

Nyelvvizsgarendszerek összehasonlító elemzése a vizsgázói teljesítmények alapján

Összefoglaló a
2014 július – 2015 január
között végzett kutatásról



TÁMOP-3.1.8-09/1-2010-0004



A nyelvvizsgarendszereket összehasonlító kutatást lezáró részletes jelentés alapján, az Oktatási Hivatal kérésére készült az alábbi összefoglaló. Ebben a kutatás legfontosabb eredményeit ismertetjük és ezek okait tárjuk fel. Az összefoglaló készítése során szem előtt tartottuk, hogy annak tartalma a pszichometria területén kevésbé jártas érdeklődő számára is érthető legyen, a kutatás statisztikai és pszichometriai részleteinek taglalása nélkül.

Tartalom

Bevezetés	3
A kutatás menete	3
Kutatási kérdések.....	4
A kutatás módszerei	4
A vizsgák nehézsége.....	7
A vizsgarészek nehézsége	8
A vizsgák minőségének mutatói.	11
A vizsgajelentkezési és vizsgaszervezési folyamat tapasztalatai	15
Következtetések és javaslatok	16

Bevezetés

Ma Magyarországon nemzetgazdasági érdek a magas szintű idegennyelv-tudással rendelkező munkavállalók képzése, és ez a tevékenység a közoktatás és a felsőoktatás különböző szintjein, valamint a felnőttoktatásban teljesebb ki. Ezzel együtt elengedhetetlen az is, hogy az akkreditált nyelvvizsgarendszereket működtető vizsgaközpontok a megszerzett nyelvtudást megbízhatóan, egységes standardok szerint érvényesen tudják mérni. Az állam ellenőrző, minőségbiztosító szerepe ebben a folyamatban kiemelt fontosságú.

Az elmúlt évtized során az akkreditált nyelvvizsgáztatás szakmai feltételei és követelményei alapvető változásokon mentek keresztül. Az idegen nyelvek elméleti ismeretének számonkérése helyett egységes és szakmai eszközökkel betartatott követelmény lett, hogy az akkreditált vizsgák az idegen nyelv gyakorlati használatának képességszintjét ellenőrizzék hiteles módon. Mindennek különös jelentőséget kölcsönöz az a körülmény, hogy a széles elismertséggel rendelkező vizsgák követelményrendszere közvetlen módon hat vissza a köz- és felnőttoktatásban zajló nyelvoktatásra is. A munkavállalás során, a további tanulmányokhoz vagy általában a hétköznapi feladatok ellátásához szükséges nyelvi készségek hiteles és standardizált mérése azonban komoly kihívást is jelentett és új szakmai eszközök alkalmazásának szükségességét vetette fel mind a nyelvvizsgáztatást végző intézmények, mind az ezek munkája felett szakmai felügyeletet gyakorló hivatal számára.

Jelen kutatás célja annak objektív eszközökkel történő feltárása volt, hogy a legnagyobb vizsgázópopulációt lefedő országos működésű vizsgaközpontok munkája mennyire felel meg a hitelesség és a standardizáltság fenti elvárásainak. A kutatási források hatékony kihasználása érdekében kutatási programunk - **objektív, számszerűsíthető eredményeket megcélözva** - azt vizsgálta meg, hogy a bevont vizsgázói populáció tagjai több középfokú (B2 szintű) általános nyelvvizsgán azonos eredményt érnek-e el, idegennyelv-tudásukat a különböző nyelvvizsgarendszerek azonosnak vagy közel azonosnak minősítik-e, valamint, hogy ezen vizsgák mint mérési eszközök minőségi mutatói a gyakorlati működés során megfelelnek-e az elvárásoknak.

A kutatás alapcélja mellett arra vonatkozóan is igyekeztünk adatokat gyűjteni, hogy a nyelvvizsgaközpontok mint szolgáltató intézmények működése az ügyfélkezelés terén mennyiben felel meg a korszerűség és az ügyfélközpontúság kívánalmainak.

A kutatás menete

Jelen tanulmány a B2 szinten akkreditált általános angol, illetve német nyelvű nagy vizsgázói létszámmal működő vizsgarendszerek empirikus összehasonlításának módszertanát és legfontosabb eredményeit mutatja be. A 2014 júliusában elkezdett kutatás szervesen illeszkedik az Oktatási Hivatal Nyelvvizsgáztatási Akkreditációs Központ és a Nyelvvizsgát Akkreditáló Testület korábbi vizsgálatainak a sorába. A

kutatásban újszerű elem, hogy a próbavizsgálók becsült képességszintjeire alapozó számításokat is tartalmaz.

A magyarországi nyelvvizsgák összehasonlításával foglalkozó szakirodalmi eredmények nem foglalnak egyértelműen állást konkrét vizsgaanyagok eltérő nehézségével kapcsolatban. Kiszely és Szabó (2009) alacsony létszámú próbavizsgálói mintákon keresztül öt angol és négy német általános nyelvvizsgát hasonlított össze. A teljes vizsgán megfigyelt teljesítmények a német esetében nem mutattak eltéréseket, viszont az egyik angol vizsga (1. vizsga) jelentősen alacsonyabb eredményekkel zárult. Ugyanennek a vizsgának az olvasás vizsgarésze jelentősen nehezebb volt a többi vizsgáénál. Általános nyelvvizsgákat és az emelt szintű érettségit összehasonlító elemzésükben Együd, Kiszely és Szabó (2012), valamint Szabó és Kiszely (2010) azt találta, hogy bár az angol nyelvvizsgák viszonylagos hasonlóságot mutatnak, az egyik vizsga az érettségénél lényegesen magasabb elvárásokat támaszt a vizsgázókkal szemben. Gazdasági szaknyelvi vizsgák elemzése során Kiszely és Szabó (2010) arra az eredményre jutott, hogy az olvasás és az írás vizsgarész esetén valós különbségek mutatkoztak.

