

Mitől jó egy nyelvvizsga?

E cikk azt vizsgálja, hogy az idegen nyelvek vizsgáztatásában részt nem vevő nyelvtanár hogyan teheti mérlegre az idegen nyelvi vizsgákat egy kérdéssor segítségével. Abból indul ki, hogy az idegen nyelvi mérés leginkább csak a szakértők számára jól átlátható, azok számára is akkor, ha pontos áttekintést szereznek a tervezés, feladatkészítés és vizsgáztatás teljes folyamatáról, hozzájutnak a vizsgáztatás során keletkező válaszadatokhoz, valamint elemzéseket és felméréseket végezhetnek. A legtöbb nyelvtanár nincs ebben a helyzetben, ha nem dolgozik valamelyik vizsgaközpontnak és nincs speciális képzettsége sem a nyelvtudás mérése terén. Mindezek mellett valamilyen egyszerű iránymutatásra nekik is szükségük van, hogy a Magyarországon akkreditált több mint 20 vizsgaközpont kínálatát a maguk és diákjaik számára helyesen értékelni tudják. A cikk tíz kérdőív szerű kérdést és a hozzájuk kapcsolódó magyarázatokat, javaslatokat tartalmazza, melyek, reményeink szerint, segíthetnek a nyelvvizsgák minőségét meghatározó ismérveket megvilágítani.

Olyan fontos kérdésekre, mint „Megvan-e a vizsga kellő érvényessége?”, vagy „Megfelelő-e megbízhatósága?”, nem csak az érvényesség (validitás) és megbízhatóság fogalmak elvontsága miatt nehéz válaszolni (Messick, 1989; Kane, 2006), hanem azért is, mert a válaszadás többféle és igen összetett információ, adat együttes értékelését követeli meg. Vajon helyes-e a következtetés, amit levonhatnánk, miszerint a nyelvtanároknak – mert többségük nem idegennyelvtudás-mérési szakember – aligha lehet segítséget nyújtani abban, hogy jobban megítélhessék a nyelvvizsgák minőségét és így diákjaiknak is jobb tanácsot adhassanak? Felvetődik az a kérdés is, hogy milyen elvek és ismeretek alapján/segítségével tud a nyelvtanár eligazodni az információs dömpingben. A jó és kevésbé jó nyelvvizsgák elkülönítése mellett általánosabb kérdés is felmerül: mit jelent a minőség fogalma, ha nyelvvizsgáról van szó?

Lehet érvelni avval, hogy a nyelvtanárok egy igen jelentős része tevékenykedik a nyelvvizsgáztatás terén és ily módon sok információhoz hozzáfér. Az is igaz, hogy a nyelvvizsgáztatásban szakemberként részt vevők túlnyomó többségének nyelvtanári háttere van, azonban azt is tudjuk, hogy a nyelvtanárok jellemzően a szóbelizetésben vesznek részt nagy számban. Az írásbeli értékelésben és a vizsgaanyag fejlesztésében már sokkal kevesebben vehetnek részt, a külföldi illetőségű nyelvvizsgák által külföldön készített vizsgaanyagok fejlesztésében pedig még kevesebben.

A nyelvvizsgákról sok tájékoztató jelenik meg, korábban nyomtatott (Fazekas, 2004), de egyre inkább elektronikus formában. Egyrészt rendelkezésre állnak a vizsgaközpontok internetes tájékoztatói, saját honlapjaikon, másrészt a Nyelvvizsgáztatási Akkreditációs Központ (NYAK) saját honlapján (www.nyak.hu), illetve az Oktatási Hivatal (www.oktatas.hu) megfelelő oldalán. A legfontosabb információk itt, egy helyen, „központosítva” található meg, továbbá linkek a részletesebb tájékoztatást igénylők számára a vizsgaközpontok honlapjaihoz.

Az információ mennyisége azonban nem áll egyenes arányban annak hasznosságával. A NYAK honlap jellemző, kötelező, minden vizsgánál látható eleme például a megszerezhető pontok száma, amely, mint azt alább látni fogjuk, szinte semmilyen kapcsolatban nincs a vizsga minőségével. Ugyanígy, minden vizsgánál szerepel a vizsgák hossza is, percben, amely csak kevéssé informál a vizsga minőségéről. Magyar nyelven hiányosak, de angolul legalábbis igen szűk körben elérhetők a helyes gyakorlat ('good practice') dokumentumai (kiadványai), amelyek a tudományos tárgyalásnál közérthetőbb formában és röviden foglalják össze, hogy mik tekinthetők a szakterület legszakoszerűbb eljárásainak. Ilyen például a cambridge-i nyelvvizsgák dokumentuma a helyes gyakorlatról (*Cambridge English Language Assessment*, 2013), vagy a European Association for Language Testing and Assessment (*EALTA*, é. n.) irányelvei; ez utóbbi azonban magyarul is megtalálható.¹

Korábban, a helyes gyakorlat követelményeinek hazai megfogalmazásában fontos szerepet töltött be, majd érzékeny veszteség lett az Akkreditációs Kézikönyv (*Andor, Berényi, Borgulya, Dávid, Fekete, Heltai, Lengyel és Völgyes*, 1999), amelyet a NYAK részeként működő Nyelvvizsgát Akkreditáló Testületben (NYAT) a testület tagjainak több „generációja” gondozott. A kézikönyvet azonban néhány évvel ezelőtt önálló kiadványként megszüntették, azzal az érveléssel, hogy az akkreditációs rendszer már a pályázatokhoz hasonló formában működik és nincs szükség a tanító, magyarázó részletekre (Csépes Ildikó, személyes közlés). A cím megtartása mellett néhány internetes fájl, táblázat, szövegrészlet maradt belőle a www.nyak.hu-n, illetve a www.oktatas.hu-n, de ezek információtartalma már nem elegendő sem a nyelvvizsgák működésének jobb megértéséhez, sem a helyes gyakorlat elsajátításához.

A nyelvvizsgák nehezen átláthatók, még mérési szakemberek számára is, ha nem férnek hozzá a megfelelő információhoz. Az értékelési eszközöknek például szerepe lehetne a tájékoztatásban, mert a követelmények, hogy milyen a jó nyelvvizsga, belőlük kikövetkeztethetők. A NYAT bírálatra vagy felülvizsgálat lefolytatására jelenleg használt eszköze egy kérdéssor, amely csak a felkért szakértők számára hozzáférhető. A NYAT korábban használt bírálati eszköze sem áll rendelkezésre már, amely egy pontozásos szempontsor volt (*Dávid*, 2001). Az Akkreditációs kézikönyv (AK) korábbi teljes szövegű változata pl. részletesen foglalkozott a nyelvvizsgáztatás folyamatszerű követelményeivel is (*Andor és mtsai*, 1999, 54–58. o.), amelyek azonban a redukált, internetes változatban már nehezen elérhetők. A fogalomtár megfelelő kifejezésénél található hipervivatkozásra kell kattintsunk ahhoz, hogy a fogalom definíciójával megismerkedjünk.² Nem lehet tehát teljesen véletlen, nomen est omen, hogy a közelmúltban megjelent két témába vágó cikk a *Modern Nyelvoktatás* folyóirat *Műhelytitkok* rovatában látott napvilágot (*Csépes*, 2012; *Király*, 2012).

E cikk a maga egyszerű eszközeivel a fentebb leírt kérdésekre igyekszik választ adni, még akkor is, ha minőség fogalma nyilvánvalóan olyan többleteket is felölel, melyek már nem tartoznak a mérés szigorúan vett területéhez (a vizsgaszervezés minősége, a vizsgázók kiszolgálása, elégedettségüknek elérése, a vizsga PR-je, stb). A jó nyelvvizsga ismérveinek van egy adott, nem túlságosan széles köre, amelyet némi érzékkel, megfigyelőkészséggel, átlagos nyelvtanári ismeretekkel és józan ésszel is meg lehet ítélni. Ezeket kísérlem meg itt összefoglalni.

A cikk megírásának körülményei

Ez a cikk két szeminárium és több előadás tanulságaira épül. A szemináriumokon diplomás, egyetemi mesterszintű továbbképzést (MA) folytató nyelvtanárok vettek részt, akik a szemináriumvezető tanár által adott bevezetés után kiválasztottak egy-egy nyelv-

vizsgát, velük kapcsolatos információ után kutattak, majd a félév vége felé 20 perces értékelő jellegű beszámolókat tartottak a kiválasztott nyelvvizsgáról. A beszámolók fő szempontja a vizsga értékelése kellett legyen, amit az összegyűjtött ismeret- és tényanyaggal kellett alátámasztani.

1. táblázat. A szeminaristák prezentációinak megoszlása

<i>A nyelvvizsgarendszer neve</i>	<i>Hányan választották?</i>
Arma	1
BGF	1
BME	1
ECL	2
Euro	3
KITEX	1
Origo	1
Pannon	1
PROFEX	1
társalkODÓ	1
TELC	3

Amikor a szeminárium vezetője a bemutatandó vizsgákról először kérdezte a hallgatókat, hogy melyik vizsgáról szeretnének prezentációt tartani, (az egyik csoportban legalábbis) szinte mindegyikük a TELC nyelvvizsgát választotta! Kérdésre magyarázatképp elmondták, azt a vizsgát tartják jónak, amelyik könnyű (sic!), nem tart sokáig és hamar megvan az eredmény. Mivel összesen 16 szeminaristáról volt szó, az elemszám túl kicsi ahhoz, hogy messzemenő következtetéseket vonjunk le bármilyen megoszlásból. A jelenség, az érvelés és a magyarázatok azonban fontosak, mert tükrözik a szeminaristáknak mint a hozzájuk forduló diákok tanárainak szempontjait: „Melyik vizsgát javaslom diákjaimnak?” Tükrözik továbbá azt is, milyen nagy szükség van az idegennyelv-tudás mérésének jobb és mélyebb megértésére.

Végül, témaválasztásuk véglegesítése során egy bővebb és kiegyensúlyozottabb lista alakult ki – ezt mutatja az 1. táblázat. Az így kiválasztott vizsgák megoszlása azt a pedagógiai szempontot jeleníti meg, hogy a szemináriumon a lehető legtöbb nyelvvizsga kapjon figyelmet. A kérdésre, hogy jobb-e a vizsga, amelyet többen választottak, könnyebb, rövidebb és gyorsabban közöl eredményt, e táblázat alapján sem válaszolhatunk. De akkor – kérdezhetjük – honnan, hogyan lehet egyszerű eszközökkel megtudni, melyik vizsga a jobb?

