t. Rasch az 1950-es években (1960) ezen típusú adatok elemzésére dolgozott ki egy modellt, amit azóta gyakran Rasch-modellnek neveznek. A modell elterjedt az egész világon, számos nemzetközi mérésben, illetve ítebankok felépítése során alkalmazzák (*Write és Masters, 1982*).

A Rasch-modell dichotomításánál fogva a részben jó válaszok elemzésére nem ad lehetőséget. A pedagógiai kutatásokban legtöbbször mégis elegendő, mivel azok skálái leggyakrabban dichotóm skálák. (A Rasch-modell közelítő eljárásairól és az item illeszkedésről lásd: *Write és Stone, 1979; Griffin, 1999*.) A további modellek könnyebb megértése és a Rasch-moddal való kapcsolatuk bemutatása miatt felvázoljuk a Rasch-modell egyenletét (levezetését és tulajdonságait lásd: *Molnár, 2006; Horváth, 1997*).

A tanulmányban bemutatott karakterisztikus és valószínűségi görbék, valamint különböző elemzések mind egy-egy szimulált adatbázis egy-egy ítemének tulajdonságát jellemzik. A szimulált adatbázisokban közös, hogy a diákok száma minden esetben (n) 2000, az ítemek száma pedig 10. Különbözőség csak az ítemek lehetséges pontozásában, illetve a modellek felépítésében van. Jelen esetben az ítemek kódolásánál szóba jöhető pontszám a 0 és 1 volt. Az 1. ábra egy fenti feltételeknek megfelelő adatbázis 5. ítemére adott helyes válasz valószínűségét mutatja a képességszint és az ítem nehézsége függvényében. Ebből adódóan jelen esetben az $i=5$.

Valószínűség



1. ábra. Dichotóm ítem esetén a jó válasz valószínűségi görbéje a képességszint függvényében

A helyes válasz valószínűségét az (1) egyenlet írja le:

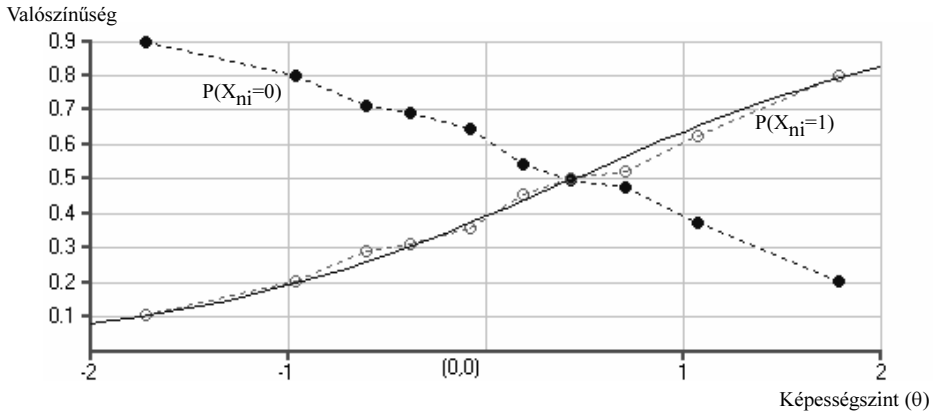
$$P(X_{ni} = 1) = \frac{\exp(\theta_n - \delta_{i1})}{1 + \exp(\theta_n - \delta_{i1})} \quad (1)$$

ahol $P(X_{ni}=1)$ az n-edik személy i-edik ítemre adott helyes válaszána valószínűsége, θ_n a személy képességparamétere, δ_{i1} az i-edik ítem jó válaszána (első lépésének) nehézségi paramétere. Hasonlóképpen írható le a helytelen válasz valószínűségét meghatározó egyenlet, majd a két egyenletet egy közös modellben felírva a (2) egyenletet kapjuk, ami a Rasch-modell matematikai formalizálása.

$$P(X_{ni} = x) = \frac{\exp \sum_{j=0}^x (\theta_n - \delta_{ij})}{\sum_{h=0}^{m_i} \exp \sum_{j=0}^h (\theta_n - \delta_{ij})} \quad (2)$$

ahol $x=0$ vagy 1 és $P(X_{ni}=x)$ az n -edik személy i -edik itemre adott helyes vagy helytelen (x értékétől függően) válaszána valószínűsége.

A modellben az item nehézségi indexének meghatározásához a fent említett szimulált adatbázis – már az 1. ábrán is elemzett – 5. itemének helyes, illetve helytelen megoldásának valószínűségi görbéit mutatjuk be a képességszint függvényében (2. ábra). Definíció szerint a két görbe metszéspontja (δ_{i1}) adja az item nehézségi indexét (jelen esetben $\delta=0,44$), ami azt a pontot jelenti, ahol a helyes és helytelen válasz valószínűsége 50–50 százalék (vö az 1. és 2. ábrát.), azaz $P(X_{ni}=0)+P(X_{ni}=1)=1$.



2. ábra. Dichotóm item esetén a jó és rossz válasz valószínűségi görbéi a képességszint függvényében

A 2. ábráról leolvasható, hogy a képességszint növekedésével egyre csökken annak valószínűsége, hogy a személy 0 pontot ér el az itemen, illetve egyre nő annak valószínűsége, hogy 1 pontot ér el. A δ_{i1} képességszintig, ami definíció szerint az item nehézségi indexét is meghatározza, a helytelen válasz valószínűségét adó görbe felette van a helyes válasz valószínűségét jellemző görbének, majd fordítva, δ_{i1} képességszint felett nagyobb annak a valószínűsége, hogy a személy jó választ ad az itemre.