Jelen kutatás előkészítő szakaszában Kiszely (2014) a magyarországi nyelvvizsgák közhiteles nyilvántartása alapján végzett előtanulmányában leíró statisztikai eljárásokkal és eloszlásfüggetlen páros próbákkal (Wilcoxon-próba) bizonyítékot talált arra, hogy az angol és a német középfokú nyelvvizsgák között is jellemzőek a statisztikai jelentőséggel bíró különbségek. Mivel azonban elemzésében a vizsgázói minták hasonlóságának elvárására alapozott, nem volt kizárható, hogy a feltárt különbségeket a vizsgaidőszakonként és vizsgaközpontokként esetlegesen eltérő vizsgázói képességszint eredményezte. Az előtanulmány végkövetkeztetései között javaslatként szerepel ennek kiküszöbölésére a próbavizsgálók teljesítményén keresztül összekapcsolt mérés megvalósítása.

Kutatási kérdések

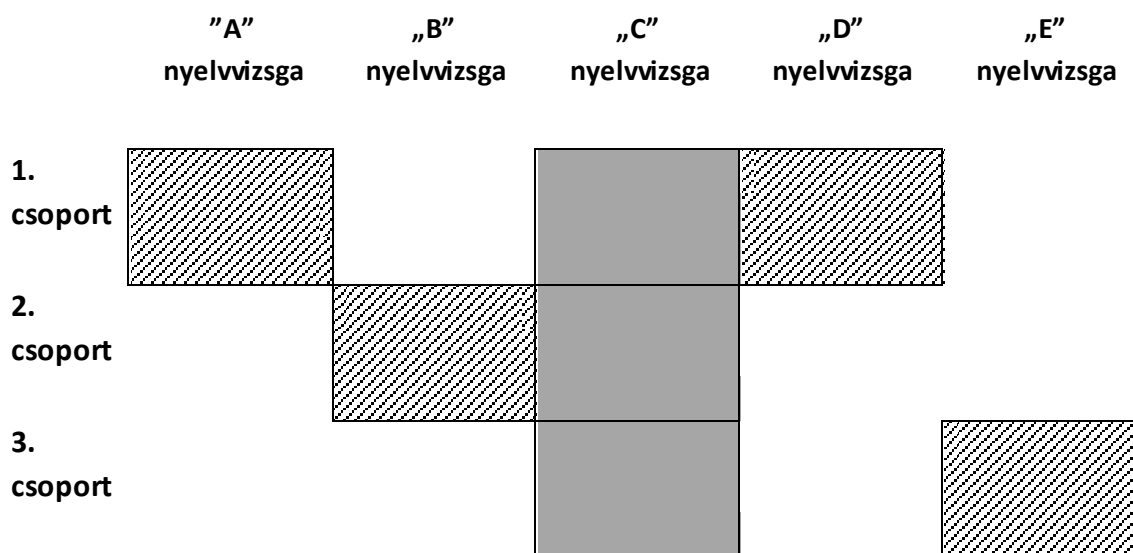
- 1. Különböznek-e egymástól – a normálisnak tekintett variancán túl – azon vizsgázók eredményei és részeredményei, akik legalább két különböző vizsgarendszerben is tettek vizsgát?*
- 2. Milyen erejű és irányú összefüggés mutatható ki a több vizsgarendszerben próbavizsgáló résztvevők azonos vizsgarészen szerzett részeredményei között?*
- 3. Azonos nehézségűek-e a B2 szinten akkreditált általános nyelvvizsgák azonos vizsgarészei?*

A kutatás módszerei

A kutatásban a Magyarországon nyelvvizsgáztatási joggal rendelkező és éves átlagban, országosan jelentős vizsgázószámmal rendelkező akkreditált központok közül – ABC sorrendben – a BME, az ECL, az Euro, az ITK és a TIT/TELC vett részt.

Anonimitásuk tiszteletben tartása érdekében a vizsgaközpontokra a továbbiakban kódokkal utalunk. A részvétel nem önkéntes alapon történt; a vizsgaközpontok utólag értesültek arról, hogy az Oktatási Hivatal és a Nyelvtudásért Egyesület által lefolytatott összehasonlító elemzésben adataikat felhasználjuk. A kutatás objektív és eredményes elvégzéséhez erre a részleges titkosításra feltétlenül szükség volt. Az angol nyelv esetében öt („A” nyelvvizsga, „B” nyelvvizsga, „C” nyelvvizsga, „D” nyelvvizsga, E” nyelvvizsga), a német esetében pedig – a források korlátozott volta miatt – két vizsgaközpont („A” nyelvvizsga és „C” nyelvvizsga) adatait vetettük össze.

Az összehasonlítás módszertani alapját egy olyan kapcsolt elrendezésű adatfelvétel alkotta, amelyben az összekötő elemek próbavizsgázói csoportok voltak. A véglegesített kutatási terv az alábbi ábrán látható.



A kutatási terv mintavétele

Az ábra sorai a tervezetten azonos létszámú ($n = 30$) próbavizsgázói csoportokat (1, 2, 3. csoport), az oszlopok pedig az egyes vizsgaközpontokat, illetve vizsgaalkalmakat jelölik (A-E). Az ábrázolt vizsgák az egyes vizsgaesemények időrendi sorrendjét követik. A szürkével kiemelt „C” nyelvvizsga minden próbavizsgázó számára kötelező volt. Ez, a valamennyi vizsgázó esetében rendelkezésre álló közös adat biztosította annak lehetőségét, hogy minden vizsga minden más vizsgával összevethető legyen. Az egyes vizsgák között eltelt időt minimalizáltuk azért, hogy tartható legyen a próbavizsgázók állandó képességszintjének elvárása. A német nyelvű összehasonlításban egyetlen próbavizsgázói csoport két vizsgájára hagyatkoztunk.