Az, hogy a nyelvtanárok jobban értsék a nyelvtudás mérését, véleményem szerint igen fontos. Ha jobban megértik, el tudják magyarázni diákjaiknak, miért nem feltétlenül az a jobb nyelvvizsga, amelyik könnyű (vagy annak tűnik), miért nem feltétlenül az a jobb, amelyik elég rövid (tehát hamar haza lehet menni), illetve miért nem feltétlenül az a jobb, ahol az eredmény is azonnal megvan, például mert kifüggesztik a vizsgaterem ajtajára (Benke Eszter, személyes közlés).

Mit tehet a nyelvtanár?

Saját tájékozottságának erősítésére csak korlátozott eszközök állnak a nyelvtanár rendelkezésére, de az alábbiakat viszonylag könnyen megvalósíthatja:

- Tanulmányozhatja az általában könnyen elérhető vizsgaleírásokat (specifikáció) vagy más nyomtatott és elektronikus tájékoztató kiadványokat, szimulált vizsgák felvételeit, más szóval a hivatalos tájékoztató kiadványokat, továbbá tanulmányozhatja a vizsgaszabályzatokat, amelyek szintén hivatalos dokumentumok.
- Információt gyűjthet, személyesen vagy másoktól, megfigyeléseket végezhet. Különösen értékes lehet azoknak a kollégáknak a beszámolója, „személyes közlése”, akiket bevontak valamely vizsgaközpont munkájába. Az összegyűjtött információ természetesen kritikailag értékelni kell. Ellenőrizni kell helyességét és végig kell gondolni, mit jelent, milyen más információ támasztja alá, vagy kerül ellentétbe vele.

A fenti pontok csak nagyon általánosan fogalmazzák meg a lehetőségeket, ezért a cikk további részében, két csoportban sorra vesszük a nyelvvizsgák értékeléséhez használható (a szemináriumok során is alkalmazott, alakított) kérdéssort. A kérdéssor tíz kérdésből áll. A cikk további részében ez a kérdéssor alkotja a téma tárgyalásának sorrendjét.

Az elemzésekhez öt vizsgarendszert választottunk, azon a szinten, amelyet Magyarországon a legtöbben választanak. Ezek nem feltétlenül azok a vizsgák, amelyek a legtöbb vizsgázót fogadják, de nem is a legkisebbek, viszont együtt feltétlenül a vizsgázók jelentős hányadát fedik le. Lényegében példák, amelyek a teljesség igénye nélkül jelzik, milyen egyszerűbb technikákkal lehet a vizsgarendszerek mérési különbségeit tetten érni. Ha az olvasó ezek között nem találna meg az őt közelebből érdeklő vizsgát, bátorítjuk, hogy annak vonatkozó adataival a maga számára egészítse ki az alábbi 2, 3. és 5. táblázatot és végezze el az összehasonlítást.

Cikkünk szempontjából fontos, hogy miként határozható meg az idegen nyelvtudás. Már egy ideje hagyománynak, ortodoxiának számít, hogy a nyelvtudás egyszerre egységes és egyszerre áll jól elkülöníthető „részekből” is. Egységes, mert a nyelv adott szintű általános „tudása” (nyelvismeret, nyelvi kompetencia) áll a háttérben és egyben részekre tagolt, mert a kompetencia egyes kommunikációs irányokban, médiumokban, módzatokban és eltérő célokból történő alkalmazásai alkotják. Ezeket az elkülönülő részeket, az idegen nyelvtudás mérése hagyományainak megfelelően, általában a nyelvi készségek szerinti felosztásban mérik.³ Cikkünk szempontjából fontos továbbá a kompetencia és performancia szintén régóta használt fogalmi párja (*Chomsky*, 1965; *Hymes*, 1972). Az előbbi, a kompetencia rokonítható leginkább a nyelvtudás mint a nyelv alkalmazására való képesség fogalmával, míg az utóbbi a nyelvtudás adott kontextusban bekövetkező konkrét megvalósítását, létrehozását jelenti – ami ténylegesen elhangzik, illetve leírják. A kompetencia (nyelvtudás) az, aminek megmérésére minden nyelvvizsgának törekednie kell, azonban a kommunikatív nyelvtudás-mérés (‘communicative language testing’) elvei szerint ezt olyan feladatok alapján kell elvégezni, amelyek alapján a vizsga körülményei között performancia jöhet létre. A performancia-teszt (vizsga) tehát olyan mérési eszköz/tevékenység, amely a nyelvtudást annak megvalósulása alapján méri és ebből következtet a nyelvtudásra (kompetenciára).

A vizsgaleírás (specifikáció), mintafeladatok és a vizsgaszabályzat tanulmányozásával megválaszolható kérdések

Kellően széles és kellő mélységű mintát vesz-e a vizsga a célnyelvből?

A legtöbb nyelvvizsga több vizsgarészből áll és a négy nyelvi alapkészséget (beszédértés, beszéd-készség, írott szöveg értése és írásbeli kompozíciós készség) egymástól elkülönülten méri. Ezek a vizsgák őrzik az 1980-as évek kommunikatív mozgalmának lenyomatát is, amennyiben a beszéd-készség és írásbeli kompozíciós készség vizsgarészeket mint performancia-tesztet szervezik meg oly módon, hogy a vizsgázó életszerű feladatszabásra válaszként szóban vagy írásban kommunikációs szempontból is értékelhető szöveget alkot, összetett nyelvi (szöveges) produkcióval válaszol, amelyet értékelők (vizsgáztatók) értékelnek, összetett szempontok alapján. (*Morrow*, 1977, 1979; továbbá jó áttekintést tartalmaz *McNamara*, 1996). A fenti kérdés tehát gyakorlatiasabban és egy kicsit egyszerűbben is megfogalmazható: Méri-e a vizsga legalább a négy alapkészséget?

A vizsgarész elnevezése, például „írás-készség”, nyilvánvalóan nem nyújt arra garanciát, hogy a készségi szintet a vizsgarész valóban méri-e. Ha valamelyik vizsgarendszer négy részből áll és ezek nevei megegyeznek a négy alapkészség neveivel, még nem lehetünk biztosak abban, hogy a vizsga az idegennyelv-tudást e készségek szerint meg is méri. Egy kommunikatív idegen nyelvi mérés módszertana szerint elkészített produktív íráskészség vizsgarész általában egy (például ITK vagy BME írásbeli B2 szinten), vagy két komplex feladatot tartalmaz (például Euro), hármat vagy még többet már nagyon ritkán. Ha az egyetlen feladat mindig csak például egy levél megírása, tudnunk kell, hogy a vizsga nem méri az íráskészséget más szóba jöhető szövegfajták tekintetében. A vizsgázó nem bizonyította tudását feljegyzés, jelentés, esszé stb. megírásával. A problémát az okozza, hogy a vizsgázó idegen nyelven végrehajtja e korlátozott feladatot, viszont a vizsgaközpont értékelése az íráskészségre általánosítva kell vonatkozzék (és az érintettek így is értik). A szóbeli vizsga beszéd-készség részében hasonló korlátozottság tapasztalható, amikor a vizsgáztatóval folytatott párbeszéd és két kép egyike alapján megvalósított, önálló témakifejtés mellett egyetlen szituációs szerepjáték valósul meg (például ITK B2), holott a szóba jöhető és az adott szinten releváns szituációk típusainak száma igen nagy. A döntéshozót, lehetséges munkaadót, stb. nem csak az érdekli, hogy a bizonyítványt felmutató személy tud-e magánlevelet írni, vagy egy interaktív helyzetben megfelelni, hanem – ez érdekli csak igazán –, hogy rendelkezik-e az adott készség terén az adott szinten általános jártassággal. Ez az általánosíthatóság problémája.

A konstruktum alulreprezentáltsága

Minden mérési-értékelési, teszt és vizsgafeladat, de különösen a kommunikatív elgondolások szerint szerkesztett feladat jellemző hátránya, hogy a vizsga nyilvánvaló idő- és egyéb logisztikai korlátai közepette az idegen nyelvből vett minta és a belőle levonható/levonandó következtetések kötelezően általános jellege közötti kapcsolat problematikus. A méréssel, értékeléssel foglalkozó szakemberek tudatában vannak a problémának. Messick (1989), nem elsősorban a nyelvvizsgákra vonatkoztatva, ezt a fajta korlátozottságot a mérendő (konstruktum) alulreprezentáltságának nevezi ('construct-underrepresentation'). A mérendő alulreprezentáltsága problémájának fontosságára utal, hogy mint kíváncsi vagyok még a jóval későbbi, mérési-értékelési szakterületre rövid bevezetésnek szánt könyvecskébe is bekerült (*McNamara*, 2000, 53. o.). Az alulreprezentáltság a nem általánosítható vizsgaeredmények egyik oka lehet.

A vizsgában ténylegesen meg nem mért (meg nem mérhető) feladattípusok, műveletek és nyelvismereti elemek tekintetében csak feltételezhetjük, hogy a vizsgázó azokat is ismeri. Ha az írott szöveg értése vizsgarészben részletes értést mérő kérdésekre helyes választ ad a vizsgázó, lényegében csak következtetjük, de nem mérjük, hogy a szöveg vonatkozó részének szókinccse, nyelvtana, azaz a nyelvismeret, nyelvi kompetencia, vagy „nyelvi tartalom” (Bárdos, 2002) elemei is ismertek számára.

A vizsgaközpontok érzik azt az ellentmondást, amely az egy feladatra korlátozódó vizsgázói teljesítmény és a célul tűzött általánosító következtetés között feszül, ezért az első, kézenfekvő megoldás az (lenne), hogy több produktív feladatot kínálnak, amelyekből a vizsgázó egyet választ ki. Ilyen például a BME B2-es vizsgája, ahol a választási lehetőség hivatalos vagy magánlevél, azonban a vizsgázó ténylegesen csak egy levelet ír meg. A teljesítmények sem összevethetők így, mert nincs olyan vizsgázó, aki mindkét íráskészség feladatát megírta volna. Továbbmegy a feladat differenciálásában például az Euro, amely egy minden vizsgázóval egy kötelező feladatot oldat meg (levél), azonban a vizsgázónak egy másik feladatot is meg kell oldania.

