

KISZELY ZOLTÁN

Egy online vizsgáztató- továbbképzés tapasztalatai

Lessons from an online standardisation training for examiners.

This study discusses the methods, results and conclusions of an online standardisation training for examiners organised by the BME's Language Examination Centre. The training consisted of two parts: in the first, the examiners had to complete a test related to their knowledge of the levels of the CEFR (Common European Framework of Reference), while in the second they had to evaluate an audio file recording of an oral language performance. The aim of the study is twofold: first, to provide a detailed account of the CEFR test results using both classical and modern test theories, and second to draw some conclusions from the evaluation task -- though only from a bird's eye view, as some of the data collected were sensitive and classified, and therefore cannot be published by the language examination centre. However, these conclusions shed light on the examiners' preferences and attitude to language, which will be of interest to a wider audience of language teachers.

Bevezetés

Az idegennyelv-tudást igazoló, államilag elismert nyelvvizsgáztatás rendjéről és a nyelvvizsga bizonyítványokról szóló 137/2008. (V. 16.) Kormányrendelet előírja, hogy minden, Magyarországon államilag elismert nyelvvizsga bizonyítványt kiadó intézménynek évente vizsgáztató-továbbképzést kell tartania az értékelői számára a standardok folyamatos fenntartása érdekében. A 2017-es évben ezt a továbbképzést a BME Nyelvvizsgaközpont online módon oldotta meg. Ez az online képzés két fázisból állt. Az első egy magyar nyelvű KER (Közös Európai Referenciakeret 2002) familiarizációs teszt, melynek célja a szintrendszer ismeretének felfrissítése, a második pedig vizsgázói beszédkészség teljesítményminták értékelése volt angol, német, francia, spanyol és olasz nyelvekből.

Az alábbi tanulmányban a szóbeli értékelők számára szervezett továbbképzés módszereit, eredményeit és legfontosabb tanulságait ismertetjük. Az első részben részletesen leírjuk a KER familiarizációs teszt módszereit, eredményeit és tanulságait, és ehhez az elemzéshez felhasználjuk mind a klasszikus, mind pedig a modern tesztelmélet eszköztárát. Ez a rész a KER egyfajta speciális értelmezésének is felfogható, aminek különös aktualitást ad az a tény is, hogy 2017 szeptemberében megjelent a KER továbbfejlesztett változata (North— Goodier—Piccardo 2017). A második részben röviden, a részletek mellőzésével, a tendenciákra koncentrálva ismertetjük a szóbeli teljesítményminták értékelésének tanulságait. Az elemzés a fentiekből következően kicsit aránytalan lesz, ugyanis a szóbeli teljesítményminták értékelésének részletes eredményei érzékeny adatokat is tartalmaznak, amelyeket egy működő vizsgaközpont

nem adhat ki, azonban az említett tendenciák és összefüggések olyan információkkal szolgálnak az értékelők preferenciáiról és nyelvszemléletéről, amely a nyelvtanárok szélesebb köre számára is tanulságos lehet. A tanulmányban a *vizsgáztató* és az *értékelő* szavakat szinonimaként fogjuk használni.

Módszerek

A továbbképzésre 2017 novemberében került sor: a vizsgáztatóknak a vizsgaközpont által összeállított két online feladatlapot kellett kitölteniük. Az első rész egy háromopciós feleletválasztós tesztet tartalmazott, amelyben a vizsgáztatóknak a KER különböző skáláiból származó 35 deskriptorról (szintleírásról, itemről) kellett eldönteniük, hogy ezek a három vizsgáztatott szint közül (B1, B2, C1) melyhez tartoznak. Jelentős könnyítés volt a feladatban, hogy sem a + szinteket leíró (A2+, B1+, B2+), sem pedig a nem vizsgáztatott szinteket leíró (A1, A2, C2) deskriptorok nem szerepeltek a feladatlapban. Ennek az volt az oka, hogy egyrészt ne nehezítsük feleslegesen a vizsgáztatók dolgát, másrészt pedig a KER deskriptorok közismert egyenetlenségei (Dávid 2012) felesleges terhet jelentettek volna a kitöltésnél.

Két további engedményt is tettünk a vizsgáztatóknak. Az egyik az volt, hogy használhatták a KER-t a kitöltéshez, ugyanis mivel online képzésről van szó, a kitöltés önállóságát a vizsgaközpont úgyszemint tudta volna ellenőrizni, másrészt pedig a KER olyan olvasmány, amit még akkor is érdemes forgatni, ha valaki azt gondolja, hogy már teljesen ismeri. A KER használatát még annyiban is megkönnyítettük, hogy belinkeltük a KER magyar nyelvű weboldalát, és arra is felhívtuk a figyelmet, hogy a részletes skálák a 4. és az 5. fejezetben találhatóak.

A feladatlapot korlátlan számú alkalommal lehetett kitölteni, azaz rengeteg alkalom nyílt a gyakorlásra még azután is, ha valaki már sikeresen megoldotta a feladatokat, ugyanis a program a deskriptorokat minden egyes új kitöltés alkalmával eltérő sorrendben jelenítette meg. A sikeres feladatmegoldás azt jelentette, hogy a vizsgáztató kollégáknak legalább 80%-os teljesítményt kellett nyújtaniuk, azaz 28 helyes megoldást kellett adniuk. A többszöri próbálkozások közül mindig a legjobbat vettük figyelembe. A feladatlap kitöltésére hozzávetőleg 30 percet irányoztunk elő, de szigorú időkorlátot nem iktattunk a rendszerbe. Az online feladatlapot két hétig tartottuk nyitva a vizsgáztatók számára, akiknek ennyi idő állt rendelkezésére a feladatok megoldására. A megoldások beküldése után, a rendszer azonnali visszajelzést adott a résztvevőknek az elért összpontszámukról, így mindenki tisztában lehetett azzal, hogy megfelelő-e az eredménye.

A KER feladatlap sikeres, azaz minimum 80%-os megoldása előfeltétele volt a második továbbképzési rész elvégzésének, amely B2 szintű beszédkészséget mérő vizsgafeladatokhoz tartozó vizsgázói teljesítmények értékelése volt. A feladatlapon, nyelvtől függően, egy próbavizsga vagy egy éles vizsga hangfelvétele szerepelt, amely annyiban különbözött az igazi éles vizsgától, hogy mivel audio anyagról volt szó, sem a vizsgázót sem a vizsgáztatókat nem lehetett látni. A feladatlapon a vizsgán (azaz a hangzó anyagban) felhasznált feladatokat és az értékelési kritériumokat is feltüntettük, így a hallgatás közben a feladatokat és az értékelési szempontokat is egyszerre lehetett tanulmányozni. A feladat teljesítése akkor volt sikeresnek tekinthető, ha a vizsgáztató kolléga maximum +/-4 ponttal tért el a standard összpontszámtól, amelyet egy szakértői csoport korábban meghatározott. Ha például az egyik nyelv

esetében ez a pontszám 40 volt, akkor a 36–44 közötti pontok bizonyultak sikeresnek. A +/-4 pont megállapítását nem előzte meg empirikus validálás; ezt a határolóértéket a vizsgaközpont nyelvi felelőseinek konszenzusa alakította ki.