A próbavizsgázók toborzását és vizsgaközpontokhoz irányítását a Nyelvtudásért Egyesület munkatársai végezték. A résztvevők kiválasztásának első lépésében egy előzetes kérdőívet kellett a jelentkezőknek kitölteniük és szintfelmérőn vettek részt. A

háttér-információk között különösen fontos szerepe volt annak, hogy (a) a próbavizsgálónak még ne legyen középfokú nyelvvizsgálója a célnyelven, (b) ne vegyen részt a kérdéses vizsgarendszer vizsgaelőkészítő tanfolyamán, (c) ugyanakkor kellő mértékben ismerje a szint követelményeit, és (d) sikertelenül megkísérelt korábbi éles vizsga esetén másik vizsgarendszerben szerepelhessen. További elvárás volt, hogy a próbavizsgáló intenzív nyelvtanfolyamon ne vegyen részt, de tartsa szinten nyelvtudását, valamint hogy a szerződésben rögzített komplex nyelvvizsgák mindegyikén jelenjen meg és szerepeljen legjobb tudása szerint. A kötelezettségeket vállaló és betartó próbavizsgálók nyelvvizsgálóinak költségét a program fedezte.

A próbavizsgálók egyes vizsgákra történő felkészülését a program megvalósítása során kérdőívek segítségével monitoroztuk. Szintén kérdőíves valamint telefoninterjú adatgyűjtéssel szereztünk részletes információt a jelentkezési, tájékoztatói és vizsgaadminisztrációs folyamat vizsgázókat közvetlenül érintő aspektusairól valamint a vizsgák lebonyolításának technikai és szervezési körülményeiről.

A vizsgaadatok elemzésére az Oktatási Hivatal három nyelvtudásmérési szakértőt kért fel. A szakértők feladata a valós körülmények között letett vizsgák adatainak Klasszikus Tesztelmélet (KTE) és a Modern Tesztelmélet (item-válasz elmélet, IRT) eszköz- és módszertana szerinti feldolgozása, az itemek (elemi részfeladatok), vizsgarészek, részvizsgák és teljes vizsgák statisztikai, pszichometriai elemzése, valamint a próbavizsgálók teljesítményének összehasonlítása, illetve mindezeket keresztül az azonos szinten akkreditált vizsgarendszerek nehézségének összehasonlító elemzése volt.

Az elemzések elvégzéséhez nem csupán a próbavizsgálók végeredményeire, hanem az adott vizsgán részt vett teljes vizsgázói populáció elemi (itemszintű) eredmény-adatbázisára is szükség volt, melyet az OH-NYAK kért be az érintett vizsgaközpontoktól. Ez az adatszolgáltatás csak részlegesen valósult meg, így – mint látni fogjuk az eredmények taglalása során – a kutatási célkitűzések is csak némi korlátozással teljesültek.

A kutatásnak nem volt és nem is lehetett célja annak igazolása vagy cáfolása, hogy a B2 szintre akkreditált rendszerek valóban B2 szintűek-e, ahogyan arra sem volt alkalmas az összevetés, hogy egyes rendszerek B2 szintje inkább a B1 vagy a C1 szinthez közelít-e. A kutatásban relatív nehézségi összehasonlítást végeztünk, a rendelkezésre álló adatok függvényében minőségi mutatók számítását végeztük el, esetenként következtetve a nyelvi mérés szakmai elvárások szerinti megvalósíthatóságának egészére.

A vizsgák nehézsége

A végeredmények tekintetében a vizsgarendszerek jellemzően hasonlóan értékelték az angolos próbavizsgázók komplex teljesítményét. Az összesen öt páros összehasonlításból három esetben nem volt jelentős¹ az eltérés. A táblázatokban keresztek (✖) jelzik a statisztikai értelemben jelentősnek minősülő eltéréseket.

vizgaközpontok	végeredmények egyezősége
„A” nyelvvizsga és „C” nyelvvizsga	✓
„A” nyelvvizsga és „D” nyelvvizsga	✓
„B” nyelvvizsga és „C” nyelvvizsga	✓
„C” nyelvvizsga és „D” nyelvvizsga	✖
„C” nyelvvizsga és „E” nyelvvizsga	✖

A különbözőségek vizsgálatából levont következtetés akkor teljes, ha azt a korrelációs elemzések eredményeivel együtt értékeljük. Mivel minden páros összehasonlításban azonos tartalmakat és változatlanok tekintett nyelvtudást vetettünk össze, az eredmények erős összefüggését vártuk. Ez a statisztikai zsargonban magas szignifikancia szintet ($p < 0,01$) és nagy korrelációs együtthatót ($r > 0,50$) jelent. A korreláció maximális értéke 1,00 lehet. A tökéletes következetesség ($r \sim 1$) nem az adatok azonosságát jelzi, hanem azt, hogy az adatképek közötti eltérés mértéke állandó. Az a vizsgázó, aki az egyik vizsgarendszerben sok pontot szerez, a másikban is jól szerepel. Ugyanígy a kevésbé jól teljesítő vizsgázó mindkét vizsgán kevés pontot kap. A korreláció szignifikáns volta ezzel szemben annak jele, hogy a kapott mérési eredményeket nem okozhatja torz vagy túlságosan kicsi minta, hanem tényleges összefüggésről vagy eltérésről van szó.

A végeredmények páros korrelációs vizsgálatokor azt találtuk, hogy az adatsorok **mindegyike nagyon erős pozitív összefüggést mutatott.**

vizgaközpontok	végeredmények korrelációja
„A” nyelvvizsga és „C” nyelvvizsga	0,984
„A” nyelvvizsga és „D” nyelvvizsga	0,964
„B” nyelvvizsga és „C” nyelvvizsga	0,868
„C” nyelvvizsga és „D” nyelvvizsga	0,982
„C” nyelvvizsga és „E” nyelvvizsga	0,840

¹ A „jelentős” szót a statisztikai szaknyelvben megszokott módon használjuk. A jelentős különbség annyit tesz, hogy az eltérés ténylegesen megjelenik és nem mintavételi hiba eredményezi. A különbözőség vizsgálatokban az elemzés eltérést keres, tehát szigorúan értelmezve nem vezethet arra a következtetésre, hogy két dolog hasonló: azt állíthatjuk, hogy az adott két dolog (vizsga) nem eltérő.