Ha végignézzük, hogy az egyes írott szöveg értése kérdések konkrétan milyen szókinccsbeli és grammatikai ismereteket feltételeznek (mi az, ami nélkül biztos nem lesznek jó válaszok), meglehetősen kevesellni fogjuk ezeket az elemeket, és úgy érezzük, a kompetencia vizsgálata nem elég szisztematikus és módszeres. A kommunikatív módszerek szerint elkészített produktív feladatokban még nehezebb a nyelvismeret mérését jó mintavételezéssel elősegíteni. Ha az íráskészség feladat például hivatalos levél szerkesztése, arra számíthatunk, hogy abban megfogalmazódik legalább egy kérés vagy követelés, amely bizonyítékot szolgáltat arról, hogy a vizsgázó birtokában van-e az ehhez szükséges nyelvi eszközöknek. Van számos más nyelvi elem, amelyet szintén mérnie kellene a vizsgának, de ezek előfordulása ebben a feladatban csak esetleges.

A vizsgaközpontok érzik azt az ellentmondást, amely az egy feladatra korlátozódó vizsgázói teljesítmény és a célul tűzött általánosító következtetés között feszül, ezért az első, kézenfekvő megoldás az (lenne), hogy több produktív feladatot kínálnak, amelyekből a vizsgázó egyet választ ki. Ilyen például a BME B2-es vizsgája, ahol a választási lehetőség hivatalos vagy magánlevél, azonban a vizsgázó ténylegesen csak egy levelet ír meg. A teljesítmények sem összevethetők így, mert nincs olyan vizsgázó, aki mindkét íráskészség feladatát megírta volna.

Továbbmegy a feladat differenciálásában például az Euro, amely minden vizsgázóval egy kötelező feladatot oldat meg (levél), azonban a vizsgázónak egy másik feladatot is meg kell oldania. A második feladat három választható feladatból áll, amely három szövegtípust testesít meg (esszé, jelentés, narratíva vagy kritika stb.), de ezekből csak egyet kell megoldani.

Az a tanulság mindenképp levonható, hogy minél több kötelező feladattal mér a vizsga, minél jobban „fedi le” az elvárt nyelvi feladatok teljes körét, a vizsgált idegen nyelv annál kevésbé lesz alulreprezentált a feladatokban. A mérendő alulreprezentáltsága problémája miatt a jobb vizsgák jele lehet, mint második megoldás, ha a négy készség mellett további vizsgarészek mérik a készségek mögött meghúzódó nyelv-

tudást (nyelvi kompetenciát), hiszen a készséget mérő vizsgarészek eredményeiből lényegében csak részlegesen és indirekt módon lehet következtetni a nyelvismeretre. Ilyen lehet a nyelvismeret (grammatika és szókincs) és a közvetítési készség mérése.

Az irodalom kritikája

Már szinte hallom a felhördülést, miszerint ezzel a cikkel vissza akarjuk forgatni az idő kerekét, hogy vissza kellene térni a strukturalista és behaviorista szemléletű, a nyelvi elemeket különálló, analitikus módon mérő feladatsorokhoz (*Lado*, 1961). Erről nincs szó.⁴ A nyelvből vett minta, a mintavétel megfelelőségével kapcsolatos hiányosságokat (nem menti, de legalább) magyarázza, hogy a kommunikatív nyelvtudásmérés teoretikusai (Carroll, Morrow és Weir) jobbára adósok maradtak a célnyelv mintavételezésének megoldásával. Carroll (1980, 1981) és Morrow (1977, 1979, 1983, 1986) szerint meg kell határozni a reprezentatív kommunikációs feladatok körét, amelyből aztán a vizsgafejlesztés során, szem előtt tartva a mintavétel szempontjait, a vizsga számára válogatnak és mint vizsgafeladatokat megírják. Sem Carroll, sem Morrow nem foglalkozik azonban eleget avval, hogy az autentikus, kommunikatív mérési feladatok hogyan alkothatnak mérési szempontból is elfogadható, a mérendő tekintetében megfelelő mintát. Morrow (1979) és Weir (1990) az extrapoláció ('extrapolation'), a mérés eredményeinek általánosíthatósága kérdésével foglalkozik és ennyiben a nyelvből vett minta és a mintavételezés kérdése is terítékre kerül. Morrow (1979) szerint minden komplex feladat ('global task') sikeres megoldása néhány részkészség ('enabling skills') meglététől függ, ezek viszont (többnyire) más releváns komplex feladat elvégzésénél is szűkségesek. Mivel következetesen megjelennek, Morrow (1979, 153. o.) szerint a nyelv mintavételezése a részkészségek szintjén lehetségessé válik. Weir (1990, 16. o.) egy évtized távlatából már látja, hogy a reprezentatív kommunikációs feladatok meghatározása csak általánosságban megfogalmazott terv maradt és óvatosságra int, hogy várni kell addig, amíg lehetséges lesz – empirikus kutatások alapján – a reprezentatív kommunikációs feladatokat körének pontos meghatározása. Látja, hogy mivel a szóba jöhető feladatok, műveletek szinte számba vehetetlenek, így az általánosítás is igen nehéz: „Minél specifikusabb a kiválasztott feladat, annál nehezebb az azon mért egyedi/egyéni performancia általánosítása.” (Weir, 1990, 16. o.). Számunkra fontos gyakorlati tanácsa az, hogy a vizsgázó nyelvtudásának mérése a lehető legnagyobb számú feladat alapján történjen meg, súrolva akár a megvalósíthatóság és kivitelezhetőség ('efficiency') határait is (Weir, 1990, 17. o.). Még népszerűbb és még egyszerűbb módon Weir (1993, 29. o.) így fogalmaz: „[annak érdekében, hogy vizsgánk reprezentatív legyen,] nem téveszthetjük szem elől azt a lehetőséget, hogy hosszú vizsgát szerkesszünk”. Újabb, több, mint 20 év távlatából e cikk írója megállapíthatja, továbbra sem készült el a szóba jöhető kommunikációs feladatok átfogó katasztere. Az így maradt ürt töltheti be a nyelvismereti, vagy a nyelvi tartalmat mérő vizsgarész, amely hozzájárulhat az idegen nyelvből vett minta szélesítéséhez, mert mint Davies (1978, 225. o.) írja, a nyelvtan minden más nyelvi elemnél erősebb az általánosíthatóságot tekintve. A mediációs vizsgarész (közvetítés) is ugyanezen okokból hasznos, mert az idegen nyelvtudás további aspektusát, egy további készségét méri ez által a vizsga. A nyelvvizsga akkreditáció kezdetekor kidolgozott feltételek (*Andor és mtsai*, 1999), alapul véve az akkor érvényes 71/1998-as kormányrendeletet, két lehetséges nyelvtudásképet (célképzetet) vettek számításba. Egyrészt lehetséges (volt) az idegen nyelvet a négy alapkészségre bontva vizsgálni – ez volt a minimum – másrészt lehetséges volt nyelvvizsgát tervezni a nyelvismerettel és/vagy közvetítéssel kibővítve.

Mindezek alapján megfogalmazható egy értékítélet. Az a vizsga lehet jobb, amelyik több vizsgarész segítségével, azaz több szempontból vizsgálja a vizsgázó nyelvtudását,

mert ily módon a célnyelvből vett minta bőséges és sokrétű lehet. Ennek megoldása lehet több produktív feladat megoldatása, vagy kevesebb (de egynél feltétlenül több) produktív feladat megoldatása mellett a nyelvismeret és közvetítési készség elkülönült mérése. Minél több információt gyűjt a vizsgázóról, a vizsgázó tudása annál nagyobb biztonsággal mérhető, számszerűsíthető.

A vizsgázó nyelvtudásából vett minta elég széles és átfogó-e?

Ahogy a nyelvvizsgának hűen kell tükröznie a mért idegen nyelvet, a vizsgázótól nyert minta (a kérdésekre adott válaszai) elég széles és átfogó kell legyen, ideértve a szóbeli és az írásbeli produkciót is. Előfordulhat azonban, hogy bár a vizsga részei minden készséget mérnek, az egyes vizsgarészek önmagukban alig tekinthetők az adott készség mérésének, mert túl kevés itemből állnak. Mielőtt továbblépnénk, meg kell határozni az 'item' szó tartalmát, mert nem mindig azonos a 'feladat' szakszó jelentésével. Nem csak az elkülönült elemekből álló tesztek állnak itemekből, hanem a kommunikatív módszertan szerinti mérésben használatos produktív feladatok is állhatnak egy vagy több itemből. Minden egyes pontszerzési lehetőség item, mert újabb és újabb esélyt ad a vizsgázónak, hogy megmutassa, mit tud, és hogy nyelvtudásának pontos megítélhetősége érdekében elegendő információt szolgáltatson. A legjobb kifejezés talán a pontszerzési esély, amint azt Hughes (2003, 44–45. o.) is 'fresh start'-nak nevezi, mert egy újabb item megválaszolása, azon túl, hogy további információt szolgáltat a vizsgázóról, újabb esélyt is ad neki arra, hogy korábbi hibáit újabb pontok megszerzésével ellensúlyozza. A *Nyelvtudásmérési terminológiai szótár* szerint is az item „minden olyan mérési pont/egység egy feladaton belül, amely külön pontot vagy pontokat kap” (Fekete és Óvári, 2004, 26. o.). Az alábbiakban néhány ismert vizsgáról a bárki által megtalálható tájékoztatást vesszük számba.

Míg a vizsgaleírások általában explicit módon meghatározzák legfeljebb hány pont szerezhető egy-egy feladatra, a pontszerzési esélyek számát inkább ki kell következtetni. Mint fentebb leírtuk, a NYAK honlapján csak a megszerzhető pontok száma szerepel, ami nem a legalapvetőbb. Az egyes vizsgaközpontok lapjain már bővebb tájékoztatást kapunk: a megszerzhető pontok száma mellett például az ITK honlapján a szóbeli vizsga részeinek és egészének időtartama szerepel, továbbá a sikeres vizsgához megkövetelt minimum pontszám. Az írásbeliről ugyanezek az információk találhatóak meg, valamint a nyelvismeret esetében szerepel az „egységek” száma, de a közvetítés, íráskészség és írott szöveg értéke esetében csak a szöveg terjedelme (leütés, sor) szerepel. A BME honlapján ugyanezeket az információkat megtaláljuk, azonban a „feladatok száma” oszlopban a „kérdések”, „gondolati egységek”, „feladat”, „címszók”, „állítások”, „információs egységek”, „kiegészítések”, „hozzárendelések” sőt, az íráskészség esetében, az értékelési „szempontok” számának meghatározása szerepel a tájékoztató táblázatban. Az Euro honlapján áttekinthető táblázatokat közöl, melyek az előbbi kettőhöz hasonló módon a megszerzhető pontok számáról és időkeretekről tudósítanak, de szöveges leírást is ad, továbbá tájékoztat a feladatok számáról, némely esetben a szövegek, illetve egyes esetekben a kérdések számáról nyújt tájékoztatást. A TELC honlapján a vizsgaleírások alatt, először táblázatos formában, hasonlóan az előbbiekhöz, az összpontszámokról, időkorlátokról és a feladatokról tájékoztat, lejjebb azonban részletesebb leírás is található, amely megadja a vizsgarészek egyes feladatait és a kérdések száma több feladatnál szerepel. Az ECL szintén az összpontszámokat közli, időtartamot csak a beszéd-készségnél határoz meg. Részlegesen megismertet a mintafeladatokkal is, amelyeket azonban teljes egészében nem tekinthet meg az érdeklődő. Az akkreditált vizsgákról való tájékoztatás sokfélesége tehát nyilvánvaló, azonban a pontszerzési esélyek száma, amely alapján e vizsgák összehasonlíthatók lennének, e vizsgáknál nem található meg.