A tanulmány további részében ismertetett modellekben felhasználjuk a Rasch-modell nehézségi indexre, képességszintekre, válaszmintázatokra, illetve a válaszok lépcsőzetes kezelésére vonatkozó meghatározásait.

A parciális kredit modell

Ahogy korábban utaltunk rá, a társadalomtudományi kutatások során nem mindig elegendő, ha adataink dichotóm skálán helyezkednek el, gyakran szükség van a több fokozatú értékelésre is. A megalkotás sorrendjét szem előtt tartva, Likert-skálán lévő adatok elemzésére alkalmas Andrich (1978) rangskálás modellje. A modell hátránya, hogy csak azon adatbázisok esetén alkalmazható, ahol minden egyes itemnek megegyezik a skálaszerkezete (Bond és Fox, 2001). Ez elég nagy hátrányt és korlátot jelentett az elemzésekben, ezért továbbfejlesztették a modellt. A parciális kredit modell (Masters, 1982) használata már nem követeli meg az azonos skálaszerkezetet. Alkalmazható például olyan adatok elemzése során, ahol az értékelés egy skálán (például 05-ös skálán) történik, vagy olyan itemeknél, ahol a válaszok egy része jobb, mint a többi (például tévképzés-kutatásokban), vagy olyan többlépcsős itemek esetében (például problémamegoldásnál), ahol a diáknak több, egymástól lehetőleg független lépést kell megoldania a feladat megoldása során (például egy matematikafeladat esetén, ahol ki kell számolni, hogy mennyi $\sqrt{8/0,2-4}$). Matematikailag a modellek közötti eltérés azok parametrizációjában van. A könnyebb

megértés kedvéért először a parciális kredit modell levezetését, majd abból a rangskálás modell levezetését mutatjuk be annak ellenére, hogy előbb a rangskálás modellt alkották meg, amelyet csak később követett a parciális kredit modell.

A parciális kredit modell levezetése a Rasch-modellből

A parciális kredit modell egyedül abban különbözik a Rasch-modelltől, hogy nem két, hanem több válaszlehetőséggel rendelkező itemek elemzésére is alkalmas. Ebből adódóan az első nem az egyedüli lépés, azaz $P(X_{ni}=0)+P(X_{ni}=1)<1$. Ahhoz, hogy valaki eljusson a második lépésig, meg kell tennie az első lépést. Ebből a gondolatból, azaz az egymás melletti kategóriákba tartozás valószínűségének meghatározásából indult ki Masters (1982) a modell felállítása során, majd az egyes kategóriákba tartozás valószínűségét leíró egyenleteket közös modellben foglalta össze (*Write és Masters*, 1982).

Konkrét példán szemléltetve, maradva a $\sqrt{8/0,2}-4 = ?$ feladatnál, a lépésre bontás a következőket jelenti:

Ha nem tette meg az első lépést:	0 pont
$8/0,2=40$ (első lépés)	1 pont
$40-4=36$ (második lépés)	2 pont
$\sqrt{36} = 6$ (harmadik lépés)	3 pont

A feladat megoldásának lépésekre bontásából adódik, hogy a második lépést nem lehet anélkül elvégezni, hogy az első lépést ne végezte volna jól el a személy, illetve a harmadik lépést sem lehet jól elvégezni az első és a második lépés helyes elvégzése nélkül. Továbbá az is leolvasható, hogy nem minden esetben igaz, hogy a későbbi lépés nehezebb, mint az azt megelőző (erre a kérdéskörre és az ebből adódó problémákra a későbbiekben még visszatérünk).

A parciális kredit modell matematikai levezetésében vegyünk egy olyan parciális kredit itemet, ahol csak két lépést (0, 1, 2 pontot lehet elérni) kell megtenni a teljes megoldásig. Elsőként megnézzük annak valószínűségét, hogy 0 vagy 1 pontot ér el a diák ezen a virtuális 3 kategóriás itemen (lásd az [1] és [2] egyenletet). Az (1) és (2) egyenlet a Rasch-modell formáját követi.

$$p_{0/0,1} = P(x = 0 / X = 0 \text{ vagy } X = 1) = \frac{P(X = 0)}{P(X = 0) + P(X = 1)} = \frac{1}{1 + \exp(\theta - \delta_1)} \quad (1)$$

$$p_{1/0,1} = P(x = 1 / X = 0 \text{ vagy } X = 1) = \frac{P(X = 1)}{P(X = 0) + P(X = 1)} = \frac{\exp(\theta - \delta_1)}{1 + \exp(\theta - \delta_1)} \quad (2)$$

ahol θ a személy képességparamétere a vizsgált látens változó képességskáláján, δ_1 az item megoldása első lépésének nehézségi paramétere ugyanazon skálán.

Hasonlóan annak valószínűsége, hogy a diák 1 vagy 2 pontot ér el az itemen, a következőképpen írható le (a [3] és [4] egyenlet is a Rasch-modell formáját követi):

$$p_{1/1,2} = P(x = 1 / X = 1 \text{ vagy } X = 2) = \frac{P(X = 1)}{P(X = 1) + P(X = 2)} = \frac{1}{1 + \exp(\theta - \delta_2)} \quad (3)$$

$$p_{2/1,2} = P(x = 2 / X = 1 \text{ vagy } X = 2) = \frac{P(X = 2)}{P(X = 1) + P(X = 2)} = \frac{\exp(\theta - \delta_2)}{1 + \exp(\theta - \delta_2)} \quad (4)$$

ahol θ a személy képességsparamétere a vizsgált látens változó képességskáláján, δ_2 az item megoldása során az első lépés után a második lépés megtételének nehézségi paramétere ugyanazon skálán.