A feladatlapot csupán egy alkalommal lehetett kitölteni, hiszen az éles vizsgán is csak egyszer értékel a vizsgáztató, de a hallgatás közben bármikor meg lehetett állítani a felvételt, és visszajátszható volt igény szerint. A feladatlap kitöltésére hozzávetőleg 25 percet irányoztunk elő, de szigorú időkorlátot itt sem iktattunk a rendszerbe. Ezt a feladatlapot is két hétig tartottuk nyitva a vizsgáztatók számára, akiknek ennyi idő állt rendelkezésére a megoldáshoz. A határidő lejárta után a rendszer adminisztrátora elküldte a vizsgáztatók által adott összpontszámot a résztvevők számára egy kísérőlevéllel együtt. Utóbbiban felhívtuk a figyelmet arra, hogy akinek az értékelése a +/-4 ponttól nagyobb mértékben tér el a standardtól, a továbbiakban is vizsgáztathat ugyan, de a következő évben a vizsgaközpont vezető szakemberei hangsúlyosabban fogják nyomon követni vizsgáztatói munkáját.

KER feladatlap

A 35 íte mből álló KER feladatlapot a következő KER skálák 12 B1, 12 B2 és 11 C1 szintű deskriptoraiból állítottuk össze: Általános beszédprodukció, Általános szóbeli interakció, *A szó átvétele*, *A szókincs terjedelme*, *A szókincs alkalmazása*, *Folyamatos monológ: Élménybeszámoló*, *Fonológiai tudás*, *Grammatikai helyesség*, *Információcsere*, *Koherencia és kohézió*, *Önellenzés és önkorrrekció*, *Szóbeli folyékonyág*, *Társalgás és Tartalmi pontosság*. A feladatlapot minden vizsgáztató e-mailben kapta meg, és két hét áll rendelkezésére a kitöltésre.

A KER feladatlapot a többszáz vizsgáztató összesen 610 alkalommal töltötte ki. Néhányan az első kitöltéskor valószínűleg még nem fordultak a KER-hez segítségért, hanem próbálták logikai alapon, korábbi ismereteiket és tapasztalataikat felhasználva megoldani a tesztet. Mások már az első kitöltés alkalmával elérték a 35 pontos maximális eredményt, ami azt jelenti, hogy vagy remekül ismerték a skálákat, vagy kitöltés közben gondosan tanulmányozták a KER-t. Az a jelenség is tapasztalható volt, hogy néhányan a megfelelési szintet jelentő 28 pont elérése után is még kitöltötték néhány-szor a tesztet vagy a minél jobb eredmény elérése, vagy a további gyakorlás kedvéért.

KER feladatlap – Klasszikus tesztelmélet: leíró statisztika

A leíró statisztika adatainak elemzése előtt néhány dolgot fontos leszögezni. A feladatokat először a klasszikus tesztelemzés (KTE) módszereivel vizsgáltuk. A KTE abból indul ki, hogy a megfigyelt pontérték a valódi pontérték és a mérési hiba összegéből adódik, ezért a célja az, hogy a mérési hiba minél kisebb legyen, mert annál közelebb kerül egymáshoz a megfigyelt pontérték és a valódi pontérték. A KTE alapján egyrészt (al)tesztszinten, másrészt pedig itemszinten lehet elemzéseket végezni. A teszt vagy tesztrész szintjén a legfontosabbak a teszt megbízhatósága, a mérési hiba és a leíró statisztika alapvető adatai (pl. átlag, módusz, medián). Az itemek szintjén az item nehézsége és az itemdiskrimináció a központi fogalmak.

Fontos megemlíteni továbbá, hogy a vizsgáztató-továbbképzés jellege miatt a KER feladatlap esetében nem lehet ugyanazokat a statisztikai adatokat megfelelőnek tekinteni, amelyeket egy nyelvvizsgateszt feladatai esetében. Jelen esetben ugyanis pél-

dául az itemek diszkriminációs indexe nem annyira fontos, hiszen egyrészt a KER deskriptorokon nem lehet javítani, ha esetleg rosszul diszkriminálnak, míg a rosszul működő vizsgafeladat esetében erre van mód. Másrészt pedig ez nem készségszintet vizsgáló teszt (proficiency), hanem egy teljesítményteszt (achievement vagy attainment), amelynek a megoldása során azt várjuk, hogy minél jobb eredményeket érjenek el a vizsgázók, hiszen csupán egy adott tananyag számonkérése történik. Ebben az esetben pedig az a jó, hogy ha a nehézségi érték magas, azaz könnyű a teszt a legtöbb vizsgázó számára, ami azt jelenti, hogy a vizsgázók nagy része a válaszok többségét tudja. Ebből pedig az következik, hogy a diszkrimináció alacsony, hiszen, ha nagyon sokan érnek el nagyon jó eredményt, akkor az itemek kevésbé tesznek különbséget (diszkriminálnak) a jobban, illetve a gyengébben teljesítők között (Green 2013).

A kvantitatív elemzés során mindig a válaszok (kitöltések) számáról és nem a válaszadók számáról beszélünk. Az elemzés során a *deskriptor* és az *item* szavakat szinonimaként fogjuk használni.

A feladatlap leíró statisztikai adatait az IteMan program segítségével elemeztük. Az adatokat kétféle módon közöljük. Az 1. táblázatban a feladatsoronkénti összesítést adjuk meg, azaz a B1, a B2 és a C1 deskriptorokat külön feladatként kezelve tesszük közzé az eredményeket, a 3. táblázatban pedig az eredményeket összesítve mutatjuk.

Szintek	B1	B2	C1
Itemek száma	12	12	11
Vizsgázók száma	610	610	610
Átlag	10.256	9.472	8.557
Variancia	4.135	4.866	5.152
Szórás	2.033	2.206	2.270
Ferdeség	-1.449	-0.769	-0.912
Csúcsosság	2.043	-0.047	0.221
Minimum	0.000	2.000	0.000
Maximum	12.000	12.000	11.000
Medián	11.000	10.000	9.000
Megbízhatóság (alfa)	0.739	0.669	0.725
Mérési hiba	1.039	1.270	1.190
Átlagos nehézségi érték	0.855	0.789	0.778
Átlagos item-összes diszkrimináció	0.498	0.459	0.515
Átlagos biszeriális diszkrimináció	0.830	0.666	0.757
Legjobb eredmény a gyenge csoportban	10	8	7
A gyenge csoport létszáma	250	190	173
Legrosszabb eredmény a jó csoportban	12	11	10
A jó csoport létszáma	218	243	266

1. táblázat. Leíró statisztika szintenként

Az 1. táblázatból kiderül, hogy a résztvevők számára a legkönnyebb a B1 deskriptorok megoldása volt (Mean P: 0,855, ami 85,5%-ot jelent). Ez nem véletlen, hiszen a nyelvtudásban a minőségi ugrás a B1 és a felette álló szintek között van, azaz a B2 és a C1 szinteket leíró deskriptorok könnyebben összekeverhetőek, tehát a B1 jobban elkülö-

nül a másik kettő szinttől. Érdekes továbbá, hogy a vizsgáztatók a B2 deskriptorokat, ugyan csak kismértékben, de mégis sikeresebben sorolták be, mint a C1 deskriptorokat. Ezt talán az magyarázza, hogy a B2 szinten a vizsgáztatók sokkal gyakrabban vizsgáztatnak, következésképpen a hozzátartozó KER skálákat is jobban ismerik. Ezt a fejtegetést az is alátámasztja, hogy az eloszlásfüggetlen páros Wilcoxon-próba (2. táblázat) (amelyet akkor alkalmazunk, ha az eredmények nem normál eloszlásúak, mint jelen esetben) mindhárom összehasonlítás esetében statisztikailag szignifikáns különbséget talált az átlageredmények között, ami azt jelenti, hogy látszólag bármilyen kicsi is a különbség a tesztrészek átlagai között, ezeket a különbségeket mégsem a véletlen okozta, azaz valós különbségeket jelentenek.

