

SZOFTVER

DÁVID GERGELY ANDRÁS

Facets, a többdimenziós mérés eszköze

Kivonat:

Az itt bemutatott Facets szoftver (Linacre, 2024) lehetőséget teremt az idegennyelv-tudás komplex, több dimenziós mérésére és – tágabb értelemben – a mérés-értékeléshez közvetett módon kapcsolódó, vagy akár nem kapcsolódó projektek eszközeinek, hatástényezőinek vizsgálatára, az érvényesítés (validálás) céljával. A Facets megvalósítja a kommunikatív nyelvtudás összetett, többdimenziós mérésének feladatát, ahogy azt Morrow, Canale és Swain és mások elképzelték: a bevett tényezők, dimenziók mint a feladat nehézsége és az értékelők szigora mellett további dimenziók azonosítására és értékelésére is van lehetőség. Viszonylag kis létszámú adatsomagok feldolgozására is alkalmas, adott eljárás követése esetén, ami szélesíti a lehetséges felhasználók körét az oktatási rendszeren belül és kívül.

Kulcsszavak: mérés-értékelés, valószínűségi mérési modell, több dimenziós Rasch-modell, kommunikatív idegennyelv-tudás mérés, Facets

Bevezetés és kontextualizáció

Akarjuk-e kompenzálni a vizsgafeladatsorok nehézségének hullámmzását? Akarjuk-e kompenzálni az értékelők, vizsgáztatók egyes csoportjainak eltérő szigorát? Munkánk, kutatásaink során szembesülünk-e avval, hogy szinte minden oktatási, társadalmi, lélektani jelenség, kérdés és probléma sokdimenziós lévén, fő és nemkívánatos melléktényezők együttes eredője? Szembesültünk-e azzal – gyanítjuk-e –, hogy még az objektív mérési eszközök is több tényező együttes hatását mutatják? Avagy kérdőívek érvényességét vizsgálva, avval a problémával szembesülünk-e, hogy az egyes változatait kevés válaszadó töltötte ki? Avagy újra kell-e kezdeni a kérdőívhez az adatgyűjtést, miután kiderült, hogy meg kell változtatni néhány kérdést? Ezekben az esetekben megoldás lehet a Facets szoftver. Legújabb elérhető változata a 4.1.3 (Linacre 2024; <https://www.winsteps.com/facets.htm>).

A kommunikatív nyelvoktatás elméleti alapvetése az 1980-as évek elejére elkészült, amint azt néhány kulcsfontosságú forrás jelzi is (Brumfit & Johnson, 1979; Johnson & Morrow, 1981; Littlewood, 1981). Az elmélet kommunikatív idegen nyelvi mérés érintő részét Morrow (1979) készítette el. Morrow hadat üzent a hagyományosnak tekinthető és a strukturalizmustól örökölt, ún. objektív, diszkrét elemekből álló nyelvi vizsgáknak vagy nyelvvizsgáknak. Morrow feltételezte azt is, hogy ezek a mérési eszközök idővel el fognak tűnni. Mindezzel nagyjából egyidőben Canale és Swain

(1980) megalkotta a nyelvtudásnak azt a többdimenziós fogalmát (konstruktumát), amely szerint a kommunikatív idegennyelv-tudás mérhetővé vált.

Ha az idegen nyelvi kommunikáció és annak oktatása is figyelembe veszi a nyelvhasználat többdimenziós jellegét (szociolingvisztikai, pszicholingvisztikai stb. tényezők elismerése), az idegennyelv-tudás mérése miért is ne lenne szintén többdimenziós mérési feladat? Szinte minden, ami a nyelvoktatásban vagy a valós nyelvhasználat körülményei közepette körbevesz bennünket, többdimenziós jelenség! A megfelelő válaszra, a Facets szoftver megjelenésére majd egy évtizedet kellett várni, amíg a matematikus Linacre azt 1989-re megtervezte, a programot megírta, megteremtve ezzel a többdimenziós Rasch-módszer, a Many-facet Rasch Modeling (MFRM) kategóriáját.