A δ_2 paraméter azonban nem mond semmit arról, hogy a személy milyen valószínűség mellett ér el 1 pontot, milyen valószínűség mellett teszi meg jól először a megoldáshoz vezető út első lépését, holott ha nem teszi meg az első lépést, nem teheti meg a másodikat sem. Ebből adódóan δ_2 paraméter függ az első lépés megtételének nehézségétől, vagyis nem független nehézségi paraméter, mintha a két lépés egy-egy független item lenne.

Ha nem párba állítva modellezzük az egyes kategóriaértékek valószínűségét, hanem a három értékkategóriát együtt kezelve, akkor a következő egyenletrendszerrel írható le a modell:

$$p_0 = P(x = 0) = \frac{1}{1 + \exp(\theta - \delta_1) + \exp(2\theta - (\delta_1 + \delta_2))} \quad (5)$$

$$p_1 = P(x = 1) = \frac{\exp(\theta - \delta_1)}{1 + \exp(\theta - \delta_1) + \exp(2\theta - (\delta_1 + \delta_2))} \quad (6)$$

$$p_2 = P(x = 2) = \frac{\exp(2\theta - (\delta_1 + \delta_2))}{1 + \exp(\theta - \delta_1) + \exp(2\theta - (\delta_1 + \delta_2))} \quad (7)$$

Általánosítva, ha i item egy nem dichotóm, 0, 1, 2, ... m_i válasz kategóriájú item, akkor annak valószínűsége, hogy n személy az i itemen x pontot ér el, megadja a parciális kredit modell általános egyenletét (lásd a [8] egyenletet):

$$P(X_{ni} = x) = \frac{\exp \sum_{j=0}^x (\theta_n - \delta_{ij})}{\sum_{h=0}^{m_i} \exp \sum_{j=0}^h (\theta_n - \delta_{ij})} \quad x=0,1,\dots,m_i \quad (8)$$

ahol $\delta_{i0} \equiv 0$ úgy, hogy $\sum_{j=0}^0 (\theta_n - \delta_{ij}) = 0$ és $\exp \sum_{j=0}^0 (\theta_n - \delta_{ij}) = 1$ (Write és Stone, 1982).

A (8) egyenletben a számláló csak a megtett x lépés nehézségi indexét tartalmazza, míg a nevező az összes lehetséges (m_i+1) számlálót magába foglalja.

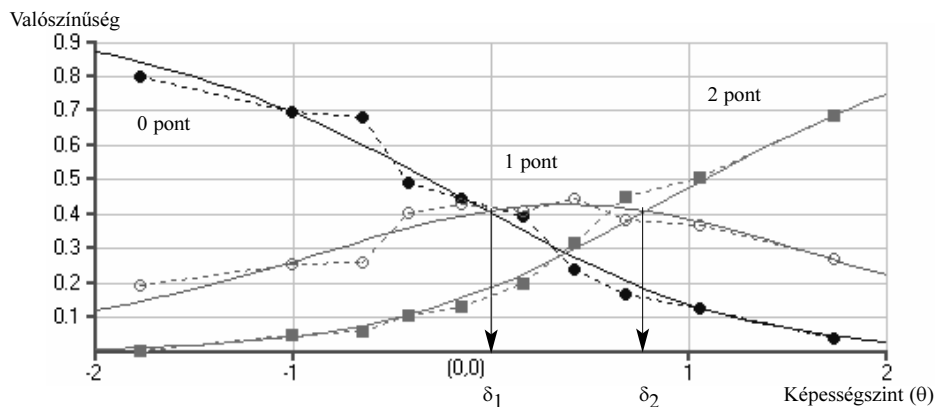
Egy egy lépcsős item esetén ($m=1$) elegendő 1 karakterisztikus görbe annak leírásához, hogy a személy milyen képességszint mellett ér el nagyobb valószínűség mellett 1, mint 0 pontot (lásd 1. ábra). Egy két lépcsős ($m=2$) item esetén már két karakterisztikus görbére van szükség ennek jellemzésére. Az első logisztikus görbe arról ad információt, hogy mi a valószínűsége annak, hogy a személy inkább 1, mint 0 pontot ér el az itemen, a második görbe pedig azt jellemzi, hogy milyen valószínűség mellett ér el a személy inkább 2, mint 1 pontot az itemen. Ezek a karakterisztikus görbék a képességskala különböző részén elhelyezkedő azonos meredekségű egyszerű logisztikus görbék (Write és Masters, 1982).

A modell alkalmazásának nem feltétele, hogy a második lépés minden esetben nehezebb legyen, mint az első lépés, viszont a második lépést csak az első valamilyen megtétele után lehet megtenni. Ha a második lépés könnyebb, mint az első, akkor a két görbe fordítva helyezkedik el a képességskálán. A következőkben ezt a problémakört járjuk körül a modell értékkategóriáinak jellemzésében.

