

## A SZINTLEÍRÁSOK NYELVÉNEK SZEREPE A KÖZÖS EURÓPAI REFERENCIAKERET MAGYAR, ANGOL ÉS NÉMET NYELVŰ KIADÁSÁBAN

**Dávid Gergely**

*ELTE Angol-Amerikai Intézet, Angol Nyelvpedagógia Tanszék*

### A Közös Európai Referenciakeret és háttere

A *Közös Európai Referenciakeret* (KER; *Council of Europe*, 2001; *Europarat*, 2001) az ezredforduló jelentős fejleménye volt az idegennyelv-tanulás és a nyelvtanítás terén. Célja, hogy elősegítse az idegen nyelvek tanulását a nyelvtudás általános, az egyes nyelvek jellegzetességeit nem tükröző leírása segítségével. A KER az európai nyelvtanulók közös célrendszerét állítja elénk úgy, hogy az idegennyelv-tudást hat, betűvel és számmal jelölt szintre, szinttartományra bontja, a legalacsonyabb A1-estől az A2, B1, B2, C1-en keresztül a legmagasabb C2-es szintig.

Az idegennyelv-tudás szintekre bontása nem ismeretlen előttünk. A *kezdő*, a *középközvetlen* és a *haladó* jelöléseket már régóta használják a nyelvtanításban, akárcsak az *alap*-, a *közép*- és a *felsőszint* kifejezéseket. A KER értékét és fontosságát az adja, hogy míg a korábbi magyar terminológia tartalmát deklaratív, írott és közkeletű formában nem határozták meg – az idegennyelv-oktatási szakemberek azt inkább csak tapasztalataik alapján értették – addig a KER egy széles közönség számára pontosan fogalmazza meg az egyes szinttartományok tartalmát, jellemzőit.

#### A KER részei: a főszöveg és a skálák

A KER-nek két fő része van. Az egyik a leíró jellegű főszöveg, mely képet ad a szerzők nyelvszemléletéről, arról, hogy mit tartanak a kommunikatív idegen nyelvi kompetencia összetevőinek. Ez a rész általános. A másik fő részt a főszövegben elhelyezett több mint ötven skála (*illustrative scale*) alkotja, melyek – túl az alapkészségek szerinti felosztáson (beszédértés, beszédalképzés, írott szöveg értése, írásbeli kompozíciós készség stb.) – részkészségek, esetenként mikrokészségek szerint határozzák meg a szinteket. A skálákban az egyes szinteken szintleírások (vagy deskriptorok) vannak, amelyek meghatározzák, az adott szinten a nyelvtanulónak milyen kommunikatív nyelvi tevékenységek elvégzésére kell képesnek lennie. Innen ered, hogy a szintleírásokat '*Can Do*' *statementeknek* is hívják (*Council of Europe*, 2001).

A KER népszerű, sokat hivatkozott munka lett. Ennek egyik valószínű oka a skálák megléte, melyek tömören összefoglalják a leíró rész nyelvszemléletét, azokat mintegy átültetik a gyakorlatba, operacionalizálják. Ismert, hogy a nyelvtudás a részkészségek miatt többdimenziós jelenség (*Bachman, 1990; Bachman és Palmer, 1982; Bárdos, 2002; Canale és Swain, 1980*), amit a nyelvtanuló fokozatosan épít fel, de a részkészségek kiépülése az egyes dimenziókban nem szükségképp azonos ütemben és szinteken valósul meg. (A többdimenziós idegen nyelvtudás-modellek az 1980-as évek óta hegemón helyzetben vannak.) Az összességében azonos KER-szintre sorolható nyelvtanulók esetében is részkészségek szerint más és más besorolás lehetséges, viszont ezek az egyéni profilt alkotó eltérések folyamatos szövegben nem tükrözhetők elég jól. Mindez magyarázza azt, miért érdemes és kell a KER érvényessége kapcsán a skálákkal, azok érvényességével behatóbban foglalkozni (*North, 2000, 2004*).

A KER egyben értékelési eszközzé is vált – igaz, a szerzők egy része akarata ellenére (*Brian North és John Trim, személyes közlés*). Főszövegének gondolatai, leírásai a skálákban változnak át értékelési eszközzé, melynek segítségével az olvasó a saját maga vagy mások idegennyelv-tudását valamely szintre besorolhatja. Hangsúlyosan fogalmazva, a KER magyar fordításának érvényességét döntően a benne lévő skálák érvényessége határozhatja meg. E kutatási projekt középpontjában a KER magyar nyelvű kiadásában megtalálható skálák, szintleírások érvényessége áll (*PTMIK, 2002*). Arra kerestem a választ, hogy a KER magyar nyelvű fordításának mint a nyelvvizsgák illesztésére használt eszköznek az érvényessége mennyire alátámasztható az illesztés során felvett válaszadatok empirikus feldolgozása alapján.

### **A kutatás kontextusa és a magyar illesztési eljárások 2007-2009**

A KER megjelenése, 2001 után felmerült az igény az idegen nyelvi vizsgák illesztésére, azaz a KER szintjeivel való összevethetőségük megalapozására. Az illesztés módszertanának kidolgozására az Európa Tanács (ET) Nyelvpolitikai Osztálya keretén belül egy nemzetközi szakértői csoport vállalkozott. 2004-től Európa-szerte projektek indultak az idegen nyelvi vizsgák KER-hez történő illesztésére. Mind arra keresték a választ, hogy az illesztendő nyelvvizsga azt a szintet méri-e, amit a KER, illetve a vizsga deklarált célja szerint mérnie kell. Az illesztési projektek kulcskérdése szintén az érvényesség volt.

Bár az ET módszertana (*Council of Europe, 2003, 2004*) hangsúlyozottan kísérleti volt, a magyar jogalkotás túl korán írta elő, hogy 2007-től csak illesztett nyelvvizsga működhet tovább, miközben az európai szintéren a módszertan véglegesítése 2008-ra volt várható. A végleges változat csak 2009-ben született meg. Ezért, miután a honi jogi normák már megszülettek, a hazai szintéren kellett rögtönözni. A magyarországi vizsgaközpontoknak egy erősen standardizált eljárást kellett végigvinniük, melyhez adott volt a KER és a skálák, de eltérő nyelvi változatokban. A vizsgaközpontok közül azok, amelyek több nyelvből vizsgáztattak, lingua francaként elsősorban a magyar nyelvű KER-t használták, míg ott, ahol korábban csak egy idegen nyelvből vizsgáztattak (pl. Euro Nyelvvizsga Központ) a magyar változatot nem használták. Az illesztési projektek kiindulópontja nyelvek szerint más és más lehetett, ami érvényességi problémákat okozha-

A szintleírások nyelvének szerepe a közös európai referenciakeret magyar, angol és német nyelvű kiadásában

tott, különösen, hogy a szintillesztés során a KER nyelvvel és a nyelvi változatoknak az illesztés eredményeire gyakorolt lehetséges hatásával senki nem foglalkozott. Mindezek miatt az itt leírt kutatás a KER-skálák magyar fordítása alkalmazásának kvantitatív vizsgálatának tekinthető.

### **A KER fordításai és érvényességük**

A KER-skálákat a főszöveggel együtt az ET-tagországok nyelveire már korábban lefordították. Ma már 38 nyelven áll rendelkezésre (Council of Europe, é. n.). A magyar nyelvű kiadás (PTMIK, 2002) az angol alapján készült (Council of Europe, 2001).

A KER-ben a skálák szövegezése különösen érzékeny pont, hiszen miután a fordító elvégzi feladatát, a felhasználók (vizsgáztatók, tanárok vagy az illesztési projektekből a zsűrizést végzők, az ítések) a célnyelvi változat alapján nem biztos, hogy ugyanarra a gondolati tartalomra gondolnak, mint ami eredetileg a skálák összeállítóinak szándéka volt. A fordítás megfelelése tehát fontos szempont, hiszen az illesztendő vizsga megítélése a fordítás áttételén keresztül valósul meg. Azt várjuk el, hogy a fejlesztésben közreműködő külföldi szakemberek szemléletmódját, megközelítésmódját, gondolkodását a honi felhasználó megértse és a maga számára rekonstruálhassa. Ehhez elsősorban ismerni kell a KER-t, továbbá olyan kollektív tudat kialakítása is szükséges, amellyel a KER-t az 1990-es években fejlesztő projekt résztvevőinek (szerzők, szerkesztők és műhelyfoglalkozások résztvevői, ítések) sajátos szemlélet- és megközelítésmódja, együttgondolkodása jellemezhető. Nevezhetjük ezt szinergiának is, amit ennyi évvel a fejlesztés után a hazai szakemberek számára ma már csakis a szintleírások megfogalmazásai jeleníthetnek meg, valamint az, hogy mely szintre sorolták be őket. Ha a fordítás nem sikerül elég jól, elképzelhető, hogy az átültetett szintleírások felhasználói – akiknek feladata, hogy a szerkesztők gondolkodásmódját a maguk számára újra felépítsék – a szintleírás szintbesorolását nem érzik helyénvalónak, vagy egy másik szint szintleírásával összetéveszthetik, vagy mint egy másik szinthez tartozót tételezik. Ha azonban a vizsgaközpont munkatársai kielégítő érvényességű magyar nyelvű skálák alapján dolgoznak, az általuk vizsgáztatott diákok eredményeit, a vizsga szintjeit, a pontszámban kifejezett minimumkövetelményt is más vizsgákénál megalapozottabbnak tekinthetjük.