A Rasch model a valószínűségi tesztmodellek egyik, egyparaméteresnek nevezett ága. Felvetődik, hogyan lehet a Facets a több dimenziós mérés eszköze, ha az ún. egyparaméteres megoldások közé tartozik. A legjobb válasz az lehet, hogy az adatelemzési módszerek akkor és úgy neveztek el, amikor és ahol kifejlesztették őket. Az eredeti Rasch módszer európai eredetű és csakis egyparaméteres volt (Rasch, 1960), míg az egy-, két- és háromparaméteres, felelet-válasz eleméletként (Item Response Theory, IRT) is ismert modellek később születtek Észak-Amerikában. Mindezeket követte Linacre több dimenziós Rasch-modellje (1989), amely akár 60 tényezőt is figyelembe vehet, de a Rasch-modell alapján, vagyis az itemek diszkriminációja paraméterének beszámítása nélkül.

A Facets szoftver bemutatása

A Facets tehát a valószínűségi tesztmodell alapján működik, ami azt jelenti, hogy ugyanazokból a válaszadatokból indul ki, mint bármely más, a klasszikus tesztelmélet alapján működő elemző szoftver, azonban a Facets a vizsgázó képességértékeit a betáplált változók (pl. itemek, értékelők, skálapontok stb.) figyelembevételével számolja ki, logaritmikus transzformációk útján.

Nemcsak elemző szoftverről van szó, hanem a vizsga (teszt) eredmények kiszámítására is fel lehet használni; ennek során a betáplált változók arányában korrigálja a vizsgázó eredményét. Szélesen elfogadott nézet szerint, ahogy minden iskolás is tudja, a feladat, item, feladatsor stb. nehézsége hat a pontértékekre, jegyekre, csökkentve őket. A Facets megalósítja ezt a nézetet, mert a vizsgázók átlagosnál alacsonyabb megfigyelt pontértékei esetében a feladat átlagosnál nehezebb voltára következtet és ennek arányában emel a képességszinten, míg magasabb pontértékek, vagyis könnyebb feladatok stb. esetében csökkenti azt. Hasonlóképp, az átlagosnál alacsonyabb pontokat adó, vagyis elnézőbb vizsgáztató esetében csökkenti a vizsgázók képességértékét, ill. szigorú vizsgáztató esetében növeli azt. Más változók hatásának iránya nem ilyen egyértelmű, azonban a Facets felhasználható mérési-kutatási projekteknél az érvényesítés (validálás) eszközeként is, illetve szélesebb értelemben, nem kizárólag a mérés terén, pl. kutatási projekteknél fontos tényezők (változók) meglétének, irányának és hatása mértékének feltárására és igazolására, továbbá a nem kívánatos hatások kiküszöbölésére.

A többváltozós Rasch-model képlete a következő:

1. ábra

A többváltozós Rasch-modell

$$\log (P_{nmijk} / P_{nmj}(k-1)) = B_n - D_i - A_m - C_j - F_k$$

ahol

B_n = n személy (vizsgáló) képességértéke,

D_i = i értékelési skála nehézsége,

A_m = m feladat nehézsége,

C_j = j vizsgáztató szigora,

F_k = k kategóriában a k-1 kategóriához viszonyított megfigyelhetőség akadály.

A fentiek úgy is megfogalmazhatók, hogy a Facets kiindulásként kiszámítja az adatsomagban meglévő összes varianciát (szórásnégyzetet), majd ebből kivonja a mérés befolyásoló és betáplált tényezőkhez kapcsolódó szórásnégyzeteket (pl. a feladatokhoz, értékelőkhöz, skálákhoz kapcsolódó szóráskomponensek), ily módon egyben szétosztva és csökkentve a mérési hibát is. A kivonások után fennmaradó szórás a nemkívánatos hatásoktól megtisztított mérendő, jellemzően a vizsgázók tudása, melyet logaritmikus transzformáció után kapunk meg – ennek alap mérési egysége a logit. A szóráskomponensek meghatározását azok skálázása követi, amikor is minden betáplált tényező ugyanarra a logit skálára kerül, innen az összevethetőség.