A parciális kredit modell értékkategóriáinak tulajdonsága

Az itemre adott válaszok pontozását a feladatlapok, tesztek kódolása során úgy kell kialakítani, hogy a pontszám növekedése párhuzamos legyen a vizsgált látens képesség fejlettségi szintjével, azaz minél magasabb az adott pontszám, annál magasabb kompetenciaszintet tükrözzön: a magasabb képességszintű diákok magasabb értékkategóriába, az alacsonyabbak alacsonyabb értékkategóriába tartozzanak. A két legalacsonyabb kategória a 0 és az 1. A magasabb képességszintűek nagyobb valószínűséggel tartozzanak az 1-es, mint a 0-s kategóriába. Hasonlóképpen az 1 és 2 kategória esetén a magasabb képességszintű diák nagyobb valószínűséggel kapjon 2, mint 1 pontot. Ebből következőleg, ha az összes értékkategóriára általánosítunk, a magasabb képességszintű diák nagyobb valószínűséggel kapjon több pontot; más oldalról megközelítve: több pont elérését várjuk el tőle, mint az alacsonyabb képességszintű diáktól.

A parciális kredit modell itemkarakterisztikus görbéi azt mutatják meg, hogy a különböző képességszintek mellett mi a valószínűsége annak, hogy a diák az adott értékkategóriát kapja a feladat megoldása során. A 3. ábra egy szimulált adatbázis ($n=2000$, itemek száma=10, válaszkategóriák száma=3 [0, 1, 2]) 6. itemének itemkarakterisztikus görbéit mutatja. Az ábráról leolvasható, hogy a képességszint növekedésével növekedik annak valószínűsége is, hogy a diák magasabb kategóriában van, magasabb pontszámot ér el. A görbék közül legfelül először a 0 kategóriába tartozás valószínűségét mutató görbe van, majd az 1 kategóriába tartozás valószínűségét mutató, végül a képességszint további növekedésével a 2 kategóriába tartozás valószínűségét mutató görbe húzódik.



3. ábra. Egy három válaszkategóriás item itemkarakterisztikus görbéi

A δ_k grafikus interpretációját ugyancsak a 3. ábra mutatja. Az ábrán azok a képességszint-értékek a δ_k értékei, ahol az egyes karakterisztikus görbék metszik egymást. Ez azt jelenti, hogy δ_k az a pont, ahol annak valószínűsége, hogy a diák a $k-1$ vagy a k kategóriában van, azonos. Ez a valószínűség kevesebb, mint 0,5, mivel annak valószínűsége, hogy a diák a $k-1$ és k kategóriákon kívüli kategóriában van, feltételezésünk szerint nem 0. Ez a matematikai tény adja a δ_k jelentését. (Matematikai szemszögből a δ értékek az [1–4] egyenletekből levezethetők.)

A két δ paraméter három részre osztja a képességskálát:

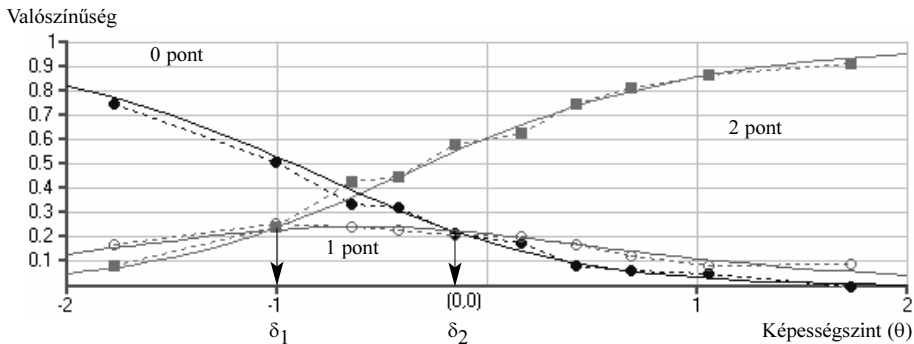
(1): $]-\infty, \delta_1[$, amilyen képességszintű diákok legnagyobb valószínűséggel a 0 kategóriában vannak, és alacsonyabb valószínűséggel teljesítenek az 1 vagy 2 kategóriában;

(2): $]\delta_1, \delta_2[$ képességszint-intervallumba eső diákok, akik legnagyobb valószínűséggel 1 pontot érnek el a feladaton, és kisebb a valószínűsége annak, hogy 0 vagy 2 pontot kapnak, valamint a

(3): $]\delta_2, \infty[$ képességszinttel rendelkező diákok, akik legnagyobb valószínűséggel 2 pontot érnek el a feladaton, és nem 0 vagy 1-et.

Ha δ_1 és δ_2 egymástól távol van a képességskálán, akkor számos képességszintű diák nagy valószínűség mellett ér el 1 pontot az itemen; ha közel vannak egymáshoz, akkor csökken ezen diákok köre: a képességszint függvényében jobban meghatározhatóvá válik az a diákcsoport, amelynek tagjai 1 pontot érnek el az itemen.