	B2 - B1	C1 - B1	C1 - B2
Z	-9,425 ^a	-15,531 ^a	-9,956 ^a
Szignifikancia szint	,000	,000	,000

2. táblázat. A Wilcoxon-próba eredményei

A feladatonkénti alacsony elemszám (12, 12, 11) miatt több értéket a feladatonkénti összevetésben nem vizsgálunk (pl. Alpha, SEM, Std. Dev.), amelyeket a teljes feladatsor esetén érdemes elemzés alá vetni.

Itemek száma	35
Vizsgálók száma	610
Átlag	28.285
Variancia	29.007
Szórás	5.386
Ferdeség	-0.633
Csúcsosság	-0.264
Minimum	9.000
Maximum	35.000
Medián	29.000
Megbízhatóság (alfa)	0.853
Mérési hiba	2.066
Átlagos nehézségi érték	0.808
Átlagos item-összes diszkrimináció	0.400
Átlagos biszeriális diszkrimináció	0.608
Legjobb eredmény a gyenge csoportban	25
A gyenge csoport létszáma	192
Legrosszabb eredmény a jó csoportban	33
A jó csoport létszáma	181

3. táblázat. Leíró statisztika a teljes tesztre vonatkozóan

A 3. táblázatban az első fontos adat a teljes teszt megbízhatóságára vonatkozó Cronbach alfa érték (Alpha), ami szórásalapú mutató, és a teszt belső konzisztenciáját méri. Ez az érték 0 és 1 között mozoghat: minél magasabb, annál megbízhatóbb a teszt. Fontos

tudni, hogy ha egy tesztben az eredmények normál eloszlásúak, akkor valószínűbb a magas alfa érték. Ha egy teszt nem normál eloszlású, hanem pozitív vagy negatív irányba hosszabban elnyúlik, akkor alacsonyabb alfa érték várható (Brown 2002). Jelen KER-teszt esetében, ahogy azt később látni fogjuk, nem beszélhetünk normál eloszlásról, következésképpen a 0,853 értékű alfa megfelelőnek mondható.

Egy másik fontos érték az átlagpontszám, amely az elvárt módon alakult (mean: 28,285), hiszen ez a teljesítési küszöbként meghatározott 80%-nak megfelelő pontszám (mean p: 0,808). A szórás értéke (Std. Dev.), azaz az átlagtól való átlagos eltérés mértéke is figyelemre méltó (5.386), ami azt jelenti, hogy az értékelők átlagosan 28,285 \pm 5,386 pontok között adták a pontszámokat, ami gyakorlatilag 23 és 34 pont közé tehető.

A mérési hiba (SEM) azt mutatja meg, hogy a megfigyelt (egyéni) pontérték milyen mértékben tér el a valódi pontértéktől, azaz mennyire bízhatunk meg a kapott eredményekben. Esetünkben ez az érték 2.066, ami azt jelenti, hogy az egyes vizsgázatók által elért megfigyelt pontérték \pm 2,066 ponttal térhet el a valódi pontértéktől. Ha tehát valaki például éppen 28 pontot ért el, ami a megfigyelt pontérték, akkor az ő valódi tudása 28 \pm 2,066 között lehet, ami gyakorlatilag 26 és 30 pont közé tehető.

A medián (közéérték), szintén fontos jelzőszám. Ez a sorba rendezett eredmények közepén elhelyezkedő érték, ami azt jelenti, hogy ettől az értéktől jobbra is és balra is ugyanannyi elemszám található. Esetünkben a medián értéke 29, ami a 35 ponthoz közeli. Ez azt jelenti, hogy a 80% feletti teljesítményt ugyanannyi kitöltés eredményezte, mint a 80% alatti.

KER feladatlap – Klasszikus tesztelmélet: itemszintű statisztika

Az itemszintű statisztika esetén a legfontosabb adat a nehézségi érték, ami azt mutatja, hogy a vizsgázatók milyen arányban adtak helyes választ az adott itemre. A két szélső-érték a 0% és a 100%, ami azt jelenti, hogy minél magasabb az érték, annál könnyebb az item. Ha az érték pl. 0%, akkor senki sem tudta megoldani az itemet, ha pedig 100%, akkor mindenki. A 4. táblázat azt tartalmazza, hogy az egyes nehézségiszint-sávokba hány deskriptor esett, és ezeknek mi volt a megoszlása a három szint között.

%	B1	B2	C1	Összesen
90–97	4	1	2	7
80–89	6	7	3	16
70–79	1	1	2	4
60–67	0	3	4	7
–56	1	0	0	1
	12	12	11	35

4. táblázat. Deskriptorok megoszlása nehézségi szintek szerint

A leíró statisztikák alapján megállapíthattuk, hogy a B1 szintű deskriptorokat találták el a legnagyobb százalékban a vizsgázatók (85,5%). Ezt a képet az itemszintű statisztika azonban árnyalja, ugyanis a legkönnyebb sávba (90-97%) csak négy B1 szintű item került a 12-ből, ráadásul a legnehezebbnek bizonyult item is egy B1 szintű volt (56%).

Ha a két legfelső sávot, azaz a két legkönnyebbnek bizonyult itemsávot tekintjük, akkor az idetartozó B1 itemek száma 10 (4+6), a B2 szintű itemek száma 8 (1+7), a C1 szintű itemek száma pedig 5 (2+3). Ezek az értékek pedig már alátámasztják a leíró statisztika adatait, mely szerint a B1-hez tartozó itemek a legkönnyebbek, a C1 szintű itemek pedig a legnehezebbek.

Az 5. táblázat a legkönnyebbnek bizonyult deskriptorokat tartalmazza. Az első oszlop a sorszámot, a második a szintet, a harmadik a nehézségi értéket mutatja százalékban, a negyedik oszlopban pedig maga az item, a KER-ből származó deskriptor található.

Sor-szám	Szint	Nehézségi érték	Deskriptor (item)
10.	B1	97%	Az előreláthatóbb helyzetekkel kapcsolatos gyakran használt „begyakorolt mondatok” és minták készletét elfogadható nyelvhelyességgel alkalmazza.
11.	B1	97%	Elegendő szókincse van ahhoz, hogy némi körülírással ki tudja magát fejezni a legtöbb olyan témában, amely összefügg saját mindennapi életével, mint például a család, a hobbi és az érdeklődési kör, a munka, az utazás és az aktuális események.
7.	B1	96%	Több rövidebb, különálló, egyszerű elemet össze tud kapcsolni összefüggő lineáris szempontsorozattá.
5.	B1	94%	Át tud adni, egyszerű, tárgyyszerű információkat, és meg tudja értetni, hogy mely pontokat érzi a legfontosabbnak. Mondanivalójának lényegét érthetően ki tudja fejezni.
33.	C1	94%	A grammatikai helyesség magas szintjét következesen fenn tudja tartani; hibák ritkán fordulnak elő, és nehezen észrevehetőek.
35.	C1	92%	Jól használja széles körű szókincsét, az esetleges hiányokat körülírásokkal könnyedén áthidalja; ritkán kell keresgélnie a kifejezéseket vagy elkerülési stratégiát alkalmaznia. Jól tudja használni az idiomatikus kifejezéseket és kollokvializmusokat.
21.	B2	91%	Viszonylag magas szintű grammatikai biztonsággal kezeli a nyelvet. Nem követ el olyan hibákat, amelyek félreértésekhez vezetnek.