A vizsgálatok fontos eredménye, hogy a két elemzési módszer egymást megerősítve mutatott rá arra, hogy az „A” nyelvvizsga, a „B” nyelvvizsga és a „C” nyelvvizsga gyakorlatilag azonosan értékeli a vizsgázók összteljesítményét. A minden próbavizsgázó által elvégzett közös vizsga lehetővé tette, hogy a meghirdetett végeredményeken keresztül az összes bevont vizsga nehézségi szintjét összehasonlítsuk. A közös vizsga, a „C” nyelvvizsga volt tehát az összehasonlítás során a viszonyítási alap.

vizsgaközpontok	végeredmények
„A” nyelvvizsga	0,31
„B” nyelvvizsga	3,15
„C” nyelvvizsga	0
„D” nyelvvizsga	-2,53
„E” nyelvvizsga	6,01

A „C” nyelvvizsga melletti érték azért nulla, mert önmagához képest nem könnyebb, vagy nehezebb. A negatív számok a „C” nyelvvizsgálathoz képest mért relatív nehézséget, a pozitív számok pedig a relatív könnyebbséget jelzik. A „C” nyelvvizsga és a „D” nyelvvizsga végeredményei között mutatkozó eltérés szignifikáns (tehát nem jelen mérésből adódó) volt ugyan, azonban a mindössze 2,53%² különbség a pszichológiai és pedagógiai mérésben tolerálható. A „C” nyelvvizsga és az „E” nyelvvizsga végeredményei közötti átlagosan 6,01%-os eltérés ennél számottevően nagyobb, itt komolyabb különbséget tárt fel a kutatás. A nehézségbeli eltérés még nagyobb, ha az egészében legnehezebb „D” nyelvvizsga és a legkönnyebb „E” nyelvvizsga közötti 8,54% eltérést figyeljük meg. A diszkrépancia okát a következő szakaszban ismertetjük.

A vizsgarészek nehézsége

A vizsgarészek tekintetében az összesen 20 darab páros összehasonlításból 12 esetben nem találtunk jelentős eltérést a részeredmények értékelésében.

² A %-ot, mint a meghirdetett értékek mértékegységét használjuk. Ez nem feltétlenül azonos a helyes válaszok arányával.

vizsgaközpontok	beszédértés ³	beszédkészség	írás	olvasás
„A” nyelvvizsga és „C” nyelvvizsga	x	✓	x	✓
„A” nyelvvizsga és „D” nyelvvizsga	✓	✓	x	✓
„B” nyelvvizsga és „C” nyelvvizsga	✓	x	x	✓
„C” nyelvvizsga és „D” nyelvvizsga	x	✓	✓	✓
„C” nyelvvizsga és „E” nyelvvizsga	x	x	✓	✓

Az olvasás vizsgarészt az összehasonlításban részt vevő mindegyik vizsgarendszer hasonlóan értékelte. Szintén nem találtunk egyetlen olyan vizsgarendszert sem, amelyik mindegyik vizsgarészen következetesen alacsonyabb vagy magasabb pontokat adott volna. A különbségek dominanciája a beszédértés és az írás esetében jelent meg. Ennek szakmai jelentősége abban áll, hogy egymástól alapvetően eltérő nyelvi készségekről beszélünk. Míg a beszédértés egy receptív készség és hagyományosan objektíven értékelhető feladatokból épül fel, az írás olyan produktív készség, amit kiképzett vizsgáztatók értékelnek. A kutatási terv szerinti elrendezésben az „A” nyelvvizsga és a „D” nyelvvizsga, valamint a „C” nyelvvizsga és a „D” nyelvvizsga mutatta a legkevésbé eltérő eredményeket.

A próbavizsgázók teljesítményét a részeredményeket illetően a vizsgaközpontok hasonlóan értékelték az összefüggés-vizsgálatok szerint is.

vizsgaközpontok	beszédértés	beszédkészség	írás	olvasás
vizsgarészenkénti korrelációértékek				
„A” nyelvvizsga és „C” nyelvvizsga	0,487	0,672	0,664	0,703
„A” nyelvvizsga és „D” nyelvvizsga	0,400	0,555	0,357	0,604
„B” nyelvvizsga és „C” nyelvvizsga	0,525	0,661	0,549	0,760
„C” nyelvvizsga és „D” nyelvvizsga	0,639	0,636	0,493	0,560
„C” nyelvvizsga és „E” nyelvvizsga	0,620	0,621	0,500	0,565

A részeredmények közötti összefüggések is jellemzően erősek voltak. A várakozásoktól az összesen 20 páros összehasonlításban 16 hasonló tendencia mellett négy kevésbé kielégítő összefüggést találtunk. Ezek közül két korrelációs együttható kevésbé maradt el a várt 0,50 értéktől (az „A” nyelvvizsga és a „C” nyelvvizsga a beszédértés, illetve a „C” nyelvvizsga és a „D” nyelvvizsga az írás tekintetében), ezért

³ A beszédértés vizsgarészre más források hallott szövegértés, hallásértés, labor vagy audi feladat néven is utalnak.

ezekkel nem szükséges külön foglalkoznunk. Ki kell emelni azonban, hogy az „A” nyelvvizsga és a „D” nyelvvizsga beszédértés, illetve írás vizsgarész-eredményei kevésbé hasonlítottak egymásra.