A Facets alkalmazása során, elemzési céllal, lehetőség van arra, hogy a feladatoknak, skáláknak, vizsgáztatóknak – és természetesen vizsgázóknak – a valószínűségi modellhez való illeszkedését értékeljük, illetve a nem illeszkedő itemek, feladatok stb. azonosítását, semlegesítését elvégezzük és adott konkrét esetben a mérésből ki is zárjuk.

Elérhetőség és a felhasználói felület használhatósága

A Facets a <https://www.winsteps.com>-on érhető el. Nem nagyon drága szoftverről van szó: 149 USD jelenleg. A felhasználói engedély (licence) lehet egy felhasználóra szóló vagy több, egy helyen (laboratóriumban, gépteremben stb.) dolgozó kutató együtt is használhatja, ill. az engedély lehet intézményi is. A szoftverhez felhasználói kézikönyv is jár (Linacre, 2017), letölthető pdf-ként ugyaninnen. Ingyenes a Facetsnek egy „könnyített”, demo változata is (Minifac), amely kevesebb adattal éppúgy fut, mint a valódi, „éles” változat. A winsteps.com hasznos Rasch irodalmat is ajánl (nem csak angol nyelven), egyes kiadványok, mint Linacre eredeti elméleti alapvetése (Linacre, 1989, 1993, 1994) le is tölthető. A szoftver Windows kompatibilis, de a korábbi, DOS alapú változat szintén ingyen használható.

A felhasználói felület előkészítéséhez nincs szükség további programra vagy szoftverre, mert kiindulásnak használható bármely szövegszerkesztő (pl. Word). A kiindulási pont mindig a lehető legegyszerűbb: szövegfájlra (.txt) van szükség, melyek jellemzően specifikációs elemeket és adatokat tartalmaznak, innen e szövegfájlok eredeti kiterjesztése (.sd). Javasolt táblázatkezelő (pl. Excel) alkalmazása az adattábla előkészítésére, amit .csv (comma separated values) fájlként lehet a .txt specifikációs és adatfájlba beszerkeszteni.

A .txt (.sd) fájl három fő egységre tagolódik:

1. alapbeállítások (cím, output fájlok, iterációs körök stb),
2. facetek (dimenziók) (labels=)
3. adatok (data=).

A letöltött szoftvercsomagban példák találhatók; Linacre javaslata szerint ezek valamelyikét, a legalkalmasabbat alakítsuk át saját céljainkra, így kevésbé valószínű, hogy valami kifejejtődik.

Az újonnan installált szoftverhez néhány beállítás azonban szükséges, ugyanis a Windows beállításai Magyarországon kapható változatban néhány tekintetben eltérnek. Egyrészt tegyük láthatóvá a fájlok kiterjesztéseit, mert azokban a felhasználók hasznos információt találnak. Másrészt a törtek jó megjelenítése érdekében a tizedes vesszőt állítsuk tizedespontra. Továbbá a cellahatárolót változtassuk pontosvesszőről vesszőre, mert a magyarországi Windows változatokban a pontosvessző a beállított érték, minek következményeképp a .csv adattáblában is pontosvessző szerepelne. A Facets adatokat, információkat mindig vesszővel választ el, viszont a kommenteket csakis pontosvessző mögé téve lehet a szoftver számára láthatatlanná tenni.