5. táblázat. A legkönnyebbnek bizonyult itemek

A legkönnyebbnek bizonyult négy B1 szintű item megfogalmazása is sugallja, hogy felsőbb szintekre nehezen lehetne sorolni őket, hiszen olyan jellemző hívószavak és kifejezések jellemzik őket, mint például a „előreláthatóbb”, „begyakorolt mondatok”, „rövidebb, különálló, egyszerű” és „közvetlen fontosságú”.

Érdekes megfigyelés, hogy a legkönnyebb hét item közül öt a *Grammatikai helyesség* és a *Szókincs terjedelme* skálák része volt (10, 11, 33, 35, 21). Ennek valószínűsíthetően két oka van: az egyik az a megfogalmazás pontossága és egyértelműsége lehet, a másik pedig az, hogy a vizsgáztatók a nyelvtan és a szókincs szintjének a megítélésében a legmagabiztosabbak.

A 6. táblázat a legnehezebbnek bizonyult itemeket tartalmazza, amely csupán annyiban különbözik az 5. táblázattól, hogy beiktattunk egy ötödik oszlopot, ami azt tartalmazza, hogy a válaszok hogyan oszlottak meg az egyes opciók között (diszt-

raktorelemzés). Például a 13. itemet a válaszok 67%-a helyesen a B2 szintre tette, a válaszok 3%-a B1 szintűnek vélte, míg a válaszok 30%-a C1 szintűnek gondolta.

Sor-szám	Szint	Nehézségi érték	Deskriptor (item)	Disztraktor-elemzés
13.	B2	67%	Megbízhatóan át tud adni részletes információt. Világosan és részletesen le tudja írni, hogyan kell egy eljárást lebonyolítani. Összegezni és jelenteni tud több forrásból származó információt és érvet.	B1: 03% B2: 67% C1: 30%
15.	B2	67%	Valódi beszélgetőpartnerként tud részt venni hosszú tárgyalásban a legtöbb általános témában, még zajos környezetben is. Kapcsolatot tud fenntartani anyanyelvű beszélőkkel anélkül, hogy akaratlanul megnevettené vagy idegesítené őket, vagy olyan viselkedésre készítené őket, ami eltér más anyanyelvű beszélőkkel szemben tanúsított viselkedésüktől. Ki tudja fejezni az érzelmek különböző fokozatait, és hangsúlyozni tudja az események és élmények személyes jelentőségét.	B1: 01% B2: 67% C1: 31%
34.	C1	66%	Alkalmanként kisebb tévesztések, de semmi jelentős szóhasználati hiba.	B1: 07% B2: 26% C1: 66%
28.	C1	64%	A beszéd menetének teljes megszakítása nélkül ki tudja kerülni a felmerülő problémát, és át tudja fogalmazni mondanivalóját.	B1: 03% B2: 33% C1: 64%
31.	C1	64%	A könnyen előhívható diskurzusfunkciók széles köréből ki tud választani egy alkalmas fordulatot megjegyzései megfelelő bevezetéséhez, annak érdekében, hogy átvegye a szót, vagy időt nyerjen, és magánál tartsa a szó jogát, amíg gondolkodik.	B1: 10% B2: 26% C1: 64%
30.	C1	63%	Véleményeket és állításokat pontosan tud minősíteni, a bizonyosság/bizonytalanság, meggyőződés/kétely, valószínűség stb. szintjének figyelembevételével.	B1: 06% B2: 30% C1: 63%
16.	B2	60%	Olyan folyékony és természetes interakcióra képes, ami lehetővé teszi a rendszeres interakciót és hosszabb kapcsolat fenntartását anyanyelvű beszélőkkel, anélkül, hogy az bármelyik fél számára megerőltető lenne. Ki tudja fejezni események és élmények személyes jelentőségét, világosan meg tudja indokolni és fenn tudja tartani nézeteit, megfelelő magyarázatot és érveket szolgáltatva.	B1: 00% B2: 60% C1: 39%
12.	B1	56%	Élményeiről részletesen be tud számolni, érzései és reakciói bemutatásával. El tudja mondani váratlan események (pl. balesetek) részleteit. El tudja mondani egy könyv vagy film cselekményét, és le tudja írni reakcióit. Leírást tud adni álmokról, reményekről és ambíciókról. Leírást tud adni valóságos vagy elképzelt eseményekről. El tud mondani egy történetet.	B1: 56% B2: 34% C1: 10%

6. táblázat. A legnehezebbnek bizonyult itemek

Az első lehetséges megállapítás az, hogy a legnehezebbnek bizonyult nyolc deskriptor nyolc különböző skálából került ki. A disztraktoranalízis alapján további érdekes megfigyeléseket lehet tenni. Alapvetően a B2 és a C1 itemeket egymással keverték. A három B2 szintű itemet (13., 15., 16.) gyakorlatilag senki nem tette B1 szintre (3%, 1%, és 0%), a négy C1 szintű itemet (34., 28., 31., 30) azonban a válaszok 7, 3, 10 és 6 százaléka helyezte B1 szintre. A 34. item esetén ezt talán azzal lehet magyarázni, hogy B1 szinten is lehet szókincsbeli hiba nélkül beszélni, tehát ez a deskriptor akár ezt a szintet is jelentheti. A 28. és a 30. itemek esetén nem találtunk magyarázatot arra, hogy ezeket az itemeket miért lehetett B1 szintűnek jelölni. A 31. item esetén esetleg az *időnyerés* illetve a *gondolkodás* szavak bizonytalaníthatták el a válaszadókat. Ezek a bizonytalanságok aláhúzzák a KER folyamatos, a konkrét vizsgáztatást megelőző tanulmányozásának fontosságát.

A legnehezebbnek bizonyult item B1 szintű volt (12. item, 56%). Ennek valószínűleg ugyanaz az oka, mint a fentebb részletezett jelenségnek, azaz ha a vizsgáztatók nem pontosan olvassák át a KER-t a kitöltés előtt, akkor ez olyan deskriptornak tűnhet, ami mintha „főlé lőne” a szintnek, például a *váratlan események*, vagy *egy film cselekményének* elmondásával, illetve álmok és *remények* leírásával. A KER skálái egyenetlenek, ugyanis nem mindegyiket kalibrálták empirikus eszközökkel (Dávid 2012), továbbá „számos szintleírást a könyv [KER] szerkesztése során, de már a kalibrálás után árttak, módosítottak, még olyanokat is, amelyekkel a kalibráció során jó eredményeket értek el” (Dávid 2012: 25), valamint néhány deskriptor nagyon rövid és lényegretörő, míg mások jóval hosszabbak (Imam—Shaw 2013). Ez az item (12.) az alacsony nehézségi értéke miatt meglehetősen „szétszórtá” a válaszokat, hiszen 34% B2 szintre sorolta, 10% pedig C1-re. Következésképpen a diszkriminációs indexe is igen magas, 0,82.