### **A KER fejlesztése**

A KER-t többnyelvű (angol, német és francia) nyelvi környezetben az ET szervezésében fejlesztették, elsősorban Svájcban. A KER szintleírásai először angol nyelven születtek meg 1994-ben, korábbi más angol nyelvű értékelési skálák egymondatos állításokra történő szétbontása útján, majd a német, francia és további angol szintleírások 1995-ben. Kezdetben körülbelül 4000 állítás tartozott a gyűjteményhez, melyeket nagyszámú nyelvtanár, szakértő műhelyfoglalkozások sorozatán válogatott és moderált, míg a szintleírások száma közel 1000-re csökkent. Ezek után az értékesnek és megtartandónak ítélt szintleírásokból kérdőíveket szerkesztettek. A kérdőívek összekapcsolását közös szintleírások tették lehetővé. Ezeket a projekt résztvevői diákok értékelésére használták videón, felvételnél vagy élőben a tanórán. A kérdőívek válaszait átfordították egy ötfokozatú

(0, 1, 2, 3, 4) mérési skálára. A válaszadatok szoftveres kalibrációja után minden szintleírásnak lett egy logit mérőszáma (értéke), melynek segítségével be lehetett sorolni a hat szint egyikére. A fejlesztők további szinteket is meghatároztak. Az A2+, B1+ és B2+ három szinttartomány felső felét jelölte (nem az e szintek feletti tartományt). Számunkra a legfontosabb, hogy az így kalibrált szintleírások egy mérési skálát alkottak – nevezzük ET-skálának – melyben az elkülönülő szinteket határoló értékek választották el egymástól (North, 2000; North és Schneider, 1998; Schneider és North, 2000).

A skála szót ebben az esetben két értelemben szükséges használni. Egyrészt a KER-ben fellapozható, a készségek és mikrokészségek szerint csoportosított és szintmegjelöléssel ellátott, táblázatos formában közölt szintleírásokat nevezem skálának. Másrészt skála az a matematikai statisztika segítségével előállított karakterisztika is, amit az összes együtt mért szintleírás logit szintértéke határoz meg. Ezt ET-skálának nevezem. A magyar nyelvű kiadással kapcsolatban ugyanez a kettősség áll fenn: a magyarított szintleírásokat tartalmazó skálák megtalálhatók a magyar kiadásban, míg az itt leírt kutatás során Magyarországon gyűjtött válaszok alapján meghatározott statisztikai karakterisztikát magyar/magyarországi skálának nevezem.

Míg a magyar nyelvű változat csupán olyan fordítás, amit nem alapozott meg érvényesítés, a skálák angol nyelvű változata, amint a fenti leírásból kitűnik, nem fordítás eredménye. Az angol nyelvű szintleírások angol egynyelvű vizsgák szintleírásaiból származtak (North, 2000). Nyelvezetük, megfogalmazásaik általában nem változtak a moderálás és a kvalitatív fejlesztési munka egyéb szakaszai során, sőt – természetesen – a szoftveres kalibráció során sem. A német skálák eredetében található ugyan ellentmondás (Schneider és North, 2000), melynek részleteire a tanulmány keretein belül nem tértek ki. Az írott szöveg értése North (2000) munkájában és kalibrációjában nem szerepelt, így azokat a KER számára valószínűleg először németül fogalmazták meg és mérték be 1995-ben.

## Elméleti háttér

### Érvényesség

Az érvényesség a nyelvvizsgáztatás terén a ponteredmények igen fontos és összetett értékmutatója, ami megmutatja, mennyiben mérvadók, fogadhatók el, illetve mennyire hitelesek az eredményekből levonható következtetések. A tanulmány keretein belül használt érvényesség-felfogás modellje Messick (1989, 1995) munkáin alapul. Messick nézetei ma a bevett nézetrendszer alapját képezik a pszichológiai mérés terén. Nézetei közül számunkra a mérendőhöz vagy konstruktumhoz (*construct*) nem köthető szórásnégyzet (*construct-irrelevant variance*) a legfontosabb.

Az érvényesség régi, de máig közkeletű meghatározása szerint a mérési eszköz azt kell mérje, amit az mérni hivatott (Kelley, 1927; Cureton, 1951). Viszont a gyakorlatban általában nem az a probléma, hogy a mérési eszköz a kiválasztott mérendőt ne mérné. Sokkal inkább az, a gyakorlat tapasztalatai szerint, hogy egyúttal mást is mér. A gyakor-

lat tapasztalatainak magyarázatát, elméleti alapját nyújtja a *messicki* felfogás. Maga a mérési eljárás olyan tényezőket hordoz magában, amelyek nem tartoznak a mérendőhöz, de feltétlenül hatnak a mérés eredményeire: varianciát generálnak és egyszerű eszközökkel nehéz hatásuk mértékét meghatározni. Ilyen tényező lehet minden, ami a mérési módszer fogalmába tartozik: bármilyen feladat, az értékelési skála, az értékelő, továbbá a válaszadók neme és ide tartozik a válaszadás vagy a fordított szöveg nyelve is. Az idegen nyelvi mérés performancia-tényezőinek is nevezik ezeket. A mérési eszközökről empirikus kutatásokból tudjuk, hogy nem semlegesek (pl. *Dávid*, 2007). Kivétel e kutatás eszközei sem jelentettek, mert a KER-ismeretet is csak valamilyen mérési eszköz, például kérdőív, teszt segítségével lehet megjeleníteni.

A *messicki* felfogás szerint a mérésre ható különféle nem releváns tényezők hatásában nehezítők vagy könnyítők, de mindenképp összeadódnak és a pluszok (nehezítő tényezők) és mínuszok (könnyítő tényezők) eredője adja a feladat végleges nehézségét. Mindezek alapján elkülöníthető a mérendőhöz nem köthető nehézség és könnyűség (*construct-irrelevant difficulty/easiness*).

A kurrens érvényesség-felfogás a korábbi, különféle érvényességfajtákat (egyetértési, prediktív, tartalmi stb.) egységes rendszerben szemléli. Ebből az is következik, hogy adott érvényesítési kérdésben, egy konkrét projekttől nem várható teljes körű érvényesítés, mert az érvényességnek csak valamely aspektusát/aspektusait vizsgálja. Nincs tehát teljesen kész érvényesítés, nem mondhatjuk el, hogy például a vizsgát érvényesítettük, legfeljebb bizonyítékot találunk, gyűjtünk az érvényesség támogatására. Ezen esetben sem várhatjuk tehát annak kimondását, hogy érvényesítettük a KER magyar fordítását, hanem legfeljebb azt, hogy az itt nyert adatok, elemzések alátámasztják, megerősítik a KER magyar fordításának érvényességét (vagy nem), az angol kifejezéssel élve *validity evidence* bemutatására vállalkozunk.

Az érvényesség *messicki* felfogásával kapcsolatban fontos az alternatív magyarázatok megdöntésének (*discounting rival hypotheses*) szükségessége is (*Messick*, 1989). Az érvényesség vizsgálatakor a kutató könnyen megfedkezhet az egyéb lehetséges, gyakran nagyon egyszerű magyarázatokról, melyek aláássák az eredmények érvényességét. Nyelvvizsgák illesztésének eredményeit kérdőjelezheti meg, ha például nem számolnak az ítések eltérő szigorával (*judge severity*) a KER egyes szintjeihez kapcsolódó követelmények tekintetében. Hasonló módon, e kutatás eredményei is megkérdőjelezhetők, ha nem számolnánk a fordítás lehetséges torzító hatásával. E tanulmánynak nem feladata az érvényesség irodalmának mélyebb ismertetése. Mégis érdemes felvillantani az érvényességgel kapcsolatos bevett nézetrendszer időről időre támadó, kétségbevonó nézeteiből is. Ilyen a *Borsboom*, *Mellenbergh* és *van Heerden* (2004) által javasolt, a kurrens nézetrendszerrel sokkal egyszerűbb érvényesség-konceptió is, miszerint a mérés érvényessége megalapozott, ha a mérendő (konstruktum) létezik és annak variabilitása okozati összefüggésben áll (*causally produce variation*) a mérés eredményeivel. *Borsboom* és munkatársai (2004) támadják az alternatív magyarázatok megdöntésének szükségességét is mint feleslegest, mivel az érvényesség józan ésszel megítélhető. Azonban úgy tűnik, nem számolnak a mérés járulékos, a mérendő szempontjából nem releváns szórást generáló módszertényezőkkel és az érvényességet mint egységes, kvalitatív és kategoriális (igen vagy nem) jellemzőt fogják fel.