Felhasználási lehetőségek mérésben – kipróbált, gyakorlati alkalmazások

Mint fentebb leírtuk, a Facets legjellemzőbb felhasználási lehetősége a teljesítmények mérése, de e lehetőségek egy részét természetesen nem csak a Facets kínálja. Más valószínűségi szoftverekhez hasonlóan klasszikus valószínűségi teszt funkciónak tekinthető a nem illeszkedő mérési elemek mint az itemek, feladatok, értékelők, skálapontok stb. azonosítása. Nem illeszkedő elemek lehetnek az itemek, skálák, vizsgáztatók és értékelők stb., ha a meg nem magyarázott szórás a +/- 20%-ot eléri vagy meghaladja. Vonzó jellemzője a szoftvernek, hogy mivel a valóság sokféle, a mérés tényezőinek szóródása nagy és hibajelenség is bőven akad, a szoftver nem „vár el” teljes mértékű illeszkedést a mérési modellhez. A nem illeszkedő elemek azonosításának fontos szerepe van a minőségbiztosításban, mert ha pretesztről van szó, az elemzések alapján a feladatok kritikus pontokon módosíthatók még az „éles” vizsga előtt, ha viszont a problémák magában az éles vizsgában jelentkeznek, a problematikus feladatok a pontozásból kizárhatók.

A horgonyzás is olyan lehetőség, amelyet más valószínűségi szoftverek is kínálnak. Lényege, hogy a különböző alkalmakkor használt feladatsorokban azonos feladatokat helyeznek el. A feladatsorok összekapcsolása révén a különbözőségek ellenére a Facets a teljesítményértékeket úgy számítja ki, veti össze mintha az összes vizsgázó egy csoportként egyetlen alkalommal oldotta volna meg az összes kérdést. A horgonyzás értelme, hogy kiegyenlítheti az évről évre vagy hónapról hónapra esedékes vizsgákat folyamatosan jellemző (látszólagos?) hullámzást; a mérésben érdekelt szakembereknek segít értelmezni a mindenkor feladatsorok nehézségét.

De akkor mégis miért érdemes szemlézni a Facetset? A rasch.org táblázata (<https://www.rasch.org/software.htm>) szerint a valószínűségi szoftverek többsége dichotom adatokat fogad, illetve skálaitemekkel (polytomous items) számol. Ezért a Facets legnagyobb előnye, hogy megvalósítja a többdimenziós Rasch mérést. Minden (skálálakkal) értékelt teljesítmény mérésére alkalmas – erre tervezték – amelyben értékelők,

ítészek vagy vizsgáztatók teljesítményeket értékelnek, jellemzően skálák és pontok segítségével. Dichotom adatok értékelésére – ez lényegében a „klasszikus” Rasch módszer – is alkalmas, bár Linacre szerint nem arra tervezték. A többdimenziós mérés arra is lehetőséget teremt, hogy a felhasználó egy teljes nyelvvizsgát feldolgozzon, tekintettel arra, hogy a modern, a kommunikatív idegen nyelvi mérés elveit követő nyelvvizsgák jellemzően erősen eltérő jellegű vizsgarészekből állnak. A receptív készségek, illetve a nyelvismeret (nyelvi rendszerismeret) mérésére jellemzően diszkrét elemekből felépülő kérdéssort alkalmaznak. Ezekre dichotom vagy skála-típusú válaszok adhatók. A másik csoportba a produktív készségek mérésére alkalmas szöveges választ kiváltó feladatok tartoznak, melyeket jellemzően analitikus skálák segítségével értékelnek. Egy teljes vizsga értékelése tehát nem kívánja meg két külön szoftver alkalmazását az eredmények számítása során.

A fentiekben túl a Facetszel végzett munka lehetőséget teremt adott dimenzió irányának megválasztására. Ilyen változó pl. a szótárhasználat a nyelvvizsgán, amelyet korábban nem engedélyeztek. Kb. 2012-óta viszont piaci nyomásra a vizsgázók döntésük szerint már használhatnak szótárt, ami lehet helyes és hasznos döntés, vagy helytelen, sőt káros is, továbbá felveti a kérdést, hogy mivel a szótárhasználat nem része az idegennyelv-tudás konstruktumának, a szótárhasználattal együtt mérve elvárható-e a vizsgázók tudásának megbízható mérése (Dávid, 2017). A dimenzió irányának meghatározhatósága fontos érték a Facetsben, mert a legtöbb dimenzió a vizsgázó ellenében hat, hasonlóan a gátfutáshoz mint metaforához: a gátfutó előtt először az itemek, feladatok „megugrása” áll, majd ezt követi az értékelő, a skálák stb., a szótárhasználat viszont felfelé lendíti a futót.