A megszerzhető pontok száma azért nem tartozik a legfontosabb információk közé, mert ha két vizsgát összevetünk, és az elérhető pontszám maximuma mindkettőn például 100 pont, a pontszerzési esélyek száma nem szükségszerűen azonos. A megszerzhető pontok számát jelentősen befolyásolja továbbá a vizsgarészek eltéréseit kiegyenlítő súlyozás is. Eltérő súlyozás nélkül az úgynevezett dichotóm feladattípusok esetében (ha a helyes válasz egy pontot ér, míg a helytelen válaszra nem adnak pontot) nincs különbség a megszerzhető pontok és a pontszerzési esélyek között, de a nem dichotóm feladatok esetében már lehet eltérés. A következő példa segíthet megvilágítani a különbséget. Két beszédértés vizsgát hasonlítunk össze. Mindkettőre maximum 20 pont adható. Az egyik 20 dichotóm feladatból áll, a pontozásban csak 1 és 0 születhet, míg a másikonál (mint például az ITK vizsgáján) a 20 pont 3–4–5 vagy 6 kérdés között oszlik meg. Az összehasonlításban világos, hogy az előbbi 20 pontszerzési esélyt nyújt, míg a másik csak kevesebb pontszerzési esélyt ad akkor is, ha kérdésenként a maximum és minimum pontok között a közbenső pontértékek is adhatók.

A pontszerzési esélyek számításának módját tehát a következőképp határozhatjuk meg: az úgynevezett objektív, elkülönült elemekből álló (dichotóm és többfokozatú) itemek alkotta vizsgarészek esetében megegyezik az itemek számával. Produktív, integratív feladatok és vizsgarészek esetében a pontszerzési esélyeket a feladatok számának és az értékelési szempontoknak, továbbá a feladatokat osztályozó értékelők (például kettős értékelés esetén) számának szorzata adja (2. táblázat 4–6. oszlopa). Például az olyan íráskészség feladatsor, amely két feladatot kötelezően tartalmaz és mindkettőt öt szempont alapján értékeli, továbbá mindkettőt kettős értékelésnek vetik alá, összesen 20 pontszerzési esélyt jelent ($2 \cdot 5 \cdot 2$).

2. táblázat. A pontszerzési esélyek összevetése néhány írásbeli nyelvvizsgán

Vizsga	Megszerezhető pontok	Rendelkezésre álló idő (perc)	Pontszerzési lehetőségek			Összesen
			Írásbeli recepció és produkció (egynyelvű)	Nyelvismeret	Közvetítés	
ITK	80	180	$8 + (1 \cdot 4 \cdot 2) = 16$	50	10?	$66 + 10? = 76$
BME	120	205	$5 + 10 + (1 \cdot 4 \cdot 2) = 23$	$20 + 30 = 50$	$(1 \cdot 3 \cdot 2) = 6$	79
Euro	75	130	$20 + (2 \cdot 6 \cdot 2) = 44$	nincs	$30 + (1 \cdot 1 \cdot 2) = 32$	76
TELC	150	120	$20 + (1 \cdot 3 \cdot 2) = 26$	20	nincs	46
ECL	200	85	$20 + (2 \cdot 5 \cdot 2) = 40$	nincs	nincs	40

A fentiek alapján megfogalmazható az, hogy az a vizsga tekinthető jobbnak, amelyik több pontszerzési esélyt nyújt a vizsgázónak, azonban a vizsgák eltérő jellege miatt az ítéletalkotás meglehetősen összetett. Az egynyelvű vizsgákkal azonosítható írásbeli recepció és produkció (interakciós) tagolásban, amely tehát a készségek szerinti mérés miatt a kommunikatív módszertan szerint a legmodernebbnek tekinthető, az Euro és az ECL a legerősebbek (2. táblázat), míg az BME, de különösen az ITK meglepően kevés, 23, illetve 16 pontszerzési esélyt nyújt. (Az ITK írott szövegértés 8-as értékét a honlapon publikált mintafeladat alapján írtuk be, de tudni kell, hogy a kérdések száma változhat és általában ennél kevesebb.) A nyelvismeret és/vagy a közvetítés hozzáadásával az ítélet változik. Megállapítható, hogy a BME és az ITK az „élre tör”, összességében a legtöbb pontszerzési esélyt az ITK és a BME nyelvvizsgák kínálják. Megjegyzendő, hogy az ITK vizsga közvetítési feladatán a pontszerzési esélyek száma nem állandó. A dolgozatok javításához a teamvezető a javítást végzők munkáját segítő tanulmányt készít, melyben mondatokra bontja a legvalószínűbb megoldást (saját mintamegoldását) és a mondatokhoz pontokat rendel. Ezek a mondatok tekinthetők az ITK közvetítés feladata

pontszerzési esélyeinek. Mivel az egyes szövegek eltérő számú mondatból állhatnak, itt csak a legvalószínűbb 10 mondatos szöveget szerepeltethetjük és a bizonytalanságot kérdőjellel fejeztük ki (2. táblázat). E táblázat alapján az a megfigyelés is megfogalmazható, hogy egyes vizsgák, ahelyett hogy az írásbeli recepciós és produkciós vizsgarészeket újabb feladatokkal egészítenék ki, inkább a nyelvismeret és közvetítés mérését vetik be a készségalapú vizsgarészek helyett. Míg az írott szöveg értése vizsgarészek 20 pontszerzési esélye amolyan „ipari normának” tűnik (a vizsgált öt vizsgából háromnál ez szerepel), az is látható, hogy a legkiegyensúlyozottabb szerkezete az ECL-nek van. Az Euro 6 értékelési szempontja segítségével maximalizálja a pontszerzési esélyek számát. Érdekes tehát a pontszerzési esélyeket a fenti bontásban más, itt nem tárgyalt vizsgáknál is értékelni, mert kiderül, mely vizsgák nyújtják a legtöbb ilyen esélyt, mint ahogy az is láthatóvá válik, hogy melyek egészítik ki mérési eszközeiket a nyelvismeret és közvetítés mérésével, és melyek azok, amelyek szinte a készségek helyett mérik a nyelvismeretet és közvetítést.

Minden mért készség, illetve kompetencia tekintetében is elegendő információt gyűjt-e a vizsga a vizsgázó nyelvtudásáról?

A kérdést úgy is megfogalmazhatnánk, hogy minden vizsgarész elég hosszú-e, azonban a hosszúság nem tűnik sem elég egzaktnak, sem pedig nagyon tudományosnak. A követelményt szabatosabban úgy fogalmazhatjuk meg, hogy minden vizsgarész elegendő információt nyújt-e a vizsgázó idegen nyelvtudásáról. Van-e olyan vizsgarész, amelynek segítségével nem gyűjthető elegendő információ a vizsgázó tudásáról?⁵

A pontszerzési esélyek mellett további jellemzője a feladatsoroknak a mérési skála/skálák potenciális információtartalma. (A skálák segítségével a vizsgázók tudását, valamilyen szempont szerint, fokozatokban, pontokkal értékelhetjük (0–1–2–3 vagy még több pont. Még a legegyszerűbb dichotóm kérdés is egy kétfokozatú (0 vagy 1 pont) skálának tekinthető.) A skálák száma, fajtái, szerkezetük döntő módon befolyásolják a különféle feladattípusok segítségével nyerhető információt. Az összegyűjthető információ mennyiségét is alapvetően a pontszerzési lehetőségek száma határozza meg, azonban az információ tekintetében az egyes feladattípusok eltérései már jelentősek. Általában igaz az, hogy az elkülönült elemekből álló úgynevezett objektív itemek (feladatok) viszonylag kevés információt gyűjtenek össze. Az állítás alapja az, hogy például a dichotóm itemek csak két fajta információt nyújtanak; vagy azt jelzik, hogy a vizsgázó helyesen válaszolt (1 pont), vagy azt, hogy helytelenül (0 pont). Mivel kevés információhoz jutunk egyetlen dichotóm, 0 vagy 1 ponttal értékelt itemből, a megfelelő mennyiségű információhoz nagyszámú dichotóm item megoldásán keresztül jutunk el. Az ECL, TELC vagy Euro vizsgák írott szöveg értése vizsgarészei 20 pontszerzési lehetőséget kínálnak. Az ebből számítható információs érték mindhárom vizsga esetében 40, mert információt a 0 pontos válasz is szolgáltat. A dichotóm itemek információs értékét tehát az itemek számának kétszerese adja. Vannak olyan elkülönült elemekből (itemekből) álló feladattípusok is, amelyek segítségével itemenként ennél több információ is gyűjthető. A példáink között szereplő ITK írott szöveg értése feladatsora ilyen feladat. A honlapon látható mintafeladat 8 kérdést tartalmaz és mivel az egész feladatra összesen 25 pont szereshető, az egyes itemek, kérdések eltérő számú pontot érnek. Minden item 0 pontjának információs értékét itt is be kell számítani, így azt a megszerezhető 25 pont és a 8 item 0 pontjai számának összege (33) adja ki.

Több, gazdagabb információ nyerhető többfokozatú ('polytomous') értékelési skálák alkalmazásával, amelyek alkalmazására produktív feladatok esetében kerül sor. A vonatkozó értékelési skáláknak több, mint két skálapontja van és legalább 0,1 vagy 2 ponttal,

de még inkább 0,1,2 és 3 vagy még több ponttal értékelhetők. A pontszerzési lehetőségek száma és a hozzájuk tartozó skálák fokozatainak száma együttesen jelezhetik, az egyszerű tanár számára is, melyik feladatsor, vizsgarész vagy nyelvvizsga nyújtja a legtöbb információt a vizsgázó idegen nyelvtudásáról. Míg a produktív feladattípusok esetében a pontszerzési esélyt a kötelezően megoldott feladatok számának, az értékelési szempontok számának és a külön pontozó értékelők számának szorzata adja, az információs értéket a pontszerzési esélyeknek az értékelési skálák fokozataival kibővített szorzata adja. Fentebb az ITK, a BME, Euro és ECL esetében az íráskészség pontszerzési esélyeit ($1*4*2$) 6-tal szorozzuk meg, mert az értékelési skála hatfokozatú, míg a TELC esetében csak 4-gyel, mert mindhárom skála csak négyfokozatú. A BME közvetítési feladatában a 3 szempontot szintén 6-tal szorozhatjuk az ötfokozatú skálák miatt. Az Euro közvetítési vizsgarésze 30 dichotóm itemet alkalmaz, továbbá ehhez hozzáadódik az egyetlen szinten hatfokozatú skála nyújtotta információ, melyet két értékelő egymástól függetlenül alkalmaz.