A *messicki* érvényesség-felfogás továbbra is az érdekelt szakemberek és szakterületek konszenzusának alapját képezi, amit *Borsboom* és munkatársai (2004) felvetései nem ingattak meg. Kritika inkább az érvényesség fogalmának a *consequential validity*-vel való kibővítését érte az elmúlt években. *Popham* (1997) szerint a következmények beemelése az érvényesség kritériumai közé túlbonyolítja, túl összetetté teszi a fogalmat, ami e kiterjesztés előtt viszonylag egyszerűen megfogalmazható volt: az érvényesség a számszerű mérési eredményekből levonható következtetések helyessége. Egy példával szemlélteti, hogy a döntéshozói magatartás miatt nem határozhatja meg alapjaiban a mérés érvényességét. Mivel *Messick* (1989) nem sok támpontot kínál a kutatóknak arra, hogyan vizsgálják a mérés érvényességét, az utóbbi években az érvényesítés módszertana (*argument-based approach to validation*) került előtérbe (*Kane*, 2012).

### Valószínűségi mérés

A kutatás során a valószínűség-számításra épülő Rasch-módszert (modellt) alkalmaztam. Kiindulópontja az értékelés során születő nyers jegyek (pl. tanári osztályzatok, a vizsgáztató pontozása), melyekből kalibrált készségértékeket számít. Alapjait *Georg Rasch* (1961) matematikus dolgozta ki, innen az eljárás elnevezése. A Rasch-módszer tagja a valószínűségi módszerek szélesebb családjának, amit a szakirodalom item-válasz elméletnek (*Item Response Theory*, IRT) nevez. (A klasszikus és modern tesztelméletekről, illetve specifikusan a Rasch-módszerről lásd még *Csapó*, 1993; *Horváth*, 1993, 1997; *Molnár*, 2003, 2005, 2006.) A Rasch-módszer továbbfejlesztése a többdimenziós *Many-facet Rasch Measurement* (MFRM; *Linacre*, 1994), ami az itemek és vizsgázók mellett több, elméletileg korlátlan számú, a vizsgázónál mért teljesítményre ható, de a mérendő szempontjából nem releváns performanciatényezőt különít el és hatásukat semlegesíti, vagyis a mérésből származó összes varianciát alkotóelemeire bontja az azt alakító tényezők szerint. A módszer jellemzői közül különösen ez utóbbi harmonizál a validitás *messicki* felfogásával, mert a kommunikatív módszerek szerinti értékelésnél elkerülhetetlen, hogy performanciatényezők (releváns és nem releváns) hatást gyakoroljanak a vizsgázó teljesítményére.

A készségértékek számítása mellett az MFRM-et operacionalizáló *Facets* (*Linacre*, 2006) szoftver által nyújtott legfontosabb, minőségbiztosítási jellegű információ az illeszkedési statisztikák, melyekből megtudhatjuk, hogy a mérés mely elemei (vizsgázók, vizsgáztatók, feladatok, értékelési skálák stb.) illeszkednek a valószínűségi mérési modellhez és melyek nem. Egy találmásra választ adó vagy puszkázó vizsgázó válaszai lehetnek a példa a gyenge illeszkedésre, akinek válaszaiból, ahogy az ítésk, a szoftver sem tud egyértelmű következtetéseket levonni, mert nehéz és könnyebb feladatokat (szintleírásokat) egyaránt eltalál, vagy nem az ismeretszintjének megfelelő nehézségű szintleírásokra ad helyes választ. A szoftver az ilyen zavaros vagy ellentmondásos válaszsorot értelmezi nem illeszkedésként, mert nem felel meg a mérési modell elvárásainak. Ami a feladatokat illeti, nem illeszkedik például a nehéz feladat, melyre több gyenge, mint jó tudású vizsgázó ad helyes választ, vagy a helyesen válaszolók között legalább annyi jó, mint gyenge tudású vizsgázó van.

A szintleírások nyelvének szerepe a közös európai referenciakeret magyar, angol és német nyelvű kiadásában

Nem utolsósorban fordítási problémák is okozhatnak nem illeszkedést. Az illeszkedés értékelése sikerrel azonosíthat olyan szintleírásokat, melyek fordítása nem megfelelő vagy a válaszadók más okból nem tudják a szintleírást konzisztens módon elhelyezni a KER-skálán, amennyiben tudható, hogy a mérési eszköz mért készséget homogén módon mér. Ha a nem illeszkedő elemeken mégsem lehet kivetnivalót találni, elképzelhető, hogy az elem mást mér, vagyis a mérési eszköz inhomogén.

### **A KER illeszkedési kérdései**

Bár a KER méltán tekinthető vonatkoztatási pontnak, mérési kérdésekben a nyomtatott verzió (bármely nyelven) csak megfelelő körültekintéssel és más kiadványokkal együtt alkalmazható. Egyrészt nem minden skálát kalibráltak. Néhány ennek ellenére bekerült a végső nyomtatott verzióba (*Council of Europe, 2001*), erre megjegyzések figyelmeztetnek egyes skálák alján. Hasonlóképp, olyan skálák is bekerültek, amelyek a bemérések alapján a nyelvtudás más dimenziójába tartoztak (pl. a *Szociolingvisztikai megfelelés* vagy a *Fonológiai tudás*). Végül, összevetve az angol nyomtatott verziót a legfontosabb háttér munkával (*North, 2000*), kiderül, számos szintleírást a könyv szerkesztése során, de már a kalibrálás után átírtak, módosítottak, még olyanokat is, amelyekkel a kalibráció során jó eredményeket értek el. Mivel az utólagos módosítások a magyar fordításba is bekerültek, hatottak saját kutatásunk technikai, módszerbeli aspektusaira, például arra, hogy az utólag átírt szintleírások referenciaként már nem voltak felhasználhatók.

## **A kutatás módszerei**

### **Célok, kérdések**

A kutatás során megválaszolandó kérdések középpontjában a KER magyar nyelvű fordításában megtalálható skálák álltak. Mivel e skálák érvényességét korábban nem értékelték, a kérdések a következők voltak: (1) Az illesztés során gyűjtött válaszadatok alapján milyen mértékben érvényesíthetők a magyar nyelvű skálák? (2) Illeszthetők-e a magyarországi válaszadatok a KER-hez, annak skáláihoz? (3) A KER skálái a magyarra fordítás során torzulhattak-e úgy, hogy később, a vizsgák illesztése során az ítések döntően más szintekre sorolták be őket?

### **Eljárás**

A kutatásban felhasznált adatok a magyarországi vizsgaközpontok egy részében (Euro, ITK és Pannon) lefolytatott illesztési eljárásokból származnak. Az adatok az illesztési projektek első, familiarizációs szakaszából valók (*Council of Europe, 2003*). Ebben a szakaszban a vizsgaközpontok munkatársainak az volt a feladata, hogy KER-rel kapcsolatos ismeretüket felfrissítsék, bővítsék és kialakítsák a hozzá tartozó kollektív tudatot,

majd több kérdőív kitöltésével adtak számot tudásukról: meghatározták az összesen 829 angol, német és magyar nyelvű szintleírás szintjét.

A szintleírások három kérdőívbe kerültek be, nyelvek szerint. Az angol nyelvű kérdőívet (273 szintleírás, forrása a *Council of Europe*, 2001) az angol nyelvből vizsgáztató kollégák, míg a német nyelvűt (270 szintleírás, forrása az *Europarat*, 2001) a német nyelvből vizsgáztató kollégák töltötték ki. Azonban válaszaik nem lettek volna összevetethetők, ha mindkét csoport tagjai nem oldották volna meg a két magyar nyelvű kérdőív legalább egyikét (összesen 286 szintleírás, forrása a *PTMIK*, 2002)<sup>1</sup>. Ily módon 154 vizsgáztató az angol vagy német kérdőívet valamelyik magyar kérdőívvel együtt töltötte ki, ami fejenként körülbelül 400 szintleírás szintjének meghatározását jelentette.