A Facets kipróbált, gyakorlati alkalmazásai terén egyértelműen az idegennyelv-tudás mérése vezet. A Magyarországon akkreditált nyelvvizsgák között az International Test of Language Competence (iTOLC) nyelvvizsga a valószínűségi tesztmodellt alkalmazza rendszeresen adatelemzésre és eredményszámításra (kalibrációra). A vizsgaeredmények számítása során operacionalizálásra kerül a feladatok, itemek és értékelők, skálák mellett az itemek formátuma (a feladat típusa), pl. feleletválasztós, többszörös választás vagy kiegészítendő, „nyílt végű” kérdés, valamint a válaszadás médiuma (a vizsga formátuma), amely lehet papíralapú, számítógépes „vizsgahelyes”, illetve számítógépes „online” (Dávid, 2023a).

Tágabb körben, de a mérés-értékelés témán belül, megemlítendő az analitikus szakdolgozati értékelő skálák beválásának és alkalmazásának mérlegre tétele (Dávid & Piniel, 2018). Ebben a projektben a skálák és értékelők (bírálok, témavezetők mint pontozók) mellett a Facets elemzés során az vált világossá, hogy a bírálok és témavezetők pontozásának eltérései nem a megbízhatóság hiányára utalnak, hanem az értékelő mellett egy önálló változó, a teljesítmény önálló tényezőjének dimenzióját jelzik (Dávid, 2023b).

Meg kell említeni a közös Európai Keretrendszer (KER) skáláinak magyar fordítását is (PTMIK, 2002) a Facets alkalmazásainak sorában. A KER nemcsak a nyelvtanulóknak és tanáraiknak, továbbá az eredeti szándékok szerint döntéshozóknak és az Európában munkát vállalóknak szól (Council of Europe, 2001), hanem legalább annyira alkalmas nyelvvizsgák (vizsgarendszerek) standardizációjára, a KER-hez való illesztésére. Tekintettel arra, hogy a KER magyar fordítása „egyszerű” fordítás volt,

amelynek megfelelőségét nem kísérte szakértői csoport ítései ellenőrzése (collection of judgements) és megfelelő elemzés, szükség volt a fordítás ellenőrzésére. Dávid (2012) arra vállalkozott, hogy a KER magyar nyelvű deskriptorainak érvényességét az angol és német nyelvű változatokkal együtt kalibrálva vizsgálja meg.

Előnyök, korlátok – összefoglalás

Az idegennyelv-tudás kommunikatív mérésével kapcsolatban McNamara (1996, pp. 48–90) és Widdowson (2001, p. 13) egyaránt a Pandora szelencéje metaforát használja. Widdowson el is magyarázza, hogy azok, akik újítként a kommunikatív nyelvoktatást egykor bevezették, aligha ismerték fel, micsoda Pandora szelencéjének, bonyolult problémák és nagyszámú tényező forrásának bizonyul majd újításuk, tekintve, hogy e tényezőket, „elszabadulván” a nyelvtanulás folyamatában, a szakemberek következő generációja már nem kezelheti mint állandókat és nem söpörhetők a szőnyeg alá. A teljesítményre ható változóként kell velük számolni.

Mint fentebb látható, a Facets alkalmazása lehetőséget teremt az idegennyelv-tudás komplex, többdimenziós mérésére és – tágabb értelemben – a mérés-értékeléshez közvetett módon (vagy akár semmilyen módon nem) kapcsolódó projektek eszközeinek, hatástényezőinek vizsgálatára az érvényesítés (validálás) céljával. További fontos előny, hogy a bevett tényezők, dimenziók mellett továbbiak azonosítása és értékelésére is van lehetőség. Viszonylag kis létszámú adatsomagok feldolgozására is alkalmas adott eljárás követése esetén, ami szélesíti a lehetséges felhasználók körét az oktatási rendszeren belül és kívül. Sok előnye mellett hátránya lehet, hogy a Facets „szigorú” szoftver, pontosságot követel meg az adattábla összeállításakor és a parancssorok megfogalmazásakor – csak azt látja, amit mi „látatni” akarunk vele. Ez egyben előnye is, véleményem szerint.