Mint látható, az információ számbavétele jelentősen árajzolja a vizsgákról alkotta képet. A pontszerzési lehetőségek tekintetében nagyjából azonos kategóriába tartozó ITK, BME és Euro közül az első kettő az információ tekintetében leszakad az Eurótól, míg az ECL elhagyja a TELC-et. Feltűnő az is, hogy a produktív készségek alapos és részletes mérése mennyivel több pontszerzési lehetőséget és információt szolgáltat. Érdekes tehát rászánni az időt és kiszámolni a pontszerzési lehetőségek számát és az információs értéket. Amennyiben a pontszerzési lehetőségek új esély jelentenek a vizsgázónak, úgy a vizsgából nyerhető bőséges információ erősíti a vizsga „fairségét” is.

3. táblázat. Az információs érték számítása néhány írásbeli nyelvvizsga esetében

Vizsga	Megszerzhető pontok	Összes pontszerzési lehetőség	Információ			
			Írásbeli recepció és produkció (egynyelvű)	Nyelvismeret	Közvetítés	Összesen
ITK	80	76	$(25+8)+(1*4*6*2)=81$	$(50*16/50)=16$	$(10*3)=30?$	$81+16+30=127$
BME	120	79	$(5*2)+(10*2)+(1*4*6*2)=78$	$(50*25/50)=25$	$1*3*6*2=36$	$78+25+36=139$
Euro	75	76	$(20*2)+(2*6*6*2)=184$	–	$(30*2)+(1*1*6*2)=72$	$184+72=256$
TELC	150	46	$(20*2)+(1*3*4*2)=64$	$(20*2)=40$	–	$64+40=104$
ECL	200	40	$(20*2)+(2*5*6*2)=160$	–	–	160

A pontszerzési esélyek, az információ és a mérési hiba összefüggése

Megfogalmazható azon összefüggés is, hogy minél több pontszerzési lehetőséget, illetve információt nyújt a vizsga, annál nagyobb pontossággal lehet a vizsgázó tudását meghatározni. Minél több információt tud a vizsga nyújtani, annál pontosabb lehet a mérés. Fordítva is igaz: minél kevesebb az információ, annál nagyobb a pontatlanság. Az alábbiakban azt mutatom be, hogy azoknak a vizsgarészeknek (feladatsoroknak), amelyeknek információs értéke magas, átlagos mérési hibaértéke általában – kívánatos módon – alacsony.

A vizsgázó nyelvtudásáról gyűjthető információ mennyisége és gazdagsága a mérési hiba fogalmához vezet bennünket. Mivel a vizsgázó tudását nem értékelhetjük minden lehetséges helyzetben, a vizsgában idő sincs erre, ezért mintát kell venni. Továbbá a nyelvtudás, más pszichológiai jelenségekhez hasonlóan, belső jellemzőnek tekinthető, amely senkinek sincs kívül a „homlokára írva”. A nyelvtudást tehát elő kell hívni, meg kell jeleníteni feladatok, kérdések, vizsgáztatók segítségével, de még így is, erőfeszítésünk eredménye nem több, mint szakszerű becslés, amelyet mindig körülleg valamilyen mértékű bizonytalanság. Ezt az utóbbit nevezzük mérési hibának (Horváth, 1993; Szokolosky, 2004). Tegyük fel, hogy diákjaink számára kiválasztjuk azt a feladattípust, amelyet a lehető legobjektívebbnek tartunk, megírunk 20 feladatot (20 pontszerzési esély, egyenként maximum 1 ponttal) és megíratjuk a diákokkal. Kijavítjuk a tesztet, összeadjuk a pontokat, érdemjegyeket adunk, de valószínű, hogy a 14 pontos és a 15 pontos teljesítmények között nincs számottevő különbség, mert a pontszerzési esélyek alacsony száma miatt a „körüllengő bizonytalanság” nagy, nagyobb lehet, mint a jelzett pontkülönbség. Ha e bizonytalanság mértéke például 2 pont, képzeletbeli körzónkkel 2 pont sugarú kört kell rajzoljunk mind a 14, mind a 15 pontos eredmény köré, mert a mérési hiba plusz/mínusz értendő. Látni fogjuk, hogy a két kör egymást jelentősen átfedi, ami annyit tesz, hogy legalább 68 százalék esély van arra, hogy a 14 pontos eredmény, másik hasonló teszt alapján 15 pontos is lehet és fordítva. A mögöttes tudást, hogy a vizsgázó tényleg mennyire jól tud az idegen nyelven, más feladatok alkalmazása esetén is ennyire jó-e (vagy gyenge) a tudása, tehát nem tudjuk elég pontosan megállapítani. A nyelvvizsgák előnyben vannak az egyszerű tanárral szemben, mert komplex módon, azaz több vizsgarészen mérhetik a nyelvtudást, továbbá sokkal több vizsgázóra is számíthatnak. Több száz, vagy több ezer vizsgázó válaszait gyűjtik össze, de a probléma alapvetően azonos a számukra is. Ha több vizsgarészből adjuk össze a pontokat, a bizonytalanság, a mérési hibaértékeket is össze kell adjuk, a nagyobb létszámok következtében mégis kisebb értékekre számíthatunk. A mérési hiba fordítottan arányos a vizsgázótól vett mintában található információ mennyiségével és sokrétűségével.

Az Euro nyelvvizsgából vett adatokkal illusztrálom az információ és a mérési hiba összefüggését. A 4. táblázatban a 2009 márciusa és 2013 augusztusa között lefolytatott B2 szintű vizsgák három vizsgarészének átlagos mérési hibaértékei szerepelnek. Az átlagos hibaérték úgy értendő, hogy az eredményszámításhoz használt szoftver (Linacre, 2013) minden egyes vizsgázónak egyéni hibaértéket számít. Az itt közölt átlagok a vizsgánként 1000–1500 vizsgázó egyénileg számított értékeinek átlagát jelentik meg. A beszédértés és írott szöveg értése vizsgarészeket azért érdemes összevetni, mert receptív készségeket mérnek és szerkezetük szerint nagyon hasonló vizsgarészekről van szó. Egymástól elkülönült objektív elemekből, dichotóm itemekből állnak. A harmadik, úgynevezett szubjektív vizsgarész jelentősen eltér az előbbi kettőtől. Produktív készséget mér, amelyet két értékelő egymástól elkülönülten értékel hat analitikus skála segítségével. A táblázatbeli hibaértékek pontoknak tekinthetők, amelyek nagysága összevethető a vizsgarészenként megszerezhető 25 ponttal. A 2009. márciusi beszédértés vizsga átlagos hibaértéke például 0.5 (fél) pont a megszerezhető 25-höz képest, míg a 2010. decemberi írott szöveg értés vizsgarész átlagos mérési hiba értéke 0.73, azaz majdnem háromnegyed pont szintén a megszerezhető 25-ből.

A vizsgából nyerhető információ tárgyalása során nem feledkezhetünk meg a súlyozás szerepéről sem. A vizsgaközpontok az általános nyelvvizsgák vizsgarészeinek egyenlő súlyozására törekednek, vagyis a vizsga egyes részei az eredmények meghatározása szempontjából egyformán fontosak. Helyes gyakorlatról van szó, mivel általános (nem szaknyelvi) vizsgát oly sokan oly sokféle céllal tesznek, így a vizsga eredménye is bármilyen célra kell felhasználható legyen. Az egyes készségek viszont jelentős eltéréseket mutatnak a tekintetben, hogy milyen eszközökkel mérhetők meg, így az egyenlő súlyo-

zást matematikai eszközökkel kell megvalósítani. Ha egyes vizsgarészek pontszámait ennek érdekében felszorozzák, információ nem vesz el, de nem is keletkezik több. Ha viszont sávosan, egy táblázat alapján lefelé váltják át a nyerspontokat vizsgapontokká, az információ egy része elvész. Az itt tárgyalt vizsgák közül ez történik az ITK és a BME vizsga nyelvismereti feladatsora esetében. Például az ITK 50 ítemes tesztjéből 0 és 15 pont közötti eredmények születnek úgy, hogy néhány szomszédos ítem (például 48, 49 és 50) mindegyike 15 pontot ér. Ez magyarázza, hogy a 3. táblázatban az ITK nyelvismereti feladatsorának miért csak 16 az információs értéke. Hasonló eljárás áldozata a BME nyelvismereti feladatai is. Ha ehhez hozzávesszük azt, amint azt fentebb láttuk, hogy a dichotóm itemekből viszonylag kevés információ nyerhető, továbbá azt a tény, hogy nehéz ilyen feladatsorokat jó minőségben előállítani, megkérdőjeleződik az eljárás értelme és egy olyan elemzésben, amelyik a vizsgából nyerhető információt helyezi a középpontba, a vizsga leértékelése következik.