A válaszadókat nem osztottam különféle csoportokba, mert – valamilyen szinten – mind ismerték a szintleírásokat magyarul és a számukra releváns idegen nyelven. A válaszadók készültek az itt tárgyalt felmérésre. Az, hogy ki melyiket ismerte (ismerhette) jobban, egyrészt függött a válaszadók egyéni tapasztalataitól, szakmai életútjától, másrészt vizsgaközpontjuk döntésétől, hogy melyik nyelvi változatot használják inkább. Nem lett volna életszerű (lehetséges) olyan válaszadói csoportokat felállítani, amelyek ismerték az egyik, de nem ismerték valamely másik nyelvi változatot, mert valószínűleg ilyen válaszadók egyáltalán nem vagy alig léteztek. Feltételezhető, hogy a válaszadók között sokan voltak olyanok, akik az egyik nyelvi változatot vizsgaközpontjuk munkája kapcsán intenzíven alkalmazták, míg a másik nyelvi változatot egy másik vizsgaközpontnál vagy máshol végzett munkájuk kapcsán ismerték.

A válaszadás folyamatát, körülményeit elemezve meghatároztam a mérési tényezőket (*facets*), melyek hathattak a felmérés eredményére. A középpontban álló KER-ismeret (ítészek) tényező, valamint a szintleírásoknak mint a kérdőívek kérdéseinek nehézsége (szintje) mellett várható volt, hogy a felmérés körülményei (vizsgaszituáció vagy otthoni munka)<sup>2</sup> és a kérdőívek három nyelve hathat az elért eredményekre. Az elemzéseket a Facets szoftverrel (*Linacre*, 2006) végeztem. Összefoglalva, az érvényesség operacionálizált kritériuma ebben a kutatásban az volt, hogy a válaszadók mennyiben sorolják a magyarított szintleírásokat az ET eredeti besorolásaival azonos szintekre – figyelembe véve a válaszadók egyéni beállítódásaiból, például az értékelői szigorból (*severity*) vagy „lágyságból” (*leniency*) és a szintleírások három nyelvből eredő különbségeket. A következők összevetése történt meg: KER-szintleírások eredeti, kalibrált ET-szintmeghatározásai mint kategóriák (A1, A2, B1 stb.) és a KER-szintleírásoknak e vizsgálat résztvevőitől származó kategoriális válaszaik (A1, A2, B1 stb.) épülő kalibrált értékeknek KER szerinti kategóriákba sorolása. Ez az operacionálizált érvényességkritérium nem szükségképp azonos azzal, amit a fordítás szakértői a jó fordítás kritériumainak tartanak.

<sup>1</sup> Egy csoport eszperantó szakos vizsgáztató csak a magyar nyelvű kérdőívekre válaszolt, mert a KER eszperantó változata akkor még éppen nem volt elérhető.

<sup>2</sup> Egy korábbi projekt során (*Dávid*, 2008), amely az illesztési módszertan próbájának (pilot) tekinthető, a válaszadók kis csoportjának engedélyeztem, hogy a kérdőíveket otthon töltsék ki. Ez nyilvánvalóan avval a következménnyel járt, hogy a felmérés körülményeit mint tényezőt be kellett kódolni az adattáblába. Az ezt követő illesztési projektek során azonban a kérdőíveket már vizsgaszituációban töltötte ki minden válaszadó.



## A vizsgálat szakaszai

A kutatás két jól elkülöníthető szakaszban zajlott. Az első szakasz célja a válaszadatok konzisztenciájának meghatározása volt. A második szakaszban történt a mérés. Mindkét szakasz a szoftver többszöri futását, több elemzési kör beiktatását követelte meg.

Az első szakasz célja a mérési modellhez nem illeszkedő elemek (válaszadók és szintleírások) azonosítása, valamint az ítések ismereteinek a felmérése volt. Valamely elem nem illeszkedése esetén a nem illeszkedő elemet semlegesíthetjük úgy, hogy kizárjuk a további elemzésekből, s a bennmaradó elemek már megfelelnek a mérési modell elvárásainak. (A kizárt elemek, *Horváth* (1997) nyomán, elhagyott elemeknek is nevezhetők.)

A nem illeszkedő elemek kizárása, az adattisztítás (*data cleaning procedures; TESOL*, é. n.), mindkét szakaszban az illeszkedés javításának, továbbá a *messicki* nézetrendszer másik fontos elemének, az alternatív magyarázatok kiküszöbölésének eszköze volt. Ha az ítései ismeretek felmérése során nem zárunk ki ítéseket, vagy nem zárunk ki nem illeszkedő feladatokat, tápot adunk a következő „ördög ügyvédje” típusú kérdéseknek: Nem modellszerűen válaszoló ítések hogyan érvényesíthetnek sikeresen szinteket? Hogyan hozhat nem megfelelően felkészült ítéscső érvényes és megbízható ítéleteket a szintleírásokról? Hogyan lehet adekvát módon megítélni a szintleírásokat, köztük a magyarra fordított leírások érvényességét nem illeszkedő ítéscső válaszai alapján?

Az adattisztítási folyamat végére olyan adattáblát sikerült előállítani, amelyben a zavaró elemeket már kiszűrtem és a szintleírások megítélése, a befolyásoló tényezők hatása nélkül, már világos, letisztult formában elvégezhető. Mivel a kutatás szempontjából a szintleírások illeszkedésének mértéke volt a legfontosabb, ebben a szakaszban először és elsősorban válaszadókat zártam ki. Amikor ez megtörtént, láthatóvá váltak, könnyebben azonosíthatók voltak a nem illeszkedő szintleírások is.

Az elemzések összesen 16 körben folytak le. A folyamat végére 752 jól illeszkedő szintleírás és 140 ítéscső maradt, így az ítéscső KER ismerete megítélhető volt és az ítéscsőktől elvárt minimumteljesítményt mint az elérhető pontok százalékát meg lehetett határozni (40%). A KER-ismeret operacionalizált formája ítéscsőként egy pontszám volt, ami a helyes válasz/helytelen válasz alapján szerzett 1 és 0 pontokkal számítottam ki a szoftver segítségével.

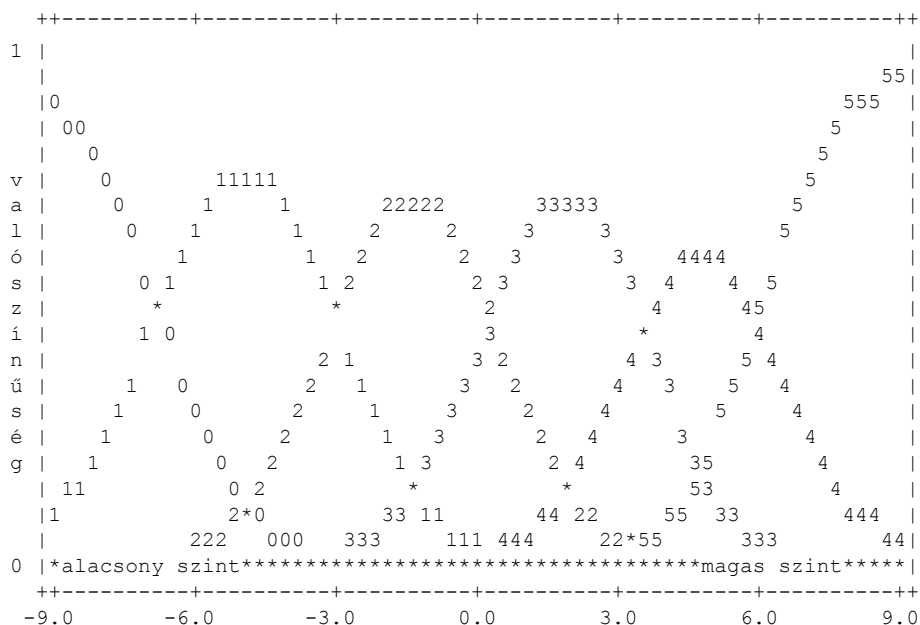
Az első szakasz előkészítette az elemzések második szakaszát, amelyben már a szintleírások álltak a középpontban. Azt vizsgáltam, hogy a válaszadók milyen mértékben helyezik a szintleírásokat ugyanazokra a szintekre, mint ahogy azt a KER fejlesztői tették. Ez a szakasz ugyanannak az adattáblának egy másik nézetére épült, mivel itt a szintleírások szintezése, a válaszadók szintítélete volt a középpontban. Tehát nem volt elvárt helyes válasz, hanem a válaszokat, a KER-szintmegjelöléseket (A1, A2, B1, B2, C1 és C2) a szoftver számára egy hatfokozatú numerikus skálára (0, 1, 2, 3, 4 és 5 pontokkal) kódoltam át. A különbség úgy is megfogalmazható, hogy míg az első szakaszban a válaszokat dichotóm módon értékeltem, addig a másodikban a hatfokozatú skálán, abban a formában, ahogy azok eleve létrejöttek. A kódolt tényezők köre nem változott az első szakaszhoz képest. A szintleírások szintje (nehézsége) állt a középpontban, amelyre hat-

hatott az ítések szigora, illetve „lágysága”, a felmérés körülményei és a szintleírások nyelve.