A korrelációs elemzés eredményei között két olyan vizsgapárt is találtunk, amelyeknél a hasonlóság véletlenszerű volt.

vizsgaközpontok	beszédértés	beszédkészség	írás	olvasás
vizsgarészenkénti korrelációértékek				
„A” nyelvvizsga és „C” nyelvvizsga	0,487	0,672	0,664	0,703
„A” nyelvvizsga és „D” nyelvvizsga	0,400	0,555	0,357	0,604
„B” nyelvvizsga és „C” nyelvvizsga	0,525	0,661	0,549	0,760
„C” nyelvvizsga és „D” nyelvvizsga	0,639	0,636	0,493	0,560
„C” nyelvvizsga és „E” nyelvvizsga	0,620	0,621	0,500	0,565

Az „A” nyelvvizsga és a „D” nyelvvizsga beszédértés és írás eredményei nem hozhatóak egyértelműen összefüggésbe. Ennek jelentősége a különbözőség vizsgálatokkal társítva igazán beszédes: a kutatás az „A” nyelvvizsga és a „D” nyelvvizsga íráskészség értékelésében talált eltérést. Az „A” nyelvvizsga írásbeli vizsgáztatói szigorúak voltak, de nem azoknak a vizsgázóknak adtak kevés pontot, akik a „D” nyelvvizsga rendszerében rosszul szerepeltek.

A fenti különbségek sok esetben statisztikai jelentőséggel bírtak – tehát nem mintavételi hibából származtak –, viszont gyakorlati jelentőségük elhanyagolható volt. **A vizsgarészek részeredményei közötti eltérések – két kivétellel - jellemzően kicsik voltak.**

vizsgaközpontok	beszédértés	beszédkészség	írás	olvasás
„A” nyelvvizsga	12,42	-1,23	-8,94	4,74
„B” nyelvvizsga	5,73	8,11	5,73	-1,97
„C” nyelvvizsga	0	0	0	0
„D” nyelvvizsga	6,38	0,17	-2,95	-4,62
„E” nyelvvizsga	11,29	20,41	8,72	-2,54

Fontos azonban kiemelni, hogy egyes esetekben lényeges nehézségbeli különbségeket tártunk fel. A beszédértés vizsgarészen az „A” nyelvvizsga feladatsora volt a legkönnyebb, 12,42%-kal magasabb részeredményeket adott, mint a készség tekintetében legnehezebb vizsgarendszer, a „C” nyelvvizsga. Ezt a részeredményt azonban a többi vizsgarész ellensúlyozni tudta. A legmarkánsabb különbségek a beszédkészségnél voltak kimutathatóak. Az „E” nyelvvizsga több mint 20%-kal haladta meg a „C” nyelvvizsga értékelését, és még ennél is nagyobb mértékben a legszigorúbb „A” nyelvvizsga pontjait. A beszédértés és a beszédkészség könnyűsége az „E” nyelvvizsga végeredményeiben is éreztette hatását, mivel a többi készség

részeredménye nem tudta kompenzálni a szóbeli vizsga megítélésénél megjelenő különbséget.

A német B2 szintű nyelvvizsgák összehasonlítását két vizsgaközpontnál végeztük el. A mintavétel szerény mérete miatt a statisztikai próbák eredményeit körültekintően kell értelmezni. A különbözőség-vizsgálatok sem a végeredmények, sem a részeredmények tekintetében nem találtak jelentős eltéréseket. A korrelációs vizsgálatok erős pozitív összefüggést mutattak ki a két német nyelvvizsga végeredményei között ($r = 0,754$; $p < 0,001$). A vizsgázói nyelvtudás egészét tehát az „A” nyelvvizsga és a „C” nyelvvizsga német nyelvű rendszere hasonlóan értékelte.

A részeredmények korrelációs vizsgálata során azt találtuk, hogy a beszédértés, az írás és az olvasás vizsgarészek erős pozitív összefüggést mutattak.

beszédértés	beszédkészség	írás	olvasás
0,694**	0,139	0,740**	0,655**

** $p < 0,01$

Az egyetlen jelentőséggel nem bíró korrelációs együttható a beszédkészség esetében volt megfigyelhető, a pontátlagok közötti eltérés azonban itt sem volt jelentős. Az „A” nyelvvizsga és a „C” nyelvvizsga német nyelvi vizsgarendszere tehát a próbavizsgázók teljesítményét hasonlóan értékelte.

A vizsgák minőségének mutatói

A nyelvvizsgák minőségi elemzésekor a feladatokat alkotó tesztlépések (itemek) minőségét, illetve a vizsgarészek megbízhatóságát vizsgáltuk. A kutatás eredményeinek értelmezésekor az adatok feldolgozási módszerei mellett azt is figyelembe kell venni, hogy a vizsgaközpont végez-e utólagos korrekciót. Ez az eljárás a kutatási beszámoló megírásának idején nem volt kötelező a vizsgázói teljesítmények értékelésekor.

Az itemek minőségi elemzésekor a pontos statisztikákhoz az elérhető legnagyobb adatbázisra volt szükség. **Mivel a „D” nyelvvizsga csak a próbavizsgázók elemi szintű adatait bocsátotta a kutatás rendelkezésére, az „E” nyelvvizsga pedig még azokat sem, itemek működését nem ismerhettük meg. Így a minőségi mutatók vizsgálatakor csak az „A” nyelvvizsga, a „B” nyelvvizsga és a „C” nyelvvizsga itemeiről alkothattunk képet. A három vizsga adatsorait elemezve általánosságban arra a következtetésre jutottunk, hogy a vizsgaközpontok tesztlépései a nyelvi mérésre alkalmasak voltak. Kevés kivétel mellett az itemek megfelelően erős összefüggést mutattak a részeredményekkel.** A kutatás elvégzésekor az itemek minőségi mutatóival kapcsolatos elvárások szakirodalmi forrásokban voltak elérhetőek. A 2015. január 1-jével hatályos Akkreditációs Kézikönyv már előírja, hogy a feladatokat felépítő elemek nehézsége és minősége milyen értékek között elfogadható.