A könnyű itemek elemzésekor említettük, hogy a vizsgáztatók még mindig a nyelvtan és a szókincs szintjének a megítélésében a legbiztosabbak. Ennek a megállapításnak ellentmond az a tény, hogy a korábban már megemlített 34. sorszámú item a *Szókincs alkalmazása* skálából származik és 66%-os nehézségi értékével meglehetősen nehéznek bizonyult. Ennek lehetséges okát fent már kifejtettük.

A 7. táblázatban azt a három itemet tüntetjük fel, amelyeket a nagyon hasonló megfogalmazás miatt, előzetesen egymással összekeverhetőnek, azaz meglehetősen nehéznek vártunk.

Sorszám	Szint	Nehézségi érték	Deskriptor (item)	Disztraktor-elemzés
20.	B2	84	Világos leírást vagy elbeszélést tud kidolgozni, mondanivalójának főbb pontjait kifejtve, megfelelő részletekkel és példákkal alátámasztva.	B1: 07% B2: 84% C1: 09%
17.	B2	83	Világos, részletes leírást és bemutatást tud adni az érdeklődési köréhez kapcsolódó témák széles skálájában, gondolatait kiegészítő adatokkal és megfelelő példákkal fejti ki és támasztja alá.	B1: 08% B2: 83% C1: 09%
29.	C1	76	Világos, részletes leírást és bemutatást tud adni összetett témákkal kapcsolatban; összekapcsolja az egyes altémákat, kifejti az egyes pontokat, és megfelelő befejezéssel zár.	B1: 01% B2: 23% C1: 76%

7. táblázat. Előzetesen egymással összekeverhetőnek várt itemek

A 7. táblázat itemei nem teljesen igazolták az elvárásokat, ugyanis nem bizonyultak túlságosan nehéznek. A két B2 szintű item a második nehézségi sávba esett, azaz olyan itemek közé, ahová a második legkönnyebben megoldott deskriptorok tartoznak (84% és 83%). A disztraktoranalízisből jól látszik az a megfigyelés, hogy mindkét B2 szintű itemet gyakorlatilag ugyanannyian tartották a B1 és a C1 szintbe tartozónak is (7% és 8%, illetve 9% és 9%). Feltétlen pozitívum, hogy ugyanakkor a C1 szintű item gyakorlatilag nem került a B1 szintre (1%), viszont meglehetősen sok válasz sorolta a B2 szintű itemek közé. Kétségtelenül összekeverhetőek ezek az állítások, ha valaki nem tudja, hogy az összetett szó az, ami a különbséget jelenti a 20., 17. és 29. sorszámú B2 és C1 szintű deskriptorok között. Ezen itemekről a következő fejezetben még lesz szó.

KER feladatlap – Modern tesztelmélet: illeszkedési mutatók

Az elemzés során az itemek modern tesztelméleti eszközökkel kiszámolt illeszkedési statisztikáit (*infit* és *outfit*) is megvizsgáltuk a Winsteps program segítségével. Ezek az értékek azt mutatják meg, hogy az egyes itemek milyen mértékben illeszkednek ahhoz a modellhez, amely szerint a vizsgázók (jelen esetben a vizsgáztatók) várható teljesítménye az itemek nehézségi szintjének és a vizsgázók tudásszintjének a függvénye. Azaz, ha egy item nehézségi értéke és a vizsgázók képességszintje azonos, akkor 50% az esélye annak, hogy a vizsgázók az adott itemre helyes választ adnak. Minél magasabb a vizsgázók képességszintje, annál nagyobb a valószínűsége a jó megoldásnak, és ez fordítva is igaz: minél alacsonyabb a vizsgázók képességszintje, annál kisebb a valószínűsége a jó megoldásnak. Ezeket a modern tesztelméleti eszközöket azért alkalmaztuk, mert az ilyen típusú vizsgálatok olyan aspektusokra is rávilágíthatnak, amelyekre a klasszikus tesztelezésen alapuló vizsgálat nem, illetve megerősíthetik a klasszikus elmélet alapján talált eredményeket.

A 8. táblázatban az első oszlop az itemek sorszámát, a második a helyes válaszok számát, a harmadik az összes válasz számát, a negyedik a logitban megadott nehézségi értéket, az ötödik pedig az itemhez kapcsolódó mérési hibát tartalmazza. Az itemek illeszkedési mutatóit (*infit mnsq* és *outfit mnsq*) a 6. és a 8. oszlop tartalmazza. Mivel a KER feladatlap tulajdonképpen egy háromopciós feleletválasztós teszt volt, ezért ilyen esetben az illeszkedésmutatók elfogadott határértékét 0,7 és 1,3 közé teszik (Linacre 2014).

610 SZEMÉLY (KITÖLTÉS) 35 ITEM WINSTEPS 3.91.2														
SZEMÉLY: REAL SZEP.: 1.56 MEGBÍZH.: .71 ... ITEM: REAL SZEP.: 7.29 MEGBÍZH.: .98														
ITEM STATISZTIKA														
SOR SZÁM	EREDM.	ÖSSZES KITÖLTÉS	LOGIT	MÉRÉSI HIBA	INFIT		OUTFIT		DISZKR. TÉNY. VÁRT		PONTOS EGYZÉ TÉNY.% VÁRT%		NEH. ÉRTÉK	ITEM
					MNSQ	ZSTD	MNSQ	ZSTD						
1	543	610	-.59	.14	.98	-.2	1.13	.7	.30	.29	88.0	87.9	.89	1
2	515	610	-.12	.12	.93	-1.0	.74	-1.8	.41	.33	83.6	83.0	.84	1
3	515	610	-.12	.12	.95	-.8	.91	-.6	.39	.33	83.2	83.0	.84	1
4	519	610	-.18	.12	.87	-1.9	.73	-1.9	.46	.33	85.1	83.7	.85	1
5	573	610	-1.31	.18	.96	-.2	.78	-.7	.27	.22	93.2	93.2	.94	1
6	511	610	-.06	.12	.93	-1.1	.74	-1.9	.41	.34	82.5	82.3	.84	1
7	587	610	-1.84	.22	.93	-.3	.86	-.3	.26	.18	95.8	95.8	.96	1
8	523	610	-.24	.13	.90	-1.3	.65	-2.4	.43	.32	84.2	84.4	.86	1
9	429	610	.91	.10	1.05	1.1	1.05	.6	.37	.42	69.2	72.4	.70	0