Előfordulnak azonban olyan itemek, ahol felcserélődnek a δ_k értékek, azaz nem rendezetten követik egymást a képességskálán. Ez akkor következik be, amikor – három kategória esetén – a középső görbe (jelen esetben az 1-es kategóriába tartozás valószínűségét mutató karakterisztikus görbe [lásd 4. ábra]) nagyon lapos, azaz nagyon kevés az olyan tanuló, aki ebbe a kategóriába sorolható. Ebben az esetben nehézkes az itemkarakterisztikus görbe interpretációja, mivel egyik képességszintre sem igaz, hogy legnagyobb valószínűséggel ebbe – jelen esetben az 1-es – a kategóriába tartoznak a diákok. A δ_1 képességszint, ahol azonos valószínűséggel van a diák a 0 és 1 kategóriában, magas érték, a δ_2 képességszint pedig, amilyen képességszintű diák azonos valószínűséggel kap 1 vagy 2 pontot, alacsonyabb érték, azaz $\delta_1 > \delta_2$. Mivel a δ értéke függ attól, hogy az egyes kategóriákban hány tanuló van, ebből fakadóan, mint korábban is utaltunk rá, a δ_k paraméter nem lehet egy független lépés nehézségének mutatója, hanem inkább az összes lépés nehézségétől függő mutató.



4. ábra. Egy három válasz kategóriás „rosszul viselkedő” item itemkarakterisztikus görbéi

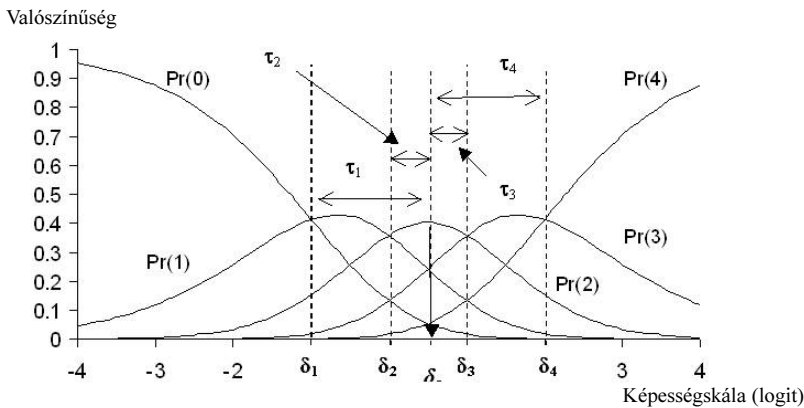
A δ_k paraméter-értékek felcserélődése gyakrabban megfigyelhető azon típusú feladatok esetén, ahol a problémát különböző lépésekben kell megoldani. A megoldáshoz vezető úton előfordulhat, hogy egy későbbi lépés könnyebb, mint egy azt megelőző. Például egy matematikai természetű probléma esetén az első lépés a formulává alakítás, a második a számítás elvégzése. Ebben az esetben a tanulók leggyakrabban a 0 vagy a 2 kategóriába tartoznak, mivel akik már helyesen lefordították a problémát a matematika nyelvezetére, vagyis formalizálták azt, ritkán követnek el számolási hibát. Másrésztől, ha holisztikusan alkalmazzuk a parciális kredit modellt, és például fogalmazások pontozása elemzésében használjuk, ritkán találkozunk ezzel a problémával.

A parciális kredit modell parametrizációjának lehetőségei

A képességskála $\delta_1, \delta_2, \dots, \delta_k$ paraméterértékei, mint korábban definiáltuk, az item egyes kategóriái karakterisztikus görbéinek metszéspontját jellemzik, azaz azokat a ké-

pességszinteket, ahol azonos annak valószínűsége, hogy a személy a k , vagy a $k+1$ -dik kategóriába sorolható. Ezzel szemben, ha valaki egyetlen paraméterértékkel, egy átlagos nehézségi indexszel szeretné jellemezni a parciális kredit item nehézségét, és elemzésében nincs szükség az egyes lépések nehézségi indexének leírására, akkor a δ_k paraméterértékek helyett használhatja azok átlagát (δ .) és a δ_k paraméterértékek δ . átlagtól való távolságát jellemző τ_k paraméterértékeket. A τ_k paraméterértékek önmagukban való interpretációja nehézkes, értelmezésük csak a δ . paraméter összefüggésében lehetséges. A τ_k paraméter egy lépésparaméter, ami megmutatja, hogy az egyes τ értékek milyen messze vannak az item átlagos nehézségi indexétől (δ .) Mivel értékük függ a karakterisztikus görbék elhelyezkedésétől, ezért ebben az esetben is találkozhatunk ugyanazzal a felcserélődés problémájával, ahogy a δ_k paraméterek esetében. Mind a δ . paraméter, mind a τ_k paraméterértékek itemről-itemre változhatnak.

Az új paraméterek más módon osztják intervallumokra a képességskála terjedelmét. Az 5. ábrán grafikusán ábrázoljuk az új (δ_i és τ_k) és a korábbi (δ_k) paraméterek tulajdonságait, azok különbözőségét.



5. ábra. Egy öt kategóriájú item karakterisztikus görbéi a két parametrizáció (δ_k és τ_k) esetén (Wu, 2006a alapján)