Adattisztításra e szakaszban is szükség volt. Ki kellett zárni azokat, akik a felmérésben nem értek el legalább 40%-os eredményt és további nem illeszkedő ítéseket is, valamint ez utóbbiak egyértelműbb azonosítása érdekében ki kellett zárni a legkevésbé illeszkedő szintleírásokat.

### Az ET-skála magyar megfelelőjének megalkotása

A második szakasz a fent leírtak mellett további magyarázatot igényel. Ha az ítések A1-től C2-ig hat szintre osztályozzák a szintleírásokat, lényegében egy hatfokozatú mérési skálát alkotnak meg közösen. A több ezer magyarországi válaszadat egy olyan karakterisztikát határozott meg, amelyben a hat fokozatnak (skálapontnak) hat egymásba érő görbe felel meg. Az 1. ábra a szintleírások szintje és a valószínűség kapcsolatát mutatja be.



1. ábra  
A magyar KER-skála valószínűségi görbéi

Az ábra vízszintes tengelyén balról jobbra haladva növekszik a mért készség, ebben az esetben a szintleírásoknak a KER szintje, míg a függőleges tengelyen felfelé a valószínűség növekszik. Látható, hogy a 0-val jelölt A1-es szinthez tartozó hullámgörbe ma-

gasról indul és lefelé tart, azaz ahogy növekszik a készség, úgy csökken annak valószínűsége, hogy A1-es nyelvtudással van dolgunk, míg az 1-gyel jelölt A2-esé már növekszik, majd a csúcstalálódási pont, a nyelvtudás adott szintje után ennek valószínűsége is csökken. Közben, továbbhaladva jobbra a felsőbb szintek felé, a 2-vel jelölt B1-es szint valószínűsége növekszik és így tovább a felsőbb szintek felé. (A Rasch-modell kiterjesztésére vonatkozóan lásd még *Molnár* (2008) vonatkozó magyarázatait.)

Míg a valószínűségi görbék íve, hullámvázása többé-kevésbé megjósolható, a skálát a görbék \*-gal jelölt metszéspontjai definiálják egzakt módon. Ezeket küszöbértékeknek (*step thresholds*) nevezzük, mert elhatárolják a skála egyes fokozatait (*Linacre*, 1998). Azt árulják el, hogy az, amit mértünk (szintleírás, ítésk, nyelv stb.) nagyobb valószínűséggel B1-nek, mint A2-nek tekinthető-e, illetve mely értékektől tekinthető inkább B2-nek, mint B1-nek és így tovább. A mérésben szereplő bármely tényező elhelyezhető a görbék valamelyikén. A hullámgörbék metszéspontjai, azaz a magyar válaszadatokra jellemző küszöbértékek táblázatosan is megjeleníthetők (l. 3. táblázat).

A magyarországi válaszadatok egy olyan A1–C2 skálát határoztak meg, amely a magyar válaszadók közös nevezőjének tekinthető és a küszöbértékei segítségével pontosan leírható. Az így nyert magyarországi skála segítségével minden szintleírás mért értéke szerint egyértelműen besorolható volt valamelyik KER-szintre, megengedve, hogy e besorolások nem biztos, hogy megegyeztek az ET-skála szerinti besorolásokkal. A magyarországi skála konstruálása után következett még a munkának az a része, amelyikben a magyarországi besorolásokat az ET-skála szerint igazítottam, vagyis ahhoz illesztettem.

## A magyar skála illesztése

Az illesztés érdekében a KER fejlesztőinek referenciaértékű kalibrációs munkájához kellett visszanyúlni. A klasszikus módszerrel, *North* (2000), illetve *Schneider* és *North* (2000) alapján gondosan összeválogatott, ott jól bevált szintleírásokkal mint referenciafeladatokkal (*anchor*) lehetett a magyar skálát az ET-skálához kapcsolni.

A második szakasz elemzése is több körben zajlottak, melyek során a referencia szintleírások egy részét ki kellett cserélni, mert a szelekció követelményei szigorúak voltak. A kiválasztott szintleírásoknak jól kellett illeszkedni mind *North* és *Schneider* (2000) méréseiben, mind a még nem illesztett magyarországi mérésekben. Nem lehettek saját szinttartományuk marginális példányai sem. A kiválogatott szintleírásoknak az illesztendő magyar mérésekben a szintek jellemző, elvárt tartományába eső értékeket kellett mutatniuk. Tehát a szintleírás, amit a KER fejlesztése során pl. B1<sup>+</sup>-ként mérték, a B1<sup>+</sup> tartományba kellett essen az illesztendő magyar mérésekben is.

Az így csoportosított 31 referencia szintleírásnak a KER fejlesztői által mért logit nehézségi értékeiket először lineárisan át kellett váltani a magyarországi adatfelvételtől eredő hosszabb skálára. A konverziót az indokolta, hogy a KER fejlesztése során a válaszokat egy ötfokozatú skálán rögzítették (0, 1, 2, 3 és 4 pont), míg a magyarországi illesztésben (így e kutatásban is) a hat KER-szint miatt hatfokozatú volt a skála, amin a

válaszokat rögzítettem. Ezt követte a konvertált értékek átlagának, azaz a csoportátlagnak a rögzítése (*group anchoring*) a magyarországi adatsomagban. Az itt alkalmazott technika nagyon finom módszer, mert csak az átlag rögzítendő, melyen belül az egyes szintleírások értékeit a szoftver állíthatja.

## Eredmények és tanulságok

Az első szakasz eredménye az ítések KER ismeretének lemérésén túl a teljesítmény feltételezett tényezői meglétének bizonyítása volt. Mind a szintleírások nyelve, mind a vizsgaszituációnak mint tényezőnek a létezése bizonyítható volt, bár hatásuk, a hozzájuk kapcsolható szórásnégyzet meglehetősen kicsi (1. táblázat). Figyelemre méltó, hogy később, amikor a 40% alatt teljesítő ítéseket kizártam az elemzésből, a fordítási tényező már nem volt szignifikáns. Ennek az eredménynek a magyarázata az lehet, hogy a KER nyelve csak a KER-t kevésbé ismerő vizsgázatóknak okozhatott igazán problémát. Ez az eredmény azt sugallja, hogy a magyar fordítás a célnak megfelelő volt és az erre épülő illesztési projektek érvényességét nem érintette: a vizsgázatók nem értek el az általuk használt nyelvi verzió miatt jobb vagy gyengébb eredményt.

1. táblázat. Az összes variancia megoszlása

<i>Varianciakomponensek</i>	<i>SD</i>	<i>Variancia</i>	
Vizsgázatók tudása	1,32	1,742	58,1%
Vizsgaszituáció	0,21	0,044	1,5%
A KER nyelvei	0,05	0,003	0,1%
Szintleírások	1,10	1,210	40,3%
Összesen	2,68	2,999	100%

### A nem illeszkedő szintleírások elemzése

Bár az egyes tényezők illeszkedése megfelelő volt mind az első, mind a második szakaszban, a teljesítménytényezők egyes elemei (pl. válaszadók és szintleírások) nem illeszkedtek. A továbbiakban áttekintjük ez utóbbiak illeszkedési problémáit, mert felvetődik a kérdés, mitől nem illeszkednek a korábban kalibrált, angol és német nyelvű szintleírások. A nem illeszkedő szintleírások számba vétele és elemzése azt a benyomást erősíti meg, hogy többségük esetében nem valószínű, hogy a fordítás miatt váltak nem illeszkedővé. Kigyűjtve a nem illeszkedő szintleírásokat, túlnyomó részük a legalsó, valamint a legfelső, vagyis a marginális szintekről való (2. táblázat). Magyarázata az lehet, hogy a magyarországi vizsgázatók érdeklődése nem irányul e szintek felé, mert az államilag elismert nyelvvizsgák rendszerében nincs A1 vagy C2 szintű vizsga. Az is lehet-

A szintleírások nyelvének szerepe a közös európai referenciakeret magyar, angol és német nyelvű kiadásában

séges, hogy a vizsgázatók kevésbé is ismerik azokat a nyelvtanulókat (vizsgázókat), akik az A1 vagy C2 szintekre sorolhatók be. Ugyanakkor a KER-t épp ezen szintek kapcsán érte kritika, miszerint az A1, de még inkább a C2 szinten eleve túl kevés szintleírás található, azaz e szintek alulspecifikáltak (Milanovic és Weir, 2010).