4. táblázat. Az átlagos mérési hiba az Euro vizsga három vizsgarészében

Vizsga		Beszédértés	Írott szöveg értéke	Íráskészség
2009	március	0,5	0,73	0,55
	június	0,66	0,66	0,53
	szeptember	0,79	0,73	0,54
	december	0,8	0,85	0,63
2010	március	0,79	0,76	0,5
	június	0,71	0,67	0,49
	szeptember	0,79	0,77	0,56
	december	0,99	1,09	0,53
2011	március	0,82	0,86	0,55
	június	0,91	1,05	0,61
	szeptember	0,76	1,12	0,52
	december	0,81	0,86	0,59
2012	március	0,96	1,07	0,62
	június	0,99	0,87	0,64
	július	0,84	0,83	0,54
	szeptember	0,98	1,04	0,57
	december	0,89	1,21	0,51
2013	március	0,86	0,96	0,62
	június	0,95	1,2	0,62
	július	0,73	1,01	0,5
Átlag		0,827	0,917	0,561

Milyen megfigyeléseket tehetünk a 4. táblázat alapján? A táblázatban szereplő 21 Euro vizsga alapján számított legalacsonyabb átlag az íráskészség esetében mutatkozik (0.561 pont). Ennél nagyobb a beszédértéshez kapcsolható átlag (0.827), míg a legnagyobb az írott szöveg értéke esetében (0.917). Mielőtt mélyebben értelmeznénk a megfigyelt különbséget, felvetődik a kérdés, hogy ezek a különbségek mennyire számottevőek: előfordulhat-e, hogy a megfigyelt különbségek csupán a véletlen műve? Erre a kérdésre a statisztikai szignifikancia vizsgálata adhatja meg a választ, először egy ANOVA, majd három t-próba segítségével. Az ANOVA a három átlagot együtt vizsgálja, és csak arra keresi a választ, hogy a megfigyelt vizsgarészek számsorainak három átlaga számottevő különbség lehet-e. Az ANOVA eredménye szignifikáns volt ($F(2) = 43,296$ $p < 0,01$), ami viszont még nem zárja

ki, hogy a három számsor közül akár csak egy kiugró átlag miatt szignifikáns az eredmény és a három közül kettőnek a különbsége viszont ne legyen számottevő. Mindhárom t-próba is szignifikánsnak bizonyult (t írott szöveg értés-írás-készség (19) = 9,427 $p < 0,01$; t beszédértés-írás-készség (19) = 3,004 $p < 0,01$; t beszédértés-írás-készség (19) = 10,865 $p < 0,01$), ami annyit tesz, hogy a megfigyelt átlagkülönbségek elég nagyok ahhoz, hogy a véletlen szerepének ne engedjünk több, mint 5 százalék esélyt.

Az átlagos mérési hiba eltérései magyarázhatók a vizsgarészek szolgáltatott információi különbségei alapján és alátámasztják az információról fentebb mondottakat. A legkisebb mérési hiba, nem csak az átlagot tekintve, hanem a vizsgált időszak tendenciáját is figyelemmel kísérve, az íráskészséghez köthető. Ugyanakkor ennek a vizsgarésznek a legmagasabb az információs értéke (144). Ennél nagyobb mérési hibát figyelhetünk meg a beszédértés esetében, amelynek információs értéke már jóval kisebb (50). A legnagyobb, de a beszédértésnél nem sokkal nagyobb átlagos hibaérték az írott szöveg értéséhez kapcsolódik, amelynek információs értéke 40. A hibaérték tehát fordítottan arányos az információs értékkel. Az utóbbi kettő viszonylag kis különbsége avval a szerkezeti különbséggel magyarázható, hogy a beszédértés vizsgarész 25, míg az írott szövegértés vizsgarész csak 20 pontszerzési esélyt kínál. Hozzá kell tennünk, más tényező is hat a mérési hibára. Ilyen pl. a vizsgázók tömegének, a populáció kis vagy nagy szóródása. Ha eleve kicsik a vizsgázók közötti különbségek, nem várható el, hogy azokat a vizsgafeladatok hozták majd létre, így a mérési hibaérték is szükségképp nagyobb lesz. E tesztek esetében azonban azt nem szabad elfelejteni, hogy jelentős részben ugyanazok a vizsgázók tették le mindhárom vizsgát. Tény az is, hogy a mérési hibaértékek a 4. táblázatban az egyes vizsgaidőszakok között hullámzást mutatnak, valószínűleg a vizsgázói csoportok eltérő szórása függvényében, esetleg az ítemek minősége tekintetében, azonban a vizsgák száma (21) ellensúlyozza ennek a hullámzásnak az esetlegességeit.

A nyelvtanár nyilvánvalóan nem tudja a mérési hiba értékét megállapítani, és így nem vonhat le abból következtetést a vizsga minősége tekintetében, mert az adatokhoz nem fér hozzá. A weblapon vagy nyomtatott tájékoztatókban kiadott információhoz viszont hozzáfér – e sorok írója tudatosan nem is használt más forrást a fenti néhány vizsga bemutatásához. A nyelvtanár nem tud megbízhatóságot sem számítani, viszont ki tudja számítani és össze tudja vetni a pontszerzési lehetőségek számát, az információs értékeket és ennek alapján megbecsülheti, mennyiben végezhetnek alapos mérést az őt érdeklő vizsgán.

Tapasztalati úton, de figyelemmel a mérési hiba fenti, jellemző értékeire is, a komplex vizsgák esetében, minimum 25, de inkább még több pontszerzési esély kívánatos vizsgarészenként, illetve készségenként. Az Euro íráskészség vizsgája mutatja a hibaérték (és az információ) kívánatosan alacsony értékét. A fél pont körüli érték azért megfelelően alacsony, mert ha a ténylegesen mért egyszerű, egész pontértékhez (nyerspontok) hozzáadjuk (vagy kivonjuk, tehát beszámítjuk) a hibaértéket, a pontszám még nem kell megváltozzon. Ehhez képest a fél pontnál nagyobb érték felveti annak a kérdését, hogy a mért, például 16 pontos teljesítmény 17 (vagy 15) pontos teljesítménynek tekinthető-e inkább, hiszen a mérési hiba annak becslése, hogy a vizsga inherens tökéletlenségei következtében mennyivel „mérhettük el” a vizsgázó tudását: mennyivel kaphatott volna több vagy kevesebb pontot, ha a mérés eszköze tökéletes lenne?⁶ A 4. táblázatban az Euro íráskészség vizsgarészének információs értéke jelenleg 144. Mivel az Euro 2013-ban változott, helyesebb talán, ha a célul tűzhető mérési hiba nagyságának számításánál az a szerkezetet vesszük figyelembe, amely a vizsgált időszak nagyobb részében érvényben volt, és amely időszakban az itt felhasznált adatok keletkeztek. Ebben az időszakban a $(2*3*6*2+2*1*11*2)$ képlettel kell számoljunk, ami 116-os információs értéket jelent. Elgondolkodtató, hogy ugyanezt az információs értéket csak 58 dichotóm ítemmel érnék el (vesd össze: ITK nyelvismereti feladatsor)! Ha 0–1–2 pontokkal osztályozott ítemeket

használnánk, ugyanennek az értéknek az eléréséhez, 39 item is elegendő lenne. Ennyi kell, mert az idegen nyelvből, ill. a vizsgázó tudásából kivett mintának nemcsak a nyelvtudás egésze, hanem az egyes készségek, nyelvismereti elemek tekintetében is kellően átfogónak kell lennie és az egyik készségnél (vizsgarészben) kivett minta erényei nem kompenzálják a másik készségből vett minta gyengeségeit.

A fentiek alapján megállapítható az is, hogy a legtöbb magyarországi nyelvvizsga beszédértés és írott szöveg értése vizsgarésze sajnálatos módon messze elmarad a vizsgázótól vett minta nagysága tekintetében. Kevés itemmel vajon jól meg lehet-e mérni a vizsgázó tudását? A hazai vizsgák szerény kivitele különösen szembeötlő, ha az amerikai illetőségű TOEIC vizsga megfelelő részeivel vetjük össze, amelyek egyenként 100 pontszerzési esélyt kínálnak! Annak ellenére, hogy a TOEIC drága vizsga, felvethető, hogy az alacsonyabb árfekvésű magyarországi vizsgák díját megfizető vizsgázó korrekt elbánásban részesül-e. Nem állítjuk a TOEIC példa alapján, hogy a magyarországi vizsgáknak is készségeként 100 pontszerzési esélyt kellene nyújtani, de az információs értékről fentebb leírtak alapján azt igen, hogy 20-nál feltétlenül többet, 40-50-60 pontszerzési esélyt kínálniuk illik.

A „lakmusz”: mennyire jó a beszédértés vizsga?

Amint a kémiaóráról kölcsönzött metafora jelzi, végezzünk „gyorstesztet”, ha nincs időnk az összes vizsgarészt végigkutatni. Nézzünk meg egyetlen vizsgarészt; azt, amely ma, Magyarországon, a nyelvvizsgáztatás terén, sűríti a lehetséges előnyöket és hátrányokat. Beszédértés vizsgát előállítani komoly feladat és „drága mulatság”. A feladatanyagok különféle médiaforrásokat, szövegtípusokat, kommunikációs módozatokat (monológok, dialógusok) kell felölelnie ahhoz, hogy kellően átfogó és mély mintát vegyen a vizsgázó tudásából. A feladatlapnak szinkronban kell lennie a hanganyaggal, mert a vizsga során a vizsgázó nem haladhat saját sorrendje, munkatempója szerint. Munkamódszerének ahhoz a sorrendhez, sebességhez kell igazodnia, amivel a hanganyagot (régebben kazettáról, újabban CD-ről, illetve hangfájlokból) lejátszzák neki. A hangminőség, a hallhatóság, teremakusztika, vagy adott esetben a fejhallgatók minősége mind fontos és igen komoly szakmai, módszertani kihívás elé állítja a feladatfejlesztőket. Nem csoda, hogy a vizsgák egy része a beszédértés racionalizálására, minimalizálására törekszik. Az egyik lehetőség az, hogy a vizsga eleve kevés szöveg- és hangzóanyagra épül, ami magával vonja azt is, hogy alacsony az itemek és pontszerzési esélyek száma is. Ennek kevésbé nyilvánvaló változata az, amikor a kevés szöveghez egyenként több (sok) itemet kapcsolnak: az ECL például összesen 2 szöveghez egyenként 10 itemet kapcsol, vagy az Euro, amely egy szövegéhez 9, egy másik szövegéhez pedig 10 itemet. Az ilyen gyakorlat azért problematikus, mert a készség és a hangzóanyagok természete szerint gyakran csak viszonylag kevés (3–4–5) kérdés adódik egy egyébként minden tekintetben megfelelő szövegnél. Az ideális az lenne, ha vizsgánként viszonylag sok, akár 6–7, vagy még több szöveget alkalmaznának a vizsgaközpontok. Ha ez teljesülne, elég lenne szövegenként körülbelül 5 item is ahhoz, hogy a pontszerzési esélyek száma és az információs érték megfelelőbb legyen. Egy másik lehetőség az, mint az korábban gyakorlat volt például a BME-n, hogy adott hangzóanyaghoz két feladatsort is készítettek (a megfelelő terem hiánya és a hanganyag elkészítésének magas költsége miatt) és a vizsga során a hanganyagot két egymást követő vizsgázói csoportnak játszották le. A vizsgabiztonság fenntartása érdekében a két csoportot „zsilipeléssel” különítették el (*Nyelvvizsgát Akkreditáló Testület*, 2003).