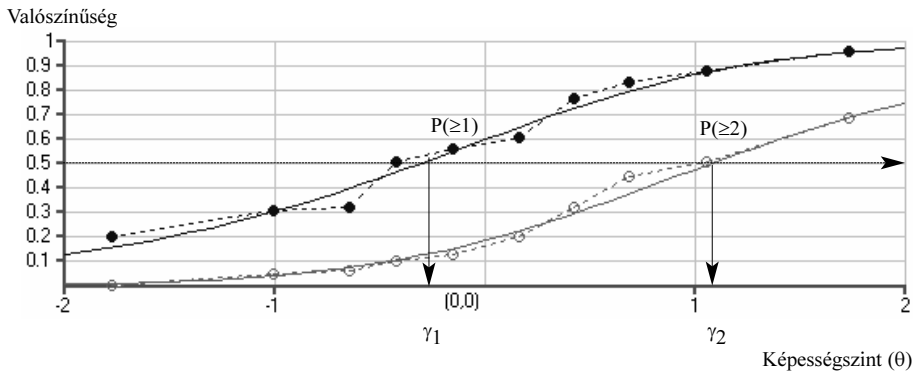
A thurstoni küszöb

A parciális kredit modellel kapcsolatban eddig tárgyalt paraméterértékek nem adnak információt arra vonatkozólag, milyen képességszint szükséges egy item adott kategóriájába való bekerüléshez. Parciális kredit itemek esetén például a két pont eléréséhez minden esetben magasabb képességszint szükséges, mint az 1 pont eléréséhez. E kumulatív teljesítmény leírására alkalmas mutató a thurstoni küszöb, amely definíció szerint azt mutatja meg, hogy milyen képességszint szükséges ahhoz, hogy valaki 50 százalék valószínűséggel elérjen egy adott pontszámot. Ebből adódóan a thurstoni küszöb az item nehézségi indexének értelmezésében játszik szerepet (Wu, 2006a).

Dichotóm item esetén az item nehézségi indexe definíció szerint az a képességszint, ahol a helyes megoldás valószínűsége 0,5. Ez a képességszint két részre – a 0 és 1 pontos részre – bontja a képességskálát. Ezt a definíciót általánosítjuk parciális kredit item esetére. A γ_1 az a képességszint, ahol az 1 pont elérésének nehézségi indexe van, a γ_2 az a képességszint, ahol a 2 pont elérésének nehézségi indexe van stb.

A 6. ábra egy 3 kategóriájú parciális kredit item esetén mutatja az adott kategóriába tartozás valószínűségi görbéjét a képességszint függvényében, grafikusán ábrázolva a thurstoni küszöb jelentését. A thurstoni küszöb értelmezhető úgy is, mint a képességskála olyan intervallumokra való felosztása, ahol értelmezhetővé válnak az itemen elért

pontszámok. A 6. ábra esetén ez azt jelenti, hogy a $\gamma_1 = -0,31$, a $\gamma_2 = 1,08$, azaz az 1 pont elérése közel átlagos képességszintet igényel, míg 2 pont eléréséhez már átlag feletti képességszint szükséges.



6. ábra. A thurstoni küszöb és a kumulatív valószínűségi görbék

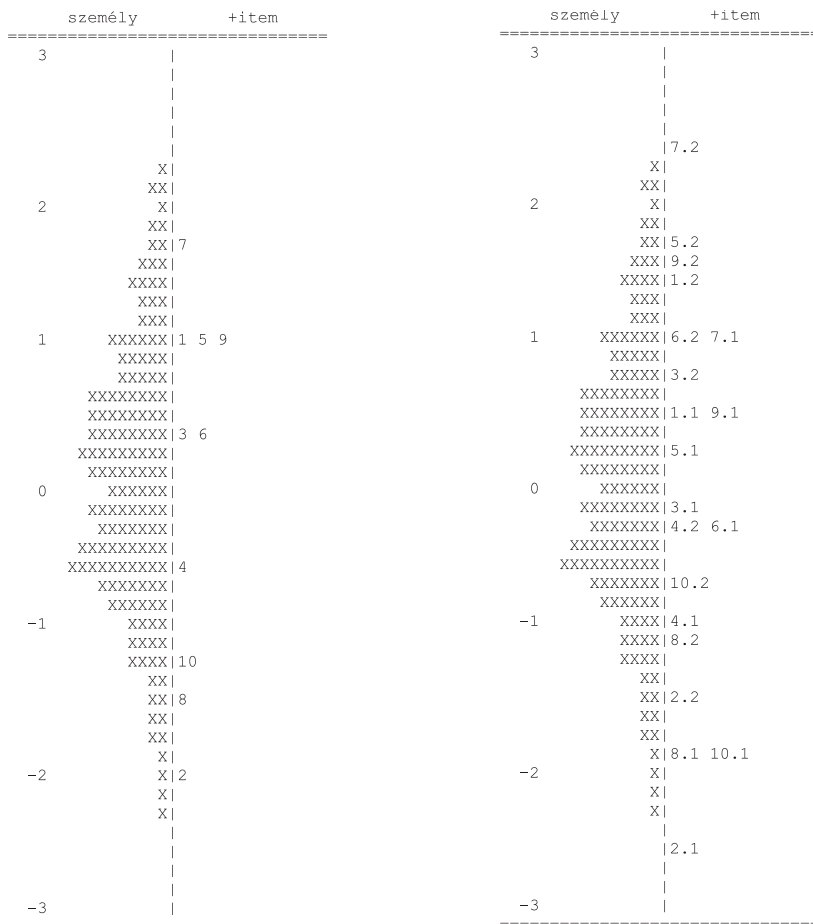
Az egyes feladatok thurstoni küszöbét és a diákok képességszint szerinti eloszlását közös képességskálán tudja ábrázolni a már említett ConQuest szoftver. Erre és az ábra interpretációjára adunk a következőkben egy példát: a 7. ábrán egy, a korábbi ábrákon is elemzett szimulált 2000 fős, 10 parciális kredit itemes adatbázis virtuális diákjainak és itemeinek személy-item térképeit ábrázoljuk, az adatokat dichotóm adatként és nem dichotóm adatként, a thurstoni küszöböt minden egyes item vonatkozásában megjelenítve.