A szakmai elvárásoktól az „A” nyelvvizsga beszédértés vizsgarészének tesztlépései maradtak el: itt az itemek 20%-a volt gyengébb minőségű. A „C” nyelvvizsga beszédértés itemei között háromnak volt alacsony a diszkriminációs ereje, azonban ez a vizsgarendszer a végeredmény előállításánál utólagos korrekcióval semlegesíti a gyenge minőségű vagy hibás itemek torzító hatását.

A kutatási jelentés publikációjának idején a hatályos szabályozás a sok vizsgázteljesítővel dolgozó vizsgaközpontok számára az adatfeldolgozás kétféle módját engedélyezte (AK. II. 12. és Folyamatszerű követelmények). Az első esetben a vizsgaközpont a vizsgaidőszak összes item szintű adatát rögzíti és a számítások során a teljes adatbázissal dolgozik. A második módszer szerint a vizsgaközpont – a résztvevők számától függően – mintát vesz a vizsgázteljesítők megoldásaiból, és ezzel a szűkebb adathalmazzal végzi a számításait. Az utólagos korrekció alkalmazása csak a teljes adatbázis rögzítése és feldolgozása esetén engedélyezett.

A mintavétellel történő itemanalízis pontosságának bemutatására az „A” nyelvvizsga beszédértés vizsgarészének adataival végeztünk statisztikai elemzéseket. A vizsgarész két, egyenként 10 itemből álló feladatból áll. Az elemzés során az itemek nehézségét vizsgáltuk. A táblázatban a p-értékeket százalékra konvertáltuk, így az item nehézsége 0 és 100 között változik és a helyes válaszok arányát jelöli. A nagyobb szám könnyebb, a kisebb érték pedig nehezebb kérdést jelent.

item	1. minta	2. minta	3. minta	Teljes
L01	83	80	82	81
L02	70	66	71	71
L03	51	62	54	54
L04	77	81	77	79
L05	46	39	34	38
L06	90	90	86	87
L07	76	77	70	71
L08	85	89	85	82
L09	28	36	23	27
L10	20	24	17	19
L11	98	99	96	97
L12	87	89	88	88
L13	99	98	95	97
L14	78	76	73	76
L15	92	92	89	88
L16	75	73	73	71
L17	58	56	42	52
L18	71	79	75	74
L19	73	72	66	71
L20	100	100	98	98
Vizsgarész	72,85	73,9	69,7	71,05

A táblázat jobb oldali oszlopa a teljes adatbázison végzett elemzés értékeit tartalmazza, az 1-3. minták pedig egy-egy véletlenszerűen választott, százfős mintán mért nehézséget jelölnek. A mintákat az IBM SPSS 22 programcsomaggal állítottuk elő.

A vizsgarész egészének nehézsége az utolsó sorban látható. Az elemzés igazolta, hogy a mintából számított globális nehézségi érték nem tér el számottevően a teljes adatbázis alapján számított értéktől. Az egyes itemek nehézségének pontos becslése viszont nem lehetséges mintavétel alapján. Egyes esetekben a minták közötti eltérés lényegesen meghaladja a 10%-ot is (L03, L05, L09 és L17). Másrészt a mintából számított nehézség alapján az item minősítése bizonytalan: a L17 item esetében az első minta inkább könnyűnek, a harmadik minta viszont nehéznek láttatja a tesztlépést. Különösen nehéz az ítélet az L09 item esetében, ahol a második minta szerint az item az előírt nehézségi spektrumon belül helyezkedik el, tehát elfogadható nehézségű, más minták és a teljes adatbázis szerint viszont túlságosan nehéz, ezért utólagos korrekció alkalmazását igényelné.

Ugyanezekon az adatokon az itemek minőségi ellenőrzését is elvégeztük. A diszkriminációs erő jelzésére az ún. korrigált item-teszt korrelációs együtthatót használtuk. Ez a szám azt mutatja meg, hogy az item mennyire hasonlóan ítéli meg a

vizsgáló tudását, mint az összes többi item együttvéve. Az Akkreditációs Kézikönyv Statisztikai elvárások című fejezete szerint az előírt legkisebb elfogadható érték $r = 0,25$. Az eredményhirdetéshez csak olyan item használható, amelynek diszkriminációja pozitív.

item	1. minta	2. minta	3. minta	Teljes
L01	0,07	0,185	0,162	0,213
L02	0,107	0,186	0,191	0,17
L03	0,259	0,357	0,494	0,298
L04	0,08	0,023	0,048	0,112
L05	0,36	0,45	0,363	0,387
L06	0,075	0,379	0,212	0,287
L07	0,266	0,314	0,219	0,287
L08	0,297	0,225	0,372	0,333
L09	0,267	0,422	0,215	0,291
L10	0,264	0,187	0,19	0,201
L11	0,174	0,029	0,277	0,134
L12	-0,011	0,045	0,293	0,198
L13	0,062	0,12	0,37	0,232
L14	-0,025	0,154	-0,039	0,082
L15	0,12	0,19	0,074	0,221
L16	0,136	0,062	0,144	0,094
L17	0,286	0,494	0,241	0,295
L18	0,414	0,146	0,219	0,226
L19	0,358	0,308	0,438	0,35
L20	0	0	0,048	0,133

A teljes adatbázison futtatott számítási sor szerint a vizsgarész minden iteme hasznos volt. Ezzel szemben a részleges mintavétel öt esetben mutatta azt, hogy az adott tesztlépés nem alkalmas nyelvi mérésre. **A részleges mintavétel az itemek minőségét egyes esetekben alul-, máskor felülbecsülte, tehát az vizsgaközpont számára megbízhatatlan képet ad a feladatsort alkotó elemek minőségéről. A vizsgálat tehát rámutatott, hogy az itemek pontos megítéléséhez a teljes vizsgázói adatbázis feldolgozására szükség van.**