10	591	610	-2.05	.24	1.03	.2	2.13	2.4	.09	.17	96.5	96.5	.97	1
11	589	610	-1.94	.23	.97	-.1	.74	-.6	.21	.17	96.1	96.1	.97	1
12	338	610	1.80	.10	.89	-2.7	.85	-2.6	.55	.46	74.2	71.0	.55	0
13	403	610	1.17	.10	1.11	2.6	1.27	3.5	.34	.43	68.0	71.2	.66	1
14	489	610	.23	.11	1.02	.3	.96	-.3	.35	.36	79.0	78.8	.80	1
15	406	610	1.14	.10	.94	-1.6	.91	-1.3	.47	.43	74.0	71.3	.67	0
16	367	610	1.52	.10	.98	-.6	.96	-.6	.48	.45	71.5	70.7	.60	1
17	505	610	.02	.12	1.08	1.3	1.54	3.4	.25	.34	81.0	81.3	.83	1
18	544	610	-.61	.14	1.05	.6	1.33	1.6	.21	.29	87.8	88.0	.89	1
19	436	610	.84	.10	1.02	.4	1.01	.1	.38	.41	71.6	72.9	.71	0
20	513	610	-.09	.12	1.02	.3	1.07	.5	.31	.33	82.5	82.6	.84	0
21	553	610	-.79	.15	.99	.0	1.10	.5	.27	.27	89.7	89.6	.91	1
22	530	610	-.36	.13	1.03	.4	.95	-.2	.28	.31	86.0	85.6	.87	1
23	532	610	-.39	.13	1.03	.4	1.11	.7	.27	.31	86.0	85.9	.87	1
24	490	610	.22	.11	1.12	2.1	1.30	2.3	.24	.36	77.7	78.9	.80	0
25	542	610	-.57	.14	1.03	.4	1.14	.8	.24	.29	88.2	87.7	.89	1
26	538	610	-.50	.13	1.02	.2	1.12	.7	.28	.30	87.1	87.0	.88	0
27	477	610	.38	.11	1.05	.9	1.13	1.1	.33	.38	76.1	77.1	.78	1
28	388	610	1.32	.10	1.09	2.3	1.12	1.8	.36	.44	68.5	70.8	.64	0
29	461	610	.56	.11	1.07	1.6	1.31	2.8	.32	.39	74.8	75.2	.76	0
30	385	610	1.35	.10	1.08	2.0	1.04	.6	.37	.44	66.9	70.8	.63	0
31	387	610	1.33	.10	1.00	-.1	.96	-.6	.43	.44	70.5	70.8	.63	1
32	493	610	.18	.11	.94	-1.0	.84	-1.3	.41	.36	80.1	79.4	.81	1
33	576	610	-1.40	.18	.94	-.4	.90	-.2	.30	.21	93.7	93.7	.94	1
34	403	610	1.17	.10	.89	-2.9	.81	-2.9	.53	.43	75.7	71.2	.66	0
35	561	610	-.98	.16	.94	-.5	.73	-1.1	.33	.25	91.3	91.0	.92	1
MEAN	491.8	610.0	.00	.13	.99	.0	1.03	.1			81.7	81.7		
P.SD	69.9	.0	1.00	.04	.07	1.3	.28	1.6			8.6	8.3		

8. táblázat. Az itemek illeszkedési statisztikái

Az összes item *infit* statisztikája (*infit mnsq*) megfelelőnek bizonyult. Ha azonban az *outfit* értékeket nézzük (*outfit mnsq*), akkor a 8. táblázat megerősíti a 7. táblázatban már korábban elemzett itemek összekeverhetőségéről szóló előzetes várakozásokat. A 17. és a 29. itemnél az *outfit* érték ugyanis meghaladja 0,7-1,3 közötti határértéket (1,54 és 1,31), ami azt jelenti, hogy ezek az itemek azon válaszadók itemválaszaira érzékenyek, akiknek a képességszintje az itemek nehézségétől távoli, azaz amikor egy alapvetően nagyon jól teljesítő résztvevő nagyon könnyű itemet ront el vagy fordítva. Ez az eredmény a találgatást bizonyítja.

A 10. item is nagyon érdekes, és erről is tettünk említést már korábban, mert ez volt az egyik legkönnyebbnek bizonyult deskriptor (lásd 5. táblázat). Kiugróan magas az *outfit* értéke (2,13), ami szintén találgatásra utal, azaz jónéhányan nem tudták helyesen besorolni azok közül, akiknek kellett volna, és akiknek nem kellett volna, azok pedig eltalálták. Ennek minden bizonnyal az item nagyon gyenge diszkriminációs ereje az oka (*Ptbiserl-ex Corr*: 0,09).

A 9. táblázat két olyan itemet mutat, amelyek esetében a klasszikus elemzés nem mutatott ki problémát.

Sorszám	Szint	Deskriptor (item)
8.	B1	Egyenes vonalvezetésű elbeszélést vagy leírást meglehetősen folyékonyan tud előadni, a főbb pontok lineáris sorba rendezésével.
18.	B2	Meglehetősen egyenletes beszédtempóval beszél. Bár időnként habozik, amikor szerkezeteken és kifejezéseken gondolkozik, csak kevés észrevehetően hosszabb szünet keletkezik emiatt. Bizonyos fokú folyékonyással és spontaneitással rendszeres interakciót tud fenntartani anyanyelvű beszélőkkel anélkül, hogy az bármelyik fél számára megerőltető lenne.

9. táblázat. Illeszkedési problémás itemek

A 8. item esetén túl alacsony volt az outfit érték (0,65), ami azonban a KER-teszt esetében nem jelentkezik problémaként, mert ez azt jelenti csupán, hogy túlságosan megjósolható válaszok születtek erre az itemre, azaz tulajdonképpen akinek el kellett találnia, az el is találta, akinek pedig el kellett tévesztenie, az el is tévesztette (túllilleszkedés). A 18. deskriptor esetén azonban szintén a találgatásra utal a magas outfit érték (1,33). Ennek minden bizonnyal újfent az item viszonylag alacsony diszkriminációs ereje az oka (Ptbisertl-ex Corr: 0,21), amit úgy is meg lehet fogalmazni, hogy megjósolhatatlan, hogy a gyengébben és a jobban teljesítők közül ki találja el. A jobban teljesítőknek el kellene találniuk, a gyengébb eredményt elérőknek nem, de ezen item esetében ez nem így van.

Szóbeli vizsgázói teljesítmények értékelése – Néhány tanulság

Ahogy azt a bevezetésben is említettük, a szóbeli vizsgázói teljesítmények értékelése során számos olyan érzékeny adat keletkezett, amelyet egy működő vizsgaközpont nem publikálhat, ezért ezzel kapcsolatban csak általános jelenségekre és tanulságokra térünk ki, ezek azonban gyakorló nyelvtanárok számára is érdekesek lehetnek. Az egyik ilyen jelenség az egyes értékelési kritériumokra adott pontszámok kérdése, a második a határesethez közeli és távoli teljesítmények értékelése, a harmadik pedig a formátum, azaz az audio hangfájlokra és a videóra rögzített teljesítmények értékelése közötti esetleges különbségek problémája.

A B2 szintű szóbeli vizsgázói teljesítmények megtekintése és értékelése öt nyelvből, angoltól, németből, franciából, olaszból és spanyolból történt. A feladatlapban szerepelt egy audio hangfájl, amelynek a meghallgatása közben lehetett tanulmányozni a feladatokat és az értékelési skálát is. A feladatlapon összesen öt kritérium alapján kellett pontozni a teljesítményt a BME értékelési rendszere szerint. Az öt kritérium a következő volt: *Feladatmegoldás*, *Kifejezőkészség*, *Nyelvhelyesség*, *Folyamatosság és koherencia*, valamint *Kiejtés*. A *Feladatmegoldás* kritérium a feladat végrehajtását, a beszédszándékok megvalósítását, a tartalmi relevanciát, a nyelvi funkciók megfelelő alkalmazását és a beszédértést jelenti. A *Kifejezőkészség* a szókincs terjedelmére és alkalmazására, a stílus és a regiszter használatára utal. A *Nyelvhelyesség* a morfológiai és szintaktikai elemek változatosságát és alkalmazását jelenti. A *Folyamatosság és koherencia* a folyamatosságot, és az információ szerkesztettségét, míg a *Kiejtés* a hangképzést, a szó-, mondat-, és beszédhangsúlyt takarja.