Az ábra bal oldali felének személy-item térképén nem jelennek meg az item egyes lépéseinek nehézségi küszöbei, hanem egy Rasch-moddal történt elemzés eredményét mutatja, ahol az itemeket átlagos nehézségi paraméterük szerint rajzolta fel a program, míg az ábra jobb oldali személy-item térképén megjelenítette itemek szerinti bontásban a thurstoni küszöbértékeket is. Az x.y megjelenítés az x-edik item y-odik lépésének küszöbét jelenti, azt a küszöböt, ahol a tanuló 50 százalékos valószínűséggel éri el legalább a jelzett itemen belüli szintet. Mindkét személy-item térkép bal oldali része a diákok képességszint szerinti eloszlását mutatja, ami jelen esetben az egyező adatbázisok miatt azonos.

A tanulmány további részében ismertetjük a rangskálás modell alap gondolatát. Bár a modell, mint korábban utaltunk rá, a parciális kredit modell megalkotása előtt megvolt, de levezetése könnyebben érthető, ha azt a parciális kredit modell egyszerűsítésével, a modell korlátainak figyelmebe vételével tesszük.

A rangskálás modell

A rangskálás modellt mindazon itemek elemzésére tudjuk alkalmazni, amelyekre adott válaszok rangsorolt válaszalternatívák, például egy attitűd-teszt esetében, amikor négy alternatíva közül kell választanunk: nagyon nem szeretem, nem szeretem, szeretem, nagyon szeretem. Ez a négy alternatíva három lépcsőfok megtételét hordozza magában. Az első lépés, amikor dönteni kell, a nagyon nem szeretem és a nem szeretem között van, a második, amikor választani kell a nem szeretem és a szeretem között stb. Miután a modell alkalmazásának feltétele, mint korábban utaltunk rá, hogy a feladatlap összes itemére adott válasz azonos számú lépésből álljon, a válasz meghozatalakor megtett lépések nehézsége közel azonos minden item esetén, ami a későbbiekben fontos szerepet játszik. Jelen esetben az attitűd-teszt minden egyes kérdésére adott válasz adásakor maximum 3 azonos nehézségű „lépést” kell megtenni.



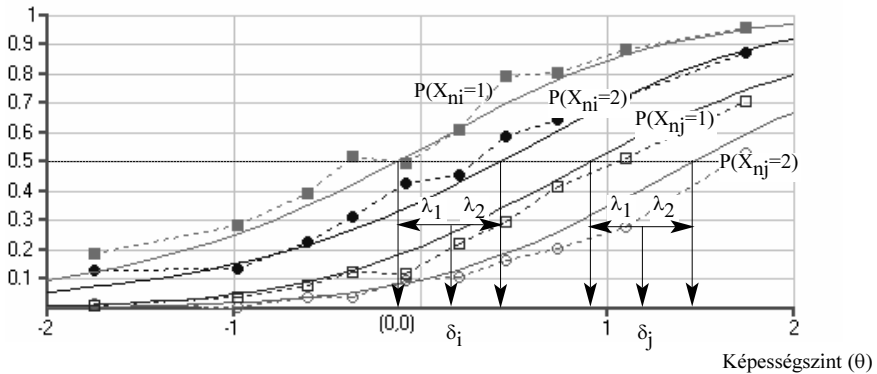
7. ábra. A thurstoni küszöb személy-ítem térképben való megjelenítése ugyanazon adatok dichotóm kezelésének fényében (mindkét ábrán minden egyes 'x' 13 tanulót reprezentál)

Ennek következtében a parciális kredit modell λ_k paramétere két komponensre, más tulajdonságokkal jellemezhető paraméterekre bontható szét (Write és Masters, 1982), amelyek bizonyos szempontból hasonlóak a parciális kredit modell alternatív paraméterezési lehetőségénél említett δ_i és τ_k értékekhez. Az azonos skálaszerkezetet és a skálákon belüli azonos lépésnehézséget figyelembe véve a következőképpen lehet parametrizálni a modellt:

$$\delta_{ik} = \delta_i + \lambda_k, \text{ ahol } \delta_i = \sum_{k=1}^{m_i} \delta_k / m$$

azaz a δ_k paraméterértékek átlaga (ez jelentésében megegyezik a korábbi δ_i paraméterrel), a λ_k küszöbérték pedig minden egyes ítem k-adik lépésének nehézsége az átlagos nehézség viszonylatában. Az ítemeket átfogó azonos lépésnehézség miatt minden egyes ítem esetén azonos a λ_1 , a $\lambda_2 \dots \lambda_k$; ez a parciális kredit modell τ_k paraméterértékeire nem igaz. Ha az átlagtól való eltérés irányát is figyelembe vesszük, és előjelesen kezeljük a paraméterek értékét, akkor egy ítem esetében a λ_k paraméterek átlaga 0, illetve $\lambda_1 = -\lambda_2$. A λ_k paraméter jelentését grafikusán a 8. ábra szemlélteti.