2. táblázat. A nem illeszkedő szintleírások

Feladatlap nyelve	Szintleír. azonosító	A skála neve, a magyar nyelvű kiadás szerinti oldalszám	Eredeti szint	A nem illeszkedéssel kapcsolatos megfigyelés
Angol (6)	17	Folyamatos monológ: élménybeszámoló, 74. o.	C2	marginális
	40	Anyanyelvű beszélők közötti társalgás megértése, 82. o.	A1	nincs leírás
	157	Információcsere, 97. o.	C2	ugyanaz, mint a B2 szinten
	190	Feljegyzések, üzentek és formanyomtatványok, 100. o.	B1	ugyanaz, mint a B2 szinten
	206	Magyarázatkerés, 103. o.	A1	nincs leírás
	252	Rugalmasság, 150. o.	A1	nincs leírás
Magyar (10)	349	Általános beszédértés, 82. o.	A2+	marginális
	441	Szóbeli folyékonyosság, 156. o.	A1	marginális
	453	Olvasás információszerzés és érvek keresése céljából, 86. o.	A1	marginális
	487	Informális eszmecsere (barátokkal), 93. o.	C1	nincs C2, illetve. ua.
	519	Információcsere, 97. o.	B1+	+ szintek
	534	Feljegyzések, üzentek és formanyomtatványok, 100. o.	A1	marginális
	547	Együttműködés, 102. o.	A1	nincs leírás
	555	Magyarázatkerés, 103. o.	A1	nincs leírás
	574	A szókinccs alkalmazása, 135. o.	A1	nincs leírás
	585	Témakifejtés, 151. o.	A1	nincs leírás
Német (11)	621	Általános írásbeli produkció, 77. o.	A1	marginális
	770	Általános írásbeli interakció, 99. o.	A1	marginális
	801	Magyarázatkerés, 103. o.	A1	nincs leírás
	813	Általános nyelvi készségek, 132. o.	A1	marginális
	812	Általános nyelvi készségek, 132. o.	C1	marginális
	820	A szókinccs terjedelme, 134. o.	C2	marginális
	833	Grammatikai helyesség, 137. o.	C2	marginális
	836	Szociolingvisztikai megfelelés, 147. o.	C2	marginális
	845	Rugalmasság, 150. o.	C2	marginális
	850	Rugalmasság, 150. o.	A2	–
	862	Szóbeli folyékonyosság, 156. o.	C2	marginális

A 2. táblázatban a *nincs leírás* megjelölések a KER egy sajátosságára utalnak: nem található szintleírás minden skála minden szintjén. Azon túl, hogy ez a jelenség is elsősorban a szélső szinteket érinti, a vizsgáztatók a felkészülés során valószínűleg nem szenteltek figyelmet e sajátosságnak, pedig a KER szerzői röviden leírják a lehetséges okokat is (PTMIK, 2002. 47. o.). Ezek közül szakmailag a legfontosabb, hogy a hiányzó leírás jelzés lehet, hogy a skálán mért nyelvtudás-aspektus az adott szinten nem releváns. A szintleírás hiányának speciális esete, amikor adott (felső) szinten a szintleírás ugyanaz, mint az eggyel vagy kettővel lejjebb lévő szinten. Jó példája ennek a *Feljegyzések, üzenetek és formanyomtatványok skála* (100. o.). Bár több oka, értelmezése lehet annak, hogy az itt vizsgált rész-készség B1 szint felett nem épül tovább, feltételezhető, hogy a hatékony üzenetírás, formanyomtatványok kitöltése nem igényel B1-es szintnél magasabb tudást. A nyelvtudásról alkotott képet árnyalhatja, ha az ilyen skálán nem került sor az adott szintnél magasabb követelményeket támaztó szintleírás megfogalmazására. Összességében a 2. táblázatból az tűnik ki, hogy a szintleírások nem illeszkedése – leszámítva egyetlen német nyelvű szintleírást (Rugalmasság, A2) – viszonylag egyszerűen, a fordítás tüzetesebb vizsgálata nélkül megmagyarázható.

### A második szakasz eredményei

A második szakasz első eredménye az ET-skálához illesztett, nem vizsgarendszer-specifikus, hatfokozatú magyarországi skála (mérési karakterisztika) megalkotása volt. Főbb jellemzőit a 3. táblázat tartalmazza. Az egyes szinteket érintő válaszok, illetve azok százalékos megoszlása, ha nem is mutat kiegyensúlyozott képet, a szintek mérhetőségét alátámasztja, mert a válaszok száma minden szinten elégedőnek bizonyult pontos küszöbértékek számítására. A C2 szintre kevesebb válasz született, ennek megfelelően a mérési hiba is nagyobb. E szint illeszkedési értéke is magasabb, mint a kívánatos érték.

3. táblázat. A KER-szintek statisztikai jellemzői

Szintek	Válaszok száma db	Válaszok száma (%)	Illeszkedés (outfit)	Küszöbérték (step threshold)	Mérési hiba (SE)	Legnagyobb valószínűség (%)
C2	3090	8	1,6	5,95	0,03	100
C1	4833	12	0,8	3,56	0,02	62
B2	8553	21	1,0	0,31	0,02	72
B1	9424	24	0,9	-2,99	0,02	72
A2	9886	25	1,0	-6,82	0,02	77
A1	4292	11	1,2	-9,94	1,01 <sup>3</sup>	100

<sup>3</sup> Küszöbértéket és mérési hibát a szoftver (Linacre, 2006) a legalacsonyabb szinthez nem számít, ezért a táblázatban legalacsonyabb logit értéket kapott szintleírást szerepeltettem.

A szakasz másik fontos eredménye annak megválaszolása, hogy a vizsgáztatók mennyiben helyezték ugyanazokra a szintekre a 829 szintleírást. A KER hivatalos szintmeghatározásai és a magyar ítések szintmeghatározásai – biztosítva az összevethetőséget – nagyon hasonlóknak bizonyultak. A kétfajta besorolás között statisztikailag szignifikáns kapcsolat van a keresztábra (Khi-próba) szerint ( $\chi(25)=2546,538$   $p<0,01$ ), továbbá magas a kettő közötti korreláció ( $r=0,925$   $p<0,01$ ), amennyiben elfogadjuk, hogy a válaszok alkotta skála intervallumskálaként is felfogható. A vizsgálat során teljesült a keresztábra azon követelménye is, hogy minden leírásnak csak egy (1 db) besorolása lehetett, mind a KER, mind pedig a magyar válaszadók adatai szerint. Fontos eredmény volt továbbá a KER és a magyar skálák közötti kapcsolat erőssége is; az, hogy mennyiben (milyen biztonsággal) jósolható meg a szintleírások besorolása a magyar skálán. Ez a *PRE, proportional reduction of error* jelentős kapcsolatot mutat ( $\lambda=0,744$ ,  $p<0,01$ ), ami annyit tesz, hogy a hivatalos KER-szintbesorolás alapján a magyarországi besorolások 74%-ban megjósolhatók. Ha a nem illeszkedő 27 szintleírás nélkül (802 szintleírás) számítjuk a magyarországi szintbesorolások megjósolhatóságát, úgy az a 80%-os biztonságot is eléri.

A 4. táblázat segít értelmezni a két skála közötti erős kapcsolat mellett az eltéréseket is. Különösen az alsóbb szintek esetében, az eredeti KER-besorolások többsége a magyar adattáblában/elemzések eredményeképp is azonos besorolást kapott. Azonban a C1 és C2 szintek esetében már jelentős eltérések tapasztalhatók. A KER-ben eredetileg C1 szintre sorolt szintleírások jelentős részét a magyar ítések a B2 szintre sorolták, továbbá a C2-re sorolt szintleírások jelentős részét is a C1 és B2 szintekre lejjebb sorolták. Az eltérés magyarázata lehet például az, hogy a tanárok a B2 szinten tapasztalható, kiugróan magas létszámokhoz képest sokkal kevesebb C1-es diákkal, vizsgázóval találkozhatnak munkájuk során, de az is elképzelhető, hogy a C1 szintleírásaihoz gyakran megtévesztően hasonló B2+ szintleírások akadályozzák a C1 szint megfelelő koncipiálását.