A vizsgarészek megbízhatóságával kapcsolatos előírást az Akkreditációs Kézikönyv II.10. pontja tartalmazza. Elemzésünk feltárta, hogy a megbízhatóság legalacsonyabb elfogadható értékét ($\alpha \geq 0,75$) – egyetlen kivétellel – mindegyik vizsgaközpont minden vizsgarészen elérte. Megbízhatósági problémát az „A” nyelvvizsga beszédértés vizsgarésze mutatott. Ahogy arra az itemek minőségi elemzésekor is utaltunk, ezt a megbízhatósági deficitet a vizsgarészt felépítő tesztlépések idézték elő. A tesztben szerepelt egy két választási lehetőséget kínáló, ún. igaz/hamis feladat is; az alacsony megbízhatóság ennek a közismerten gyengén mérő feladatnak volt tulajdonítható.

A vizsgajelentkezési és vizsgaszervezési folyamat tapasztalatai

Kutatási programunk során a jelentkezési folyamatot (megfelelési kérdőív, telefonos interjú, szintfelmérés) követően is mindvégig nyomon követtük a tesztvizgázók felkészülését és tapasztalataikat. A vizsgaeseményeket megelőzően egy héttel a résztvevők kérdőívet töltöttek ki, melyen felkészülési tevékenységükről és tapasztalataikról számoltak be. Amellett, hogy a felkészülési beszámoltatás annak ellenőrzését is célozta, hogy egyik vizsgára se készüljenek célirányosan lényegesen többet vagy kevesebbet, visszajelzést kaptunk az egyénileg jelentkező és felkészülő vizgázók számára rendelkezésre álló gyakorló és felkészítő anyagok elérhetőségére vonatkozóan is. Utóbbi valamennyi vizsga esetében kielégítő volt a többség számára. Bár az érintett öt vizsgarendszerből csak kettő esetében áll rendelkezésre a gyakorlóanyagok mellett felkészülést segítő tankönyv is, ezt a viszonylag rövid felkészülési idő alatt nem is hiányolták a résztvevők, az ingyenesen, illetve pénzért elérhető gyakorlótesztek mennyisége pedig elégséges volt.

A vizsgajelentkezések során a lehetséges fizetési módok – csekk, átutalás, készpénz – között viszonylag egyenletes volt a megoszlás. Mivel a kutatási program szoros ellenőrzéséhez és elszámolásához mind a jelentkezéseket, mind a befizetéseket szigorúan dokumentálni kellett a résztvevőknek, az ehhez kapcsolódó adminisztrációs és számviteli hibák szembeötlően jelentek meg: a „D” vizsgára jelentkezők többsége csekken, az „E” nyelvvizsga esetében a legtöbben átutalással fizették be a vizsgadíjat. Ezen vizgázói csoport esetében visszatérő, rendszeres problémát jelentett, hogy a vizgázók befizetésükről külön kérés ellenére sem kaptak névre szóló, „áfás” számlát (amit az elszámoláshoz a résztvevőktől megköveteltünk). A jelentkezéseket elvileg kezelő vizsgahelyek sok esetben nem tudtak érdemi felvilágosítást adni, de több esetben a vizsgaközponttól is csak tétova ígéretet kaptak a vizgázók arra, hogy „majd a vizsga után megküldik” vagy „kb. decemberben fognak küldeni” számlát. Mindez nem csak a számviteli előírásoknak nem felel meg, de kielégítő ügyfélszolgálatnak sem tekinthető. Találkoztunk olyan egyetemi tanszéken működő vizsgahellyel, ahol a jelentkezési folyamat során egy folyosón kihelyezett dobozba kellett a jelentkezési lapot és a hozzátűzött eredeti(!) csekket bedobni úgy, hogy személyes, érdemi tájékozódásra semmilyen lehetőség nem volt.

Bár a vizgázók maguk nem észrevételezték, nem értékeltük kimondottan „jó gyakorlatként” azt sem, hogy miközben a „C” nyelvvizsgára személyesen jelentkezőknek a nyugtát automatikusan, kérés esetén pedig a számlát minden esetben, azonnal a kezükbe adják, a papíralapon kitöltött és leadott jelentkezési lapokból másolatot nem kapnak, így a majdani bizonyítványon megjelenő adataik helyes vagy netán hibás leadásának náluk nem maradt nyoma.

Az egyes vizsgákat követően szintén kérdőíven gyűjtöttünk adatokat a vizsgaszervezéssel, technikai feltételekkel és lebonyolítással kapcsolatban. Súlyos kifogást egyetlen esetben sem fogalmaztak meg a vizgázók a több, mint 220

vizsgaalkalommal kapcsolatban. A feltételek - döntő többségük szerint - jellemzően megfelelők voltak. Több esetben érte kritika a „C” és az „E” vizsgákon a beszédértés vizsgarészekén a lejátzó eszközök hangminőségét, három esetben számoltak be résztvevők számottevő csúszásról a beszédképesség vizsgarészen. A feltételek minőségével kapcsolatosan a legkevesebb észrevételt a „D” és az „A” vizsgák kapcsán fogalmazták meg a résztvevők. Többen is jelezték azonban, hogy a három napra tagolt „D” vizsgán való megjelenés, időpont-egyeztetés komoly problémát jelentett számukra (a szóbeli és a „labor” napja, illetve időpontja jelentkezéskor nem ismert).