Valószínűség



8. ábra. Két rangkálás item kumulatív valószínűségi görbéje (jelen esetben $i=5$ és $j=6$, Write és Masters, 1982 alapján)

Az új parametrizációt behelyettesítve a parciális kredit modell egyenletébe leegyszerűsödik modellünk a rangkálás modellre:

$$P(X_{ni} = x) = \frac{\exp \sum_{j=0}^x (\theta_n - (\delta_i + \lambda_j))}{\sum_{h=0}^m \exp \sum_{j=0}^h (\theta_n - (\delta_i + \lambda_j))} \quad x=0,1,\dots,m \quad (9)$$

ahol $\lambda_0 \equiv 0$ úgy, hogy $\sum_{j=0}^0 (\theta_n - (\delta_i + \lambda_j)) = 1$.

A modellt a parciális kredit modell megalkotása óta ritkán használják, éppen fent említett korlátai miatt. Ha empirikusan össze szeretnénk hasonlítani a két modellt, akkor vegyünk egy adatbázist, aminek minden egyes itemének azonos a skálaszerkezete (például az eddig is elemzett $n=2000$, itemek száma=10, itemkategóriák száma=3 szimulált adatbázist), és elemezzük mindkét modellel. A következő eredményt kapjuk: a rangkálás modellben a közelített paraméterek száma 12, míg a parciális kredit modellben 21. Az iterációk száma mindkét modellel történt elemzés során 27, viszont a devianciában χ^2 -próbaival ($df=9$) ellenőrizve szignifikáns a különbség, a parciális kredit modell illeszkedésvizsgálata szignifikánsan jobb, mint a rangkálás modellé (az illeszkedésvizsgálatokról lásd *Wit*, 2006b).

Összességében az (1) egyenlet alapján három különböző valószínűségi modellt definiálhatunk aszerint, hogy hogyan definiáljuk a δ_{ix} -t:

- 1) ha $\delta_{ix} = \delta_i$, akkor a dichotóm esethez, azaz a Rasch-modell egyenletéhez jutunk,
- 2) ha $\delta_{ix} = \delta_{ix}$, akkor a parciális kredit modell egyenletét kapjuk,
- 3) ha $\delta_{ix} = \delta_i + \lambda_x$, akkor a rangkálás modellt írja le az egyenlet.

A δ_{ix} további parametrizálásával további valószínűségi modellekhez juthatunk (ezekről lásd *Write és Masters*, 1982). (1)

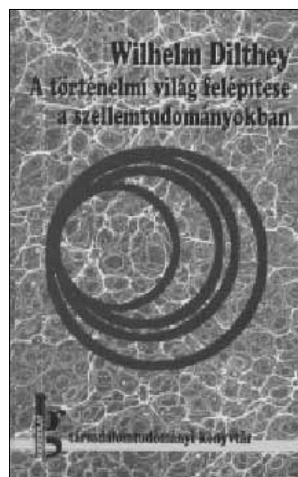
Jegyzet

(1) A tanulmány a T 046659PSP OTKA kutatási program, az Oktatásméleti Kutatócsoport és az SZTE MTA Képességkutató Csoport keretében készült. A

tanulmány írása idején a szerző Bolyai János Kutatási Ösztöndíjban részesült.

Irodalom

- Bond, T. – Fox, C. M. (2001): *Applying The Rasch Model. Fundamental Measurement in the Human Sciences*. Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey.
- Griffin, P. (1999): *Item Response Modelling: An introduction to the Rasch Model*. Assessment Research Centre Faculty of Education, The University of Melbourne.
- Horváth György (1997): *A modern teszmodellek alkalmazása*. Akadémiai Kiadó, Budapest.
- Molnár Gyöngyvér (2003): Az ismeretek alkalmazásának vizsgálata modern tesztelméleti eszközökkel. *Magyar Pedagógia*, 4. 423–446.
- Molnár Gyöngyvér (2004): Hátrányos helyzetű diákok problémamegoldó gondolkodásának fejlettsége. *Magyar Pedagógia*, 3. 319–338.
- Molnár Gyöngyvér (2005): Az objektív mérés megvalósításának lehetősége: a Rasch-modell. *Iskolakultúra*, 3. 71–80.
- Molnár Gyöngyvér (megjelenés alatt): A Rasch modell alkalmazása a társadalomtudományi kutatásokban. *Iskolakultúra*, megjelenés alatt.
- Molnár Gyöngyvér és Józsa Krisztián (megjelenés alatt): Az olvasási képesség értékelésének tesztelméleti megközelítései. In Józsa Krisztián (szerk.): *Az olvasási képesség fejlődése és fejlesztése*. Dinasztia Tankönyvkiadó, Budapest.
- Rasch, G. (1960): *Probabilistic models for some intelligence and attainment tests*. Danish Institute for Educational Research, Copenhagen.
- Write, B. D. – Masters, G. N. (1982): *Rating Scale Analysis*. MESA press, Chicago.
- Write, B. D. – Stone, M. H. (1979): *Best Test Design*. MESA press, Chicago.
- Wu, M. (2006a): *PISA Training Workshop: Application of Item Response Theory (IRT) to PISA (ConQuest)*. Hong Kong PISA Centre, Hong Kong.
- Wu, M. (2006b): *How Well Do the Data Fit the Model? Kézirat*.
- Wu, M. – Adams, R. J. – Wilson, M. R. (1998): *ACER ConQuest. Generalised Item Response Modelling Software*. ACER Press, Australia.
- Masters, G.N. (1982): A Rasch model for partial credit scoring. *Psychometrika*, 149–174.
- Andrich, D. A. (1978): A rating formulation for ordered response categories. *Psychometrika*, 561–573.



A Gondolat Kiadó könyveiből