4. táblázat. A magyar és a KER szerinti besorolások megoszlása hat szint esetén

Kalibrált magyar besorolások	A KER szerinti besorolások						Összes
	A1	A2	B1	B2	C1	C2	
A1	68	1					69
A2	28	188	1				217
B1		26	170	6			202
B2			20	155	37	4	216
C1					46	34	80
C2						45	45
Összes	96	215	191	161	83	83	829

A másik tendencia az, hogy az alsóbb szintek tekintetében bizonyos lágyság tapasztalható. Míg például a KER A1 szintre sorolt 96 szintleírásának többségét helyesen A1 szintre sorolták, addig ebből 28 szintleírást eggyel feljebb, az A2-es szintre. Hason-

lón lágý (*lenient*) tendencia érvényesül az A2 és B1 szintek esetében is, azonban e fölött a tendencia már a B2-nél megfordul és tovább erősödik a C1 és C2 szinteken. Ez utóbbi tendencia szigorúként értelmezhető, mint amikor azt mondjuk, hogy „a szintleírásban megjelenő nyelvi tevékenység nem C2 szintű, hanem annál gyengébb”.

A KER hivatalos besorolásai és a magyar ítések besorolásai közötti kapcsolat akkor is szignifikáns, ha nyelvek szerint szétbontjuk (5. táblázat). A szintleírásokat mindhárom nyelven nagyon hasonló módon sorolták be. A fentebb azonosított tendenciák (lágýság és szigor) szintén érvényesülnek mindhárom nyelv esetében. A hivatalos és a magyarországi besorolások közötti kapcsolat esetében a véletlen egyezések (az egyező besorolások véletlenszerűségének) esélye kisebb, mint 1% mind az angol ( $\chi(25)=977,076$   $p<0,01$ ), a magyar ( $\chi(25)=913,904$   $p<0,01$ ), mind a német ( $\chi(25)=710,335$ ,  $p<0,01$ ) nyelvű szintleírások esetében.

5. táblázat. A magyar és a KER szerinti besorolások megoszlása nyelv szerinti bontásban

Kalibrált magyar besorolások		KER szerinti besorolások						Összes
		A1	A2	B1	B2	C1	C2	
Angol	A1	24						24
	A2	7	67					74
	B1		4	63				67
	B2			2	52	9	1	64
	C1					18	10	28
	C2						16	16
Összes		31	71	65	52	27	27	273
Magyar	A1	24	1					25
	A2	10	64					74
	B1		8	55	2			65
	B2			9	54	10	1	74
	C1					20	9	29
	C2						19	19
Összes		34	73	64	56	30	29	286
Német	A1	20						20
	A2	11	57	1				69
	B1		14	52	4			70
	B2			9	49	18	2	78
	C1					8	15	23
	C2						10	10
Összes		31	71	62	53	26	27	270



Besorolási eltérések, hibák mindhárom nyelv esetében megfigyelhetők, azonban a hibás besorolások nyelvenkénti eltérései már szignifikáns különbségeket mutatnak. Legkevesebb (33) ilyen eltérés az angol leírásokban fordult elő. A magyar leírások között már több található (50), míg a legtöbb a német leírásoknál volt megfigyelhető (74). Ezek a különbségek a *one-way Chi-square* eljárással voltak ellenőrizhetők, választ keresve arra a kérdésre, hogy a hibás besorolások 33-50-74 arányban való tagolódása lehet-e a véletlen műve. A megfigyelt különbségek túl nagyok voltak ahhoz, hogy csupán a véletlen művének tulajdonítsuk ( $\chi(2)=16,217, p<0,01$ ).

Az adattábla részbeni átkódolását követelte meg a válaszok kilenc szint szerinti besorolása. Az A2, B1 és B2 + szintjeire vonatkozó információ a KER mindhárom nyelvi változatában fellelhető (*Council of Europe, 2001; Europarat, 2001; PTMIK, 2002*), ha nem is elég explicit formában, mert a + szintekre sorolt leírásokat csak egy vonal választja el a normál szintleírásoktól a KER-skálákban: a vonal felett látható leírások a + szint jellemzői. Bár az illesztési projektek során nem volt követelmény az, hogy az ítések meg tudják különböztetni a + szinteket, szerepeltetésük több betekintést nyújt.

A kilenc szint szerinti besorolásnál egyrészt a tendenciák ismétlődnek: egyrészt a légység átfordul szigorba (6. táblázat), másrészt a válaszadók mintha tudatában lettek volna, milyen mértékű légység vagy szigor kerüli el azt, hogy a szintleírás már egy másik szintre kerüljön át. Például az A2-es szintleírások többségét ugyan A2+-re sorolták, azonban az A2+ szintleírások többsége nem került át B1 szintre. Hasonló tendencia figyelhető meg a B1/B1+ szint esetében is, bár itt már több a besorolási hiba. Összesen 20 szintleírást minősítettek B2 szintűnek. A B2/B2+ szint leírásainak besorolása jól sikerült, alig van hiba. Végül a C1 és C2 szintek esetében a keresztábra az elemzőben megerősíti a szigor benyomását, ami az előző tábla alapján keletkezett. E mellett az is látszik, hogy a B2 szintre tett C1-es leírásokat egy kivételével a B2+ szintre helyezték.

6. táblázat. A magyar és a KER szerinti besorolások megoszlása kilenc szint esetén

Kalibrált magyar besorolások	KER szerinti besorolások									Összes
	A1	A2	A2+	B1	B1+	B2	B2+	C1	C2	
A1	68	1								69
A2	27	40	7							74
A2+	1	76	65	1						143
B1		4	20	49	6					79
B1+			2	70	45	6				123
B2				6	14	84	20	1		125
B2+						19	32	36	4	91
C1								46	34	80
C2									45	45
Összes	96	121	94	126	65	109	52	83	83	829

## Összegzés és a tanulmány jelentősége

A KER magyar nyelvű fordításának mint a nyelvvizsgák illesztésére használt eszköznek az érvényessége alátámasztható az illesztés során felvett válaszadatok empirikus feldolgozása alapján. A kutatás eredményei megerősítik a KER magyar változatával, pontosabban magyarra fordított skáláival kapcsolatban megfogalmazható tételt, miszerint a skálák alkalmasak a KER szerinti illesztés elvégzésére. A magyar nyelvű szintleírások döntő mértékben együtt mozognak, a magyar kollégák besorolásai nagymértékben hasonlóak olyan más szintleírásokhoz (angol és német), melyeknek érvényességét korábban több bizonyítékkal jobban megalapozták. Ez a közös nevező azt jelentette, hogy leszámítva a kizárt, leggyengébben teljesítő ítéseket, mind az ítések KER ismerete szintjének eltérései, mind a kérdőív kitöltésének körülményei, mind a besorolt szintleírások nyelve semlegesítve lett.

Mindemellett igaz az, hogy a KER magyar fordításának érvényességét szükséges még kvalitatív alapon is vizsgálni. Biztosak lehetünk abban is, hogy a magyar fordítás érvényessége kérdéséhez még a fordítás avatott szakértői is sok mindent hozzátehetnek. Ennek a kutatásnak az eredménye az, hogy egy lehetséges alternatív magyarázattal kevessebbel kell szembenézni: nem valószínű, hogy az illesztési eljárások eredményeit befolyásolta volna, hogy a KER magyar nyelvű változata csak egy lektorált fordítás. Arra is lehet következtetni, hogy az érintett vizsgaközpontok ítései a magyarázott KER szintleírásai alapján sikerrel alakítottak ki egy olyan közös szemléletmódot, ami megfeleltethető a KER idegen nyelvű szintleírásaiban tükröződő felfogással.

A vizsgaközpontok adattábláinak összefűzésének jelentősége abban áll, hogy így a jövőben, további adattáblák bekapcsolása esetén, még létrejöhet egy olyan magyar skála, amely a legtöbb magyarországi nyelvvizsga adatait figyelembe veszi, és amely alkalmas mind feladatsorok, mind vizsgázoí teljesítmények illesztésére. Az lenne az ideális, ha a magyarországi nyelvvizsgák illesztett KER követelményei átfogóak lennének, egy általánosan elfogadott, közmegegyezésen alapuló, illesztett „magyar” standardot jelenítené meg a saját vizsgában. Az itt feldolgozott válaszadatok többsége illeszthető a KER-hez, bár nem képviselik az összes olyan magyar vizsgaközpontot, ahol a magyar nyelvű KER-t használják, még csak nem is azok többségét. Mégis, ha meg akarjuk nevezni, milyen skáláról van szó, a válasz az lehet, hogy a skála Magyarországon működő nyelvvizsgák ítéseinek meglátásait tükrözi. Ennyiben tehát egy magyar KER-skáláról van szó.