Összességében a vizsgák lebonyolítását, szervezettségét és a technikai feltételeket a vizsgázók - jellemzően kritikus hozzáállásuk ellenére - valamennyi vizsga esetében megfelelőnek ítélték meg. Bár a vizsgázók koruknál fogva erre kevésbé érzékenyek, úgy véljük, az előzetes tájékoztatás, a jelentkezési és befizetési folyamat egyes vizsgarendszerek esetében az eljárások felülvizsgálatát és a színvonal érdemi javítását kívánná meg.

Következtetések és javaslatok

Az államilag elismert középfokú (B2 szintű) nyelvvizsgák összehasonlítása, melyet próbavizsgázók éles vizsgán mutatott produkcióinak elemzésével végeztünk el, számos tanulsággal szolgált. Ezek a kutatás természetéből adódóan legnagyobb részt statisztikai és pszichometriai természetűek, de már a szervezés, a vizsgalebonyolítás és az adatok bekérésének szakaszaiban is rengeteg fontos és hasznos tapasztalattal gazdagodhattunk.

A kutatás egyik újszerűsége az volt, hogy éles vizsgán felvett adatokkal dolgozott, és az egyes vizsgákat próbavizsgázók teljesítményei kapcsolták egymáshoz. Korábbi vizsgálódásokkal ellentétben a mostani kutatási terv nagy előnye, hogy így nem kellett nehezen tartható, bizonytalan, vagy biztosíthatatlan elvárásokat támasztanunk a tesztalanyok képességszintjével kapcsolatban. A tapasztalat azonban azt mutatja, hogy a nehéz ugyanazt a próbavizsgázót három különböző nyelvvizsgán szerepeltetni. A probléma a szervezés összetettségén túl részben abban is rejlik, hogy a kutatás éles vizsgahelyzetét a sokszori szereplés módosíthatja.

Az elemzés célja volt, hogy az item-válasz elmélet alapján összehasonlítsa a kutatásba bevont vizsgarendszerek egy-egy tesztjét. A számítások pontossága érdekében ennek elvégzéséhez a lehető legnagyobb vizsgázói item-válasz adatbázisra volt szükséges. **Egyes vizsgaközpontok segítőkézsége példás volt; megértették a projekt jelentőségét és lehetőségként fogták fel a részvételt. Mások azonban a hivatalos megkeresés ellenére sem adtak ki megfelelő mennyiségű és részletességű adatot.**

Munkánk során tapasztalhattuk, hogy a magyarországi vizsgaközpontok adatfelvételi gyakorlata nagymértékben eltér. Bizonyos vizsgaközpontok rutinszerűen rögzítenek minden elemi teszt-választ, míg mások csak az adatok töredékét veszik fel.

Ez nemcsak vizsgabiztonsági, de adatfeldolgozási és vizsgafejlesztési szempontból is aggályos.

A vizsgarendszereket összehasonlító elemzés feltárta, hogy a vizsgázói teljesítményeket a központok összességében hasonlóan értékelték. Minden vizsgarészt figyelembe véve könnyű vizsga nem volt. Az egyes könnyebb vizsgarészek hatását más, nehezebb vizsgarészek általában kompenzálni tudták. Jó példa erre az „A” nyelvvizsga esete, ahol a viszonylag könnyebb beszédértés és olvasás feladatokat ellensúlyozta a viszonylag nehéz beszédkézség és írás vizsgarész. Ellenpélda az „E” nyelvvizsgarendszer, ahol a beszédkézség könnyűségét a többi vizsgarész nem tudta kiegyenlíteni.

A kutatás rámutatott arra is, hogy az összetett adatfeldolgozás és eredményszámítás segít az eredmények pontosításában. Kutatásunkban az „A” nyelvvizsga, a „B” nyelvvizsga és a „C” nyelvvizsga központja elégséges mennyiségű adatot szolgáltatott ahhoz, hogy az itemek nehézségét és a vizsgázók vizsgarészenkénti képességszintjét modern tesztelméleti eljárásokkal meg tudjuk becsülni. Az elemzés fontos tanulsága, hogy a becsült képességszintekkel végzett páros összehasonlítások egyetlen esetben sem mutattak jelentős eltérést. Ez tehát annyit jelent, hogy az adatfelvétel eszközei – a nyelvvizsgák – hasznos információt adnak a vizsgázói nyelvtudásról, és komplex adatfeldolgozási eljárásokkal a nyelvi mérés pontatlansága jelentősen csökkenthető. Az ilyen számítások bonyolultak és jól felkészült szakembert igényelnek, előnyük azonban, hogy pontos visszajelzést adnak a nyelvtudásról.

Az összehasonlító elemzés több korrigálható hibát is talált. Egyes könnyebb vagy nehezebb itemek mellett voltak minőségi problémákat mutató tesztkérdések is. Helyes gyakorlat ezek akár utólagos törlése, hogy az eredményeket már ne torzíthassák. A vizsgázók nyelvtudásának rendszeres és szakszerű tesztelése jelentős hatással bír. A nyelvvizsgáztatás a köz- és felsőoktatásban vitathatatlanul fontos szerepet tölt be, a nyelvtanulás motorjaként működik. A munka világában az előléptetéseket és a bérezést sok esetben nyelvvizsgához kötik. Mindezek, valamint a hazai nyelvvizsgáztatás európai viszonylatban is magas szintje és elismert gyakorlata miatt **fontos, hogy a jelen kutatáshoz hasonló vizsgálatok rendszeresek legyenek, és eredményeik hatással legyenek a vizsgáztatás folyamatára. A kutatás során tapasztalt adatszolgáltatási problémák miatt szükségesnek látjuk, hogy a vizsgaközpontok, az OH-NYAK által lebonyolított, rendszeres minőségbiztosítási eljárások keretében folytatott ellenőrzések során elérhetővé tegyék a teljes vizsgaadatbázisukat. Az így nyert információ nem csak nekik, hanem a magyar nyelvvizsgáztatás egészének, ezen keresztül a nyelvtanítás eredményességének szempontjából is fontos, tanulságos és hasznos.**