A beszéd-készség mérése az adott vizsgán három feladat segítségével történik. Az elsőben a vizsgázónak néhány kifejtő kérdés alapján kell a mindennapi életben előfor-

duló témák széles körében saját személyével kapcsolatban beszélgetnie a vizsgáztatóval. A második részben a vizsgázó önállóan fejt ki gondolatait egy adott témáról, amelyhez képi stimulus tartozik. A harmadik részben a vizsgázónak egy szerepjátékot kell eljátszania a vizsgáztatóval, célnyelven leírt szituációs feladat alapján. A három feladatot külön-külön a *Feladatmegoldás*, a *Kifejezőkészség*, és a *Nyelvhelyesség* kritériumokkal értékeljük egy 0-tól 5 pontig tartó skálán. A *Folyamatosság és koherencia* kritériumot az első és a harmadik feladat esetében együttesen használjuk szintén egy 0-tól 5 pontig tartó skálán, mert mindkét feladat interaktív jellegű dialógus, míg a második feladat esetén külön értékeljük ezt a szempontot, mert ez egy monológ. A *Kiejtés* kritériumot a három feladaton átívelően egyszer értékeljük szintén egy 0-tól 5 pontig tartó skálán.

A 10. táblázat az első négy legszigorúbban értékelt azaz a legalacsonyabb pontszámokkal értékelt kritériumot, a 11. táblázat pedig a négy legenyhébben értékelt, azaz a legmagasabb pontszámokkal értékelt kritériumot tartalmazza. A kritériumok utáni szám azt mutatja, hogy az a kritérium a három feladat közül melyikre vonatkozik.

	Angol	Német	Francia	Olasz	Spanyol
1.	Nyelvhelyesség3	Kifejezőkészség1	Nyelvhelyesség3	Kifejezőkészség2	Nyelvhelyesség3
2.	Nyelvhelyesség2	Nyelvhelyesség1	Nyelvhelyesség2	Nyelvhelyesség2	Foly és koh2
3.	Foly és koh2	Feladatmegoldás1	Foly és koh 1-3	Foly és koh2	Kifejezőkészség3
4.	Nyelvhelyesség1	Nyelvhelyesség3	Nyelvhelyesség1	Nyelvhelyesség1	Feladatmegoldás3

10. táblázat. Legszigorúbban értékelt kritériumok

	Angol	Német	Francia	Olasz	Spanyol
9.	Kifejezőkészség1	Kifejezőkészség2	Kifejezőkészség1	Foly és koh 1-3	Kifejezőkészség2
10.	Kiejtés	Kiejtés	Kifejezőkészség2	Feladatmegoldás3	Feladatmegoldás2
11.	Feladatmegoldás3	Foly és koh2	Feladatmegoldás1	Feladatmegoldás1	Kifejezőkészség1
12.	Feladatmegoldás1	Feladatmegoldás2	Feladatmegoldás2	Kiejtés	Feladatmegoldás1

11. táblázat. Legenyhébben értékelt kritériumok

A 10. táblázatból az látszik, hogy a vizsgáztatók (nyelvtől függetlenül) a legszigorúbban a nyelvismereti kritériumokat kezelték, azaz a *Nyelvhelyességre* és a *Kifejezőkészségre* megítélt pontokkal bántak leginkább szűkmarkúan. Ez a megfigyelés két ok miatt jelentős. Egyrészt az egyes nyelveken nagyon eltérő standard pontszámú teljesítménymintákat kellett értékelni, tehát a nyelvismereti kritériumokkal kapcsolatos szigorúság nem annak volt köszönhető, hogy esetleg gyenge nyelvi szintű volt az összes értékelt teljesítmény, azaz a vizsgáztatók teljesítménytől függetlenül értékelték szigorúan a nyelvismereti kritériumokat. Másrészt pedig érdekes összefüggés, hogy a KER feladatlapon is a B2 szintű *Grammatikai helyesség* volt az egyik legkönnyebb

item, a vizsgáztatók ezt ismerték a legjobban és ez volt az a kritérium, amely esetén a legerősebben nyomták meg a ceruzát a vizsgateljesítmények értékelésekor.

A kommunikatív kritériumok alig tudtak bekerülni az élemezőnybe, ez alól csak a német és a spanyol nyelvek volt kivétel: a *Feladatmegoldás1* (német) és *Feladatmegoldás3* (spanyol) ugyanis bekerült a legszigorúbban értékelt kritériumok közé. A spanyol nyelv egy másik szempontból is kissé eltér a többi nyelvtől. A legszigorúbban ugyan itt is a *Nyelvhelyességet* (3. feladat) értékelték és a harmadik helyen is nyelvi kritérium szerepel (*Kifejezőkészség3*), de a kettő közé ékelődik a *Folyamatosság és koherencia*, a negyedik helyen pedig már a *Feladatmegoldás3* szerepel. A legenyhébben értékelt kritériumok köre már hasonlatos a többi nyelvhez, ugyanis az utolsó három helyből kettőt itt is a *Feladatmegoldás* (1. és a 2. feladat) foglal el.

A fenti jelenség ellentéte tapasztalható a legenyhébben értékelt kritériumok esetén, amit a 11. táblázat mutat, itt ugyanis a kommunikatív kritérium, azaz a *Feladatmegoldás* túlsúlya figyelhető meg. Ez nem támasztja alá azt a vélekedést, amely szerint a vizsgázók viszonylag gyenge kifejezőkészsége befolyásolhatja a kommunikatív érték, azaz a *Feladatmegoldás* megítélését. Ezt az is alátámasztja, hogy a vizsgáztatók nem estek a halo-effektus csapdájába, azaz az egyik kritériumra adott pontszám nem befolyásolta a többi kritériumra adott pontszámokat. Ez a megfigyelés azonban statisztikai módszerekkel még további ellenőrzéseket igényel.

Egy másik érdekesség is nyilvánvalóvá vált a vizsgáztató-továbbképzés során. Azokon a nyelveken, ahol határesetbe tartozó vizsgázói teljesítményt kellett értékelni, jobban szórtak az eredmények, azaz nem volt olyan mértékű egyetértés a vizsgáztatók értékelései között, mint azokon a nyelveken, ahol egyértelműbben a megfelelő tartományba eső teljesítményt kellett értékelni. A jövőben érdemes lenne a mintateljesítmények kiválasztásakor ezt kompenzálni: azokon a nyelveken, ahol most határesetnek számító teljesítményt kellett értékelni, egyértelműen megfelelt vizsgázói teljesítményt kellene értékelteni, míg azokon a nyelveken, ahol most nagyon sikeres vizsgázói teljesítményt kellett értékelni, a jövőben inkább határesethez közeli teljesítményt kell választani. Ez gyakorlatilag ellenpróbája lehetne a jelenlegi eredményeknek.