## Irodalom

- Bachman, L. F. (1990): *Fundamental considerations in language testing*. Oxford University Press, Oxford.
- Bachman, L. F. és Palmer, A. S. (1982): The construct validation of some components of communicative language ability. *TESOL Quarterly*, 16. 4. sz. 449–465.
- Bárdos Jenő (2002): *Az idegennyelvi mérés és értékelés elmélete és gyakorlata*. Nemzeti Tankönyvkiadó, Budapest.

A szintleírások nyelvének szerepe a közös európai referenciakeret magyar, angol és német nyelvű kiadásában

- Borsboom, D., Mellenbergh, G. J. és van Heerden, J. (2004): The concept of validity. *Psychological review*, **111**. 4. sz. 1061–1071.
- Canale, M. és Swain, M. (1980): Theoretical bases of communicative approaches to language learning and testing. *Applied Linguistics*, **1**. 1. sz. 1–47.
- Council of Europe (2001): *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press, Cambridge.
- Council of Europe (2003): *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment. Manual, Preliminary Pilot Version*. The Council of Europe, Strasbourg.
- Council of Europe (2004): *Reference supplement to the preliminary pilot version of the manual for relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment*. Language Policy Division, Strasbourg.
- Council of Europe (é. n.): 38 language versions of the CEFR. Letölthető: [http://www.coe.int/t/dg4/linguistic/Publications\\_EN.asp#TopOfPage](http://www.coe.int/t/dg4/linguistic/Publications_EN.asp#TopOfPage).
- Cureton, E. E. (1951): Validity. In: Lindquist, E. F. (szerk.): *Educational measurement*. American Council on Education, Washington, DC. 621–694.
- Csapó Benő (1993): Tudásszintmérő tesztek. In: Falus Iván (szerk.): *Bevezetés a pedagógiai kutatás módszereibe*. Keraban kiadó, Budapest. 277–317.
- Dávid, G. (2007): Investigating the performance of alternative types of grammar items. *Language testing*, **24**. 1. sz. 65–97.
- Dávid Gergely (2008): Az emelt szintű érettségi és az államilag elismert nyelvvizsgák a vizsgázói teljesítmények tükrében. Összegző tanulmány. Kézirat. Nyelvvizsgát Akkreditáló Testület. PH Nyelvvizsgázatási Akkreditációs Központ. <http://www.euroexam.org/index.php?page=590&l=2>
- Europarat (2001): *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Langenscheidt, Berlin, München.
- Horváth György (1993): *Bevezetés a tesztelméletbe*. Keraban Kiadó, Budapest.
- Horváth György (1997): *A modern tesztmodellek alkalmazása*. Akadémiai Kiadó, Budapest.
- Kane, M. (2012): Validating score interpretations and uses: Messick lecture. The language testing research colloquium. Cambridge, April 2010. *Language testing*, **29**. 1. sz. 3–17.
- Kelley, T. L. (1927): *Interpretation of educational measurements*. Macmillan, New York.
- Linacre, J. M. (1994): *Many-facet Rasch measurement*. Mesa Press, Chicago.
- Linacre, J. M. (1998): *A user's guide to facets and facform*. [Software manual] Mesa Press, Chicago.
- Linacre, J. M. (2006): Facets: Rasch measurement computer program. Version 3.59 [Computer software] Mesa Press, Chicago.
- Messick, S. (1989): Validity. In: Linn, R. L. (szerk.): *Educational measurement*. American Council on Education/Macmillan, New York. 13–103.
- Messick, S. (1995): Validity of psychological assessment. *American Psychologist*, **50**. 9. sz. 741–749.
- Milanovic, M. és Weir, C. J. (2010, szerk.): *Studies in language testing. 15<sup>th</sup> Anniversary edition*. University of Cambridge ESOL Examinations, Cambridge.
- Molnár Gyöngyvér (2003): Az ismeretek alkalmazásának vizsgálata modern tesztelméleti eszközökkel. *Magyar Pedagógia*, **103**. 4. sz. 423–446.
- Molnár Gyöngyvér (2005): Az objektív mérés megvalósításának lehetősége: a Rasch-modell. *Iskolakultúra*, **15**. 3. sz. 71–80.
- Molnár Gyöngyvér (2006): A Rasch-modell alkalmazása a társadalomtudományi kutatásokban. *Iskolakultúra*, **16**. 12. sz. 99–113.

- Molnár Gyöngyvér (2008): A Rasch-modell kiterjesztése nem dichotóm adatok elemzése: a rangskálás és parciális kredit modell. *Iskolakultúra*, **18**. 1–2. sz. 66–77.
- North, B. (2000): *The development of a common framework scale of language proficiency*. Peter Lang Publishing, New York.
- North, B. (2004): Relating assessments, examinations and courses to the CEF. In: Morrow, K (szerk.): *Insights from the Common European Framework*. Oxford University Press, Oxford.
- North, B. és Schneider, G. (1998): Scaling descriptors for language proficiency scales. *Language testing*, **15**. 2. sz. 217–263.
- Popham, W. J. (1997): Consequential validity: Right concern – wrong concept. *Educational Measurement: Issues and Practice*, **16**. 2. sz. 9–13.
- PTMIK (2002): *Közös európai referenciakeret. Nyelvtanulás, nyelvtanítás, értékelés*. Pedagógus-továbbképzési Módszertani és Információs Központ Kht., Budapest.
- Rasch, G. (1961): On general laws and the meaning of measurement psychology. In: Neyman, J. (szerk.): *Proceedings of the Fourth Berkeley symposium on mathematical statistics and probability. Vol. 4.*, University of California Press, Berkeley. 321–334.
- Schneider, G. és North, B. (2000): *Fremdsprachen können - was heisst das? Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit*. Verlag Rüegger AG, Chur/Zürich.
- TESOL (é. n.): TESOL Quantitative research guidelines. [On-line] Available: <http://www.tesol.org/read-and-publish/journals/tesol-quarterly/contact-tq-editorial-staff/tesol-quarterly-submission-guidelines/because-tq-reports-sub>.

## ABSTRACT

GERGELY DÁVID: THE ROLE OF LANGUAGE IN LEVEL DESCRIPTORS IN THE HUNGARIAN-, ENGLISH- AND GERMAN-LANGUAGE EDITIONS OF THE COMMON EUROPEAN FRAMEWORK OF REFERENCE

This study deals with the validity of the Hungarian translation of the Common European Framework of Reference (CEFR) with a special focus on the Hungarian version of the illustrative scales. The scales have special relevance in a validity study as they operationalise the concepts of language proficiency for a wider audience and a variety of uses. Based on the English-language version, the Hungarian translation was the starting point for a number of CEFR linking projects in 2007, following the Council of Europe initiative and subsequent Hungarian legislation to establish a link between language proficiency tests and the CEFR. The linking scheme was pushed through, even though the Hungarian translation only involved a handful of professionals and was not validated at all. Further, the linking projects operated on the basis of different language versions of the CEFR. In this project, thus, quantitative data were collected in three languages (English, Hungarian and German) from 154 CEFR linking project judges by three language test producers based in Hungary. The research questions concerned the extent to which the validity of the Hungarian scales could be supported with appropriate validity evidence, whether the Hungarian scales could be linked to the CEFR scales in English and German and whether rival hypotheses (alternative explanations) could be rejected. These rival hypotheses included the effect of different languages (translation), the raters' severity and the conditions of filling in the questionnaire. The data consisted of ratings of the CEFR level (A1, A2, etc.) of 829 descriptors from the illustrative scales. These data were drawn from questionnaires originally meant as a survey of the judges' competence in the CEFR. On the basis of the ratings, the descriptors, including the Hungarian translations, were scaled and classified, and the results were compared to the original CEFR calibrations-based level classifications. The results confirmed the presence of a language-of-the-scales facet. Overall, they show a good degree of fit between the way Hungarian judges rated the descriptors in the three languages and the official CEFR level categories. In comparison with CEFR classifications, Hungarian judges tended to be lenient concerning requirements at levels A1 to B2 and strict concerning levels C1 to C2.

Magyar Pedagógia, 112. Number 1. 19–39. (2012)

Levelezési cím / Address for correspondence: ELTE, Angol-Amerikai Intézet, Angol Nyelvpedagógia Tanszék, H-1088 Budapest, Rákóczi út 5.