Irodalom

- Brumfit, C. J. & Johnson, K. (eds.). (1979). *The Communicative Approach to Language Teaching*. Oxford University Press.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge University Press.
- Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to language learning and testing. *Applied Linguistics*. 1(1). 1-47, <https://doi.org/10.1093/applin/1.1.1>
- Dávid, G. (2012). A szintleírások nyelvének szerepe a közös európai referenciakeret magyar, angol és német nyelvű kiadásában. *Magyar Pedagógia*. 112(1). 19–39.
- Dávid, G. (2017). Szótárhatás az idegen nyelvi mérésben. In Károly, K. és Homonnay, Z. (szerk.). *Mérési és értékelési módszerek az oktatásban és pedagógusképzésben*. Eötvös kiadó. pp. 271–281.
- Dávid, G. A. (2023a). Three mediums of test-taking and performance in the measurement of foreign language competence at level B2. *Alkalmazott Nyelvtudomány*, 23/1. 18–35. <http://dx.doi.org/10.18460/ANY.2023.1.002>
- Dávid, G., A. (2023b). The supervisor’s contribution. Is there a separate role for the supervisor in the rating of OTAK/MA-in-ELT theses? Powerpoint prezentáció.

- Dávid, G. & Piniel, K. (2018). Establishing categories in the design of rating scales for MA-in-ELT theses. *Working Papers In Language Pedagogy*, 12. 55–82. <https://doi.org/10.61425/wplp.2018.12.55.82>
- Johnson, K. & Morrow, K. (1981). (Eds.) *Communication in the Classroom*. Longman Group Ltd.
- Linacre, J. M. (1989, 1993, 1994). *Many-facet Rasch Measurement*. Mesa Press.
- Linacre, J. M. (2017). *A User's Guide to FACETS Rasch-Model Computer Programs*. Program Manual 3.80. Winsteps.com.
- Linacre, J. M. (2024). Facets: Rasch Measurement Computer Program. Version 4.1.3 [Computer software] Winsteps.com.
- Littlewood, W. (1981). *Communicative Language Teaching: An introduction*. Cambridge University Press.
- McNamara, T. (1996). *Measuring second language performance*. Longman.
- Morrow, K. (1979). Communicative language testing: Revolution or evolution? In Brumfit, C. J. and Johnson, K. (Eds.). *The Communicative Approach to Language Teaching*, (pp. 143–159). Oxford University Press.
- PTMIK (2002). *Közös Európai Referenciakeret. Nyelvtanulás, nyelvtanítás, értékelés*. Pedagógus-továbbképzési Módszertani és Információs Központ Kht.
- Rasch, G. (1960). *Probabilistic Models for some Intelligence and Attainment Tests*. Danmarks Paedagogiske Institut.
- Widdowson, H. (2001). Communicative language testing: The art of the possible. In Elder, C., Brown, A., Grove, E., Hill, K., Iwashita, N., Lumley, T., McNamara, T. & O'Loughlin, K. (Eds.) *Experimenting with Uncertainty: Essays in honour of Alan Davies*, (pp. 12–21). Cambridge University Press.

Facets, a tool for multi-dimensional measure

Linacre's Facets (2024) software not only makes the complex, multi-faceted measurement of the foreign language proficiency possible, but it also allows for the statistical analysis and validation of research instruments, factors (facets) of influence and their effects, more distantly related to (or unrelated to) assessment. Facets operationalises communicative language testing and assessment, as envisioned by Morrow (1979), Canale and Swain (1980) and others around 1980. Apart from calibrating the usual facets of performance, such as the difficulty of the tasks and items and the leniency or severity of the raters, Facets allows for the identification of further dimensions of performance. If special procedures are followed, Facets can also calibrate relatively small datasets, which in turn should increase potential interest inside and outside the world of education.

Keywords: assessment and testing, probabilistic measurement, Many-facet Rasch Model, Communicative Language Testing and assessment, Facets