A harmadik tanulság pedig az, hogy most nem videóra hanem audiófájlra rögzített teljesítményeket kellett értékelni, és nem tudható, hogy ez milyen mértékben befolyásolta az értékeléseket. A legközelebbi alkalommal érdemes lenne videóra rögzített vizsgateljesítményeket használni a továbbképzés során, hiszen más *csak hallani* és más *hallani és látni* a vizsgát. A továbbképzésen ugyan jó minőségű felvételeket lehetett meghallgatni, de bizonyára kisebb koncentrációt igényelt volna, ha látni is lehetett volna a vizsgázót, hiszen például a látható nonverbális reakciók a beszéd közbeni szüneteket is tolerálhatóbbá tehetnék volna.

Összefoglalás

A tanulmány a BME Nyelvvizsgaközpont szóbeli értékelői számára szervezett online továbbképzés módszereit, eredményeit és legfontosabb tanulságait tárgyalta. Az első részben részletesen ismertettük a KER familiarizációs teszt módszereit és eredményeit, a második részben pedig a részletek mellőzésével, csupán néhány fő tendenciára koncentrálva ismertettük a szóbeli teljesítményminták értékelésének tanulságait. Utóbbi esetben a részletek mellőzésének az volt az oka, hogy a szóbeli teljesítmény-

minták értékelésekor olyan érzékeny adatok is keletkeztek, amelyeket egy működő vizsgaközpont nem adhat ki.

Összegezve a KER skálák ismereteit ellenőrző tesztet megállapítható, hogy a vizsgáztatók KER ismeretei jók, és ezt egy megbízhatónak bizonyult teszt segítségével mértük fel. Ha az eredményeket a szintek felől közelítjük meg, akkor a nehézség tekintetében a három szint deskriptorai között kicsi, ám statisztikailag szignifikáns különbségeket találtunk: a B1 skálákat ismerték a vizsgáztatók a legjobban, ezt követte a B2, majd a C1. Ez a sorrend nem véletlen, hiszen a B1 szintű deskriptorok meglehetősen jól elkülöníthetőek a B2 és a C1 szintű itemektől. Ennek ellenére a legnehezebb deskriptor is a B1-esek közül került ki, ami elsősorban annak tartalmának, illetve megfogalmazásának, nem pedig a vizsgáztatók hibájának róható fel.

Ha az eredményeket az egyes KER skálák felől közelítjük meg, akkor a 91% feletti nehézségi értékű (azaz legkönnyebb) hét item közül öt a *Grammatikai helyesség* és a *Szókincs terjedelme* skálák közül került ki. Nagyon tanulságos, hogy ez a tudás visszaköszönt a szóbeli teljesítmények értékelésekor nyelvfüggetlenül is, hiszen a vizsgáztatóink a *Nyelvhelyesség* és a *Kifejezőkészség* kritériumokat értékelték a legszigorúbban, tehát azokat, amelyeket a KER skálák közül a legjobban ismertek.

A legnehezebbnek bizonyult nyolc deskriptor nyolc különböző skálából került ki. Ezen itemek körében érdekes megfigyelés volt, hogy néhány C1 szintű itemet jónéhányszor B1 szintűnek tüntettek fel a vizsgáztatók, míg a B2 szintű itemeket jóval ritkábban tették a B1 szintre.

A szóbeli teljesítmények értékelésekor három jelenségre utaltunk. Az első az egyes értékelési kritériumokra adott pontszámok különbsége, a második a határesethez közeli és távoli teljesítmények problémája, a harmadik pedig a formátum, azaz az audio hangfájokra és a videóra rögzített teljesítmények értékelése közötti esetleges különbségek kérdése volt. A vizsgáztatók a legszigorúbban a nyelvismereti jellegű kritériumokat értékelték; a határesetekhez közeli vizsgázói teljesítmények esetén nagyobb szórást mutattak a vizsgáztatók által megítélt pontszámok, mint a határesetektől távoli teljesítmények esetén; míg egy jövőre vonatkozó tanulság, hogy a felvételek formátuma (audio vs video) is befolyásolhatta az értékelők munkáját.

A fentiekén kívül még számos megfontolandó tanulsága volt a továbbképzésnek mind a vizsgaközpont, mind pedig az egyes vizsgáztató kollégák számára. Ezen tanulságok leszűrése és a vizsgáztatásba történő visszaforgatása elsősorban a vizsgázók miatt elemi érdek.

IRODALOM

- Brown, J. D. (2002): Statistics corner. Questions and answers about language testing statistics: The Cronbach alpha reliability estimate. *Shiken: JALT Testing & Evaluation SIG Newsletter*, Vol. 6 No. 1. (p. 17 - 19) [ISSN 1881-5537]. http://hosted.jalt.org/test/bro_13.htm Letöltve 2010. december 12.
- Dávid Gergely (2012): A szintleírások nyelvének szerepe a *Közös Európai Referenciakeret* magyar, angol és német nyelvű kiadásában. *Magyar Pedagógia* 112/1, 19–39.
- Green, R. (2013): *Statistical analyses for language testers*. Palgrave Macmillan: Basingstoke.

- Imam, H.—Shaw, S. (2013): *The CEFR: over-utilised or under-utilised? Perspectives from Cambridge Assessment*. <http://www.cambridgeassessment.org.uk/images/the-cefr-stuarly-shaw-helen-imam-presentation.pdf>
- Linacre, M. (2014): *A user's guide to WINSTEPS: Rasch-model computer programs*. Chicago: Winsteps.com. <http://www.winsteps.com/winsteps.htm>
- PTMIK (2002): *Közös Európai Referenciakeret. Nyelvtanulás, nyelvtanítás, értékelés*. Budapest: Pedagógustovábbképzési Módszertani és Információs Központ Kht.
- North, B. —Goodier, T. —Piccardo, E. (2017): *Common European Framework of Reference for Languages: Learning, teaching, assessment: Companion volume with new descriptors*. Provisional edition. <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/168074a4e2-teaching/168074a4e2>

KISS JENŐ, PUSZTAI FERENC (FŐSZERKESZTŐ)

A MAGYAR NYELVTÖRTÉNET KÉZIKÖNYVE

548 oldal 6990 Ft

Nemzeti kultúránk legfőbb hordozója a magyar nyelv. Több évezredes története sokakat érdekel. A magyar nyelvtörténet kézikönyve közérthető módon foglalja össze az eddigi vizsgálatok megállapításait, kiegészítve őket a legújabb kutatások eredményeivel.

A kötet a magyar nyelv három és fél ezer éves történetét a következő öt nagy korszakra bontva tárgyalja: ősmagyar kor; ómagyar kor; középmagyar kor; újmagyar kor és az újabb magyar kor.

A kézikönyvben az elméleti kérdések felvetése (a nyelvi változás mikéntje, a változás és a nyelvtörténet kapcsolatának mibenléte) és tisztázása után az érdeklődők részletes képet kapnak a legfontosabb magyar nyelvemlékekről, valamint a magyar helyesírás és hangjelölés történetéről.

Megvásárolható a kiadóban:

TINTA KÖNYVKIADÓ

1116 Budapest, Kiskőrös utca 10.; tel.: (1) 371-0501; fax: (1) 371-0502

E-mail: info@tintakiado.hu; honlap: www.tintakiado.hu