Ítéletalkotásunk során ne veszítsük szemünk elől azt a tény, hogy a beszédértés készsége még ma is egy nem minden tekintetben megértett, viszonylag keveset kutatott készség (*Buck*, 2001), azonban feltehető, hogy ha a vizsgaközpont a beszédértés vizsgálatának nehéz (és drága) feladatát jól oldja meg, a vizsga többi részében sem végez rossz

munkát. Bár ma már nem lehet nyelvvizsgát akkreditálni a nélkül, hogy bármelyik szintje ne tartalmazna külön beszédértés (hallott szöveg értése) feladatsort (vesd össze: Fazekas, 2004, 42. o.), a beszédértés vizsgák még mindig nagy különbségeket mutatnak (5. táblázat). Az ITK egyetlen (!) szöveghez kapcsolódó itemsor alapján értékeli, míg a BME és az ECL kettő, a TELC és az Euro három szöveget alkalmaz. E két utóbbi mintafeladat-soraiból az is kitűnik, hogy az egyik szöveg valójában több rövidebb „szövegecske”, amely tovább szélesíti az idegen nyelvi készségből vett mintát, tükrözve annak összetettségét.

A megszerezhető pontok száma a beszédértés esetében sem eléggé informatív. Csak azt mutatja, hogy a pontokat általában felszorozzák, hogy illeszkedjenek a szóbeli interakciós részéhez. A pontszerzési esélyek száma, de még inkább az információ már sokkal többet elárul. Az ITK honlapról is letölthető mintafeladata meglepően kevés esélyt nyújt. Ezt némileg javítja, hogy a 7 item mindegyike többfokozatú skálának tekinthető, de az információs érték még így is csak 27! Ehhez képest az 5. táblázatban a vizsgák többségének információs értéke 40, míg a legtovább, 50, az Eurót jellemzi.

5. táblázat. Néhány beszédértés vizsga pontszerzési esélyei és információs értékei

Vizsga	Megszerezhető pontok	Pontszerzési esély	Információ
ITK	20	7	$2+4+6+2+1+3+2+(7*1)=27$
BME	60	20	$(10*2)+(10*2)=40$
Euro	25	25	$(6*2)+(9*2)+(10*2)=50$
TELC	75	20	$(5*2)+(10*2)+(5*2)=40$
ECL	25	20	$(20*2)=40$

Véleményünk szerint az a jobb beszédértés vizsga, amelyik vállalja sok hangzóanyag/feladat elkészítését, sőt hangzóanyagot még más vizsgarészekben is alkalmaz. Ilyenre is van (volt) példa. Konkrétan, a hangminőségen, a kérdések jóságán, megfelelőségén, stb. túl érdemes megvizsgálni, hogy a beszédértés vizsga egyes kiadásai tartalmazzanak-e legalább három feladatot (megfelelő hangzóanyaggal, illetve hangfelvétellel). Ezt a minimumot a mérési feladat fent leírt technikai-tartalmi összetettsége adja ki.

A beszédértés vizsga példája egyben a minta mélységével és átfogó jellegével kapcsolatos gondolatok összefoglalására is alkalmas. A beszédértés vizsgarész az akkreditációs rendszerben a szóbeli (rész)vizsga alkotóeleme, tehát az eredményeket a beszéd-készség vizsga eredményeivel együtt számítják. Ez egyben az is jelenti, hogy aki ilyen vizsgára jelentkezik, hátrányban lehet másokkal, akik a komplex nyelvvizsga írásbeli részére jelentkeznek. Miért? Míg a szóbeli legfeljebb két vizsgarészből áll, addig az írásbeli legalább kettő (írott szöveg értése, írásbeli kompozíciós készség), de inkább több vizsgarészből, amennyiben az adott vizsga nyelvi rendszerismeretet és közvetítést is tartalmaz. Logikus, hogy a vizsgázó több készség tekintetében, több feladaton bizonyíthat, különösen akkor, ha az írásbeli vizsga kettőnél több vizsgarészt tartalmaz. Ha a pontszerzési esélyek száma, és így az információs érték is magas, az írásbeli vizsga több esélyt nyújt, a vizsgázóra nézve „fairebb”, mert a szóbeli beszéd-készség részének megvannak a maga korlátai. A beszéd-készség mérésekor korlátozza a pontszerzési esélyek számát, hogy még páros értékelés esetén is a vizsgáztatóktól csak néhány, az íráskészségnél mindenesetre kevesebb értékelési szempont figyelembe vételét lehet elvárni, mert a vizsgázó megszólaltatása (interlokúció) is odafigyelést kíván meg. Ezért az egyik értékelő vagy egyáltalán nem osztályoz, vagy csak az összbenyomást osztályozza. Ha a szükségképpen korlátozott beszédértés méréséhez egy „rövid” beszédértés vizsga társul (például egy feladat/hangzóanyag, csupán néhány kérdéssel), szinte kizárt, hogy a vizsgázó idegen nyelvtudását jól megmérjék: nem kap elég pontszerzési esélyt és a mérés az egész szóbeli vizsgán csorbát szenved.

Vannak-e publikált folyamatszerű követelmények az adott nyelvvizsgálóhoz?

Ez a kérdés a nyelvtudás mérésének egy fentebb nem vizsgált (csak említett) aspektusára vonatkozik. A folyamatszerű követelmények a méréssel szemben támasztható olyan követelményeket ölelnek fel, amelyek a vizsgáztatási folyamat egyes állomásaiként írhatók le (Andor és mtsai, 1999). A folyamatszerű követelményeknek, ha a vizsgaanyag fejlesztését nézzük, a 1. ábrán felsorolt tevékenységeket kell felölelnie. Ha az értékelők feladatai felől közelítjük meg, a 2. ábrán felsorolt lépések állíthatók fel mint az eljárás kötelező részei. Tudva levő, hogy míg a vizsgaleírások meglehetősen statikus képet festenek egy nyelvvizsgáról, mintegy „pillanatfelvételt” készítve arról, hogy miből áll és milyennek kell lennie a vizsgának, addig a folyamatszerű követelmények ennél dinamikusabbak: arról szólnak, milyen folyamatnak kell megtörténnie, milyen eljárást kell lefolytatni, s nem csak a vizsgán, hanem az előtt és az után is ahhoz, hogy a vizsgaeredményeknek meglegyen a kellő érvényessége.

1. feladattervezetek elkészítése (feladatírás)
2. moderálás, a feladatok bírálata (bizottsági szakasz)
3. a feladattervezetek módosítása
4. a tervezett feladatok próbatesztelése, adatelemzés, a jóságmutatók ellenőrzése
5. további módosítások, nyomdai előkészítés, ellenőrzés
6. a vizsga lefolytatása (megíratása)
7. a dolgozatok értékelése (javítás)
8. válaszadatok bevitel, a feladatsor ellenőrzése (jóságmutatók alapján)
9. az eredmények kiszámítása
10. standardizáció, a határoló pontérték meghatározása.

1. ábra. A folyamatszerű követelmények a vizsgaanyag-fejlesztés szempontja szerint

1. vizsgáztató-tréning
2. benchmarking (tréning a vezető vizsgáztató iránymutató értékelése alapján)
3. a vizsga lefolytatása
4. az eredmények elemzése, jóságmutatók alapján
5. az eredmények kiszámítása
6. a határoló pontértékek kiszámítása

2. ábra. A folyamatszerű követelmények a vizsgáztató-képzés szempontjából

A folyamatszerű követelmények betartása igen fontos szerepet játszhat abban, hogy megítélhessük a vizsgák minőségét. Csupán és önmagában az, hogy a vizsgaközpont munkatársai minden egyes vizsgaanyag fejlesztése során, minden vizsgára (vizsgaidőszakra) való felkészülés során végigjárják a folyamat egyes „stációit”, a vizsga érvényessége, jobb minősége irányában hat még akkor is, ha e folyamatok természetes jellemzője, hogy a vizsgázók, tanáraik és általában a nagyközönség előtt e folyamatok nem vagy csak részben láthatók. Ezt a fajta érvényességet procedurális (vagy diskurzív) érvényességnek nevezzük (Somlai, 1997, 128. o.). A folyamat minden egyes eleme hozzájárul a vizsga minőségéhez, viszont a publikált tájékoztatókban, weboldalakon általában nem található meg. Nem lehet tudni, hogy miért. A NYAT akkreditációs kézikönyve is mint önálló, évente kiadott kiadvány megszűnt 2008 után. Ami maradt a magyarázó, tájékoztató részek elhagyásával, az a NYAT honlapján található „váz”, amely igen kevés a folyamatszerű követelmények jelentőségét jobban megérteni kívánó tanár és szakember számára.⁷ A folyamatszerű követelmények hiányára a szeminaristák kutatási eredményei

is figyelmeztettek. Diákjaim fentebb leírt kutatásai során kiderült, hogy az áttekintett nyelvvizsgák közül csak egy publikálta a vizsgafolyamat leírását!

A folyamatszerű követelmények fontosságának elismerése, mi több, publikálása természetesen nem garantálja, hogy azokat be is tartják. Annak megválaszolása, hogy betartják-e, vagy ha nem, akkor milyen alternatív, ekvivalens eljárást követnek, a NYAT felülvizsgálatok feladatai közé tartozik és legfeljebb a mérési szakember számára, de nem átlátható a diákjait vizsgára felkészítő tanár számára. A folyamatszerű követelmények teljesülését a vizsgaközpontok dokumentációs kötelezettsége alapján rendszeresen ellenőrzi a NYAT, de biztosak lehetünk-e benne, hogy a NYAT „mindent lát”? Az, hogy az ellenőrzés a NYAT feladata, még nem jelenti azt, hogy egy ilyen cikkben tárgyalni sem érdemes. Cikkünk következő, a megfigyelésről szóló fejezetéhez vezet már át az a záró gondolat, hogy a folyamatszerű követelményekről szóló hivatalos tájékoztatást célszerű a vizsgáztatásban részt vállaló tanár kollégák által elmondottakkal ütköztetni.

Megfigyeléssel és információszerzéssel megválaszolható kérdések

A vizsgafolyamatok azért is nehezen átláthatók, mert jelentős részben nem a vizsgázó szemei előtt zajlanak, hanem a vizsgaközpont berkein belül, a háttérben. Vegyük csak a fenti ábrák néhány elemét. A feladatfejlesztés során az új feladatok bírálata (moderálása) egyéni és bizottsági szakaszokra bontható. A vizsgaközpont munkatársai megkapják a feladattervezeteket, egyénileg áttanulmányozzák, megoldják a feladatokat, mert így tudják a vizsgázó helyzetét a legjobban átérezni és így tudnak felkészülni a bizottsági szakaszra, ahol egy értekezlet vagy belső levelezés keretében vitatják meg és hoznak döntést a feladattervezetetről. Világos, hogy a külső szemlélő számára, de még a mérési szakember számára is, ez a „bizottságosdi” nem elérhető. Láttuk, a folyamatszerű követelmények igen fontosak (lennének), de a vizsgaközpontok zárt világába a nyelvtanár alig tekinthet be. A vizsgaközpontban végzett munka minősége azonban, némely aspektusát tekintve, igenis felmérhető még a nyelvtanár számára is. A következőkben erről lesz szó.

Hogyan számítják a beszédértés (vagy íráskészség) jegyeket?

A folyamat olyan aspektusa, amely bizonyos mértékig átláthatók a kívülről állók számára is például a pontszámítás módja performancia típusú vizsgarészekben, amely jó és kevésbé jó módszerekről árulkodhat. A vizsgák egy részénél, helyesen, a két egyébként köztelevő értékelő által adott pontokat összeadják, vagy átlagolják, más vizsgáknál azonban az értékelés során a pontokat egyeztetik és megállapodnak a közösen adott pontokban. Ez utóbbi módszer nem teszi lehetővé az értékelők munkájának ellenőrzését. Utólag ugyanis nem lehet megválaszolni a megbízható értékelés és a két értékelő egyetértése mértékének igen fontos kérdését, mert annak feltétele lenne, hogy a pontértékek az értékelőkhöz mint egyénekhez és nem a párokhoz legyenek kapcsolhatók. Ha nincsenek a külön adott pontértékek rögzítve, az egyeztetett pontértékeket utólag visszabontani már nem lehetséges. Nézetünk szerint az a vizsga a jobb, ahol a párban dolgozó értékelők pontértékeit külön-külön rögzítik későbbi elemzés, minőségbiztosítás céljából.

Milyen hamar hirdetik meg az eredményeket?

Nem feltétlenül az a jobb vizsga, ahol az eredményeket igen hamar publikálják. Bár érthető a vizsgázók törekvése (és tanáraik elvárása), hogy minél hamarabb lezárhassák a vizsgázást, abszurd, de valós az a fentebbi példa, amelyben a vizsgáztatók a szóbeli vizsga végén kifüggesztik annak eredményeit a vizsgaterem ajtajára. Ha a vizsgáztatók így módon hozzák nyilvánosságra az eredményeket, teljesen kizárt, hogy a folyamatszerű követelményekből, amelyek az eredmények elemzését írják elő, bármit is megvalósítsanak. Evvel a módszerrel bármiféle utólagos ellenőrzés és korrekció lehetősége kizárt, ez pedig elég egyértelműen a vizsga gyengeségét jelzi. Lényegében véve arról van szó, hogy a vizsgaközpontnak nincs minőségbiztosítása; a rossz üzenet pedig nem más, mint hogy „a díjat köszönettel elfogadjuk, de ezért a vizsgázók sokat ne várjanak”. Az a vizsga tekinthető jobbnak, ahol az eredmények publikálása előtt van idő a mérési eszközrendszer (ennek része a vizsgáztató is) működésének ellenőrzésére.

Reális-e a vizsgán a megfelelési ráta (arány)?

A nyelvvizsgákon a mérés elsődlegesen a vizsgarészekben történik, azonban a vizsgarészek eredményeit valamilyen módszerrel összesíteni is kell a vizsga komplexitása és az áttekintés miatt is. Ezt az összesített teljesítményt (összpontszám) figyeli mindenki más: a vizsgázó, a tanára, a kolléga, a vizsgaközponttal versenyben lévő másik vizsgaközpont vezetője, stb. Ilyenformán értéktételek alakulnak ki „könnyű” vizsgákról, amennyiben a megfelelési rátát magasnak értékelik, vagy „nehéz” vizsgákról, és a tanárok is ezek szerint adnak diákjaiknak tanácsot. Nyilvánvalóan a „könnyű” vizsga rendkívül vonzó, hiszen – nem vitás – senki sem szeret vizsgázni és a jó eredményért külön erőfeszítést tenni.

A megfelelési ráta vizsgálata valóban jelenthet valami fontosat. Ha 65 százalék felett vagy 35 százalék alatt van, elmondhatjuk, a megfelelés valamely oknál fogva igen magas vagy igen alacsony az adott vizsgán. A közbenső tartományt lehet a „reális” jelzővel leírni. Hogy miért a középső tartományt lehet reálisként jellemezni, a magyarázat tulajdonképpen egyszerű: nem vagyunk tökéletesek, senki és semmi sem tökéletes. A vizsgázók a felkészülés során hibáznak; van, aki nem tud a megfelelő tempóban a vizsgára felkészülni és lemarad, a vizsga idején még nem áll megfelelő szinten; van, aki „beprobálgozik”, bár tudja, vagy tanára mondja, hogy még nincs itt a vizsga ideje. Van, aki a korai vizsgatapasztalatot tartja kellő motivációs forrásnak és van, hogy maga a vizsga sem sikerül a legjobban, vagy nem jól méri fel saját tudását a jelentkezés előtt, stb. Száz és száz okot lehet felsorolni tehát, amiért az adott vizsgahelyzet nem optimális és nem időszerű. Ezért is volt, hogy elképedéssel hallgattuk a szakembert, aki öntudatosan magyarázta el, hogy a vizsga, amelyben ő akkor részt vett, igen magas (90 százalékos) megfelelési arányt teljesít (Németh Zsuzsanna, személyes közlés). Az ehhez hasonló megfelelés legfeljebb a felkészítő kurzusokon elképzelhető (ott is nehezen), ahol a tanár abban a helyzetben van, hogy tanácsot adjon. A vizsgaközpontok azonban nincsenek abban a helyzetben, hogy tanácsot adjanak. Ez nem is feladatuk, sőt az államilag elismert nyelvvizsgáztatás rendszerének lényege, hogy szintező vizsgákat (‘proficiency testing’) működtetnek, ahol a vizsgáztatás és annak sikere-kudarca nem függhet a vizsgázó tanulmányi „előéletétől”. A logikailag (statisztikailag) legvalószínűbb, az emberi létben a tökéletesség hiányát figyelembe vevő, elméletileg legvalószínűbb megfelelési ráta az 50 százalék körüli érték. Ez azonban csak egy támpont, távolról sem szabad kizárólag erre figyelni, de az 50 százalék körüli érték arra utal, hogy a követelmények reálisak, és a jól és nem elég jól felkészült vizsgázók egyensúlyban vannak.

Reális-e a ponthatár (határoló pontérték)?

A fentiekből látható, mennyire bizonytalan, milyen hozzávetőleges csupán az ítélet, amely a megfelelési ráta vizsgálatából következhet, sokan mégis ebből próbálnak következtetéseket levonni. A bizonytalanság részben abból is fakad, hogy a sikeresség minimum ponthatára (vagy határoló pontérték, lásd: *Európa Tanács*, 2007) mintegy ráhatással van a megfelelési ráta alakulásába. Elméletileg a NYAT által megkívánt és 2007–2008 során lefolytatott illesztési eljárások eredménye az kellene legyen, hogy mind a határoló pontérték, mind a megfelelési ráta a *Közös Európai Referenciakeret* (PTMIK, 2002) adott szintjéhez (B1, B2, stb.) igazodik, illetve azt számszerűsíti. Az így lefolytatott illesztési munka eredménye igaz is lehetett 2008-ban. Akkor mintegy „pillanatfelvétel” készült, amely a továbbra is engedélyezett vizsgákat a KER-hez illesztettnek mutatta. Azóta azonban sok idő telt el és nem lehet tudni, az illesztettség továbbra is fennáll-e, fenntartása mennyire sikerült a nyelvvizsgáknak. A korábban illesztett vizsga nem biztos, hogy most is illesztettnek tekinthető.

A minimum ponthatár (határoló pontérték) akkor tekinthető reálisnak, ha a komplex vizsga részeredményeit összegző skálán (legyen az százalék vagy pontszám, például 0–150-ig) a skála középső tartományában helyezkedik el. Ennél pontosabb meghatározása már nehéz. A nyelvtudásmérés nemzetközi irodalmában nem találunk fogódzót. Nincs a közismert 50 vagy 60 százaléknak, vagy más hasonló értéknek a többlet kizáró módszertani alapja (*Alderson, Clapham és Wall*, 1995). Elméleti alapja azonban a skála középső tartományának viszont lehet, de akkor sem egy vagy más konkrét értéknek (például 57 százalék). A pontskálák, ugyanis, függetlenül attól, hogy a részvizsgákat mutatják-e vagy az összeredményt, a skála középső tartományában mérnek a legjobban. Középen sok az információ, sok a ponteredmény, tehát itt lehet a legkisebb mérési hibával dolgozni, míg a skála szélei felé kevesebb a ponteredmény. Egyre kevesebb az egyre jobb vagy egyre gyengébb tudású vizsgázó és az őket mérő egyre nehezebb vagy egyre könnyebb feladat. A vizsgán a pontmaximumot teljesítő vizsgázóról pedig elmondhatjuk, hogy mérhetetlen, mert ha teljesítette a maximumot, elérhette volna a maximum +1, +2, vagy akár +3 pontot is.⁸ Mindezek alapján kimondható, hogy a ponthatár akkor lehet reális, avagy az a jobb vizsga, ha a maximális összpontszám 50–60 százaléka körül mozog. Ez az elméleti alap segíthet a figyelmes szemlélőnek.

Akkreditált-e a nyelvvizsga?

Azokban kérdésekben, amelyekre a nyelvtanár nem kaphat választ, a Nyelvvizsgáztatási Akkreditációs Központ korábbi döntéseire és munkájára kell hagyatkozni. A nyelvvizsga-akkreditáció 1998-tól, illetve 2000-től indult. Munkáját sok siker (a szakmai követelmények felállítása, az akkreditációs munka elvégzése, a magyarországi vizsgarendszereknek a KER-hez való illesztése 2007–2008 folyamán, illetve a vizsgarendszerek időszakos ellenőrzése) és néhány fájdalmas kudarc kísérte (néhány rossz döntés, testületi döntés miniszteri felülbírlása legalább egyszer), mégis az akkreditáltság státusa jelent valami támpontot a vizsga minőségére.

A KER-hez való illesztés időszaka után a vizsgán sikeresen megfelelők aránya csökkent, amit – hozzáértők szerint – csak részben magyarázhat az iskolás és egyetemista korosztályok létszámának csökkenése. A magyarázat, amelyet e sorok írója kapott, a KER-hez való illesztést jelöli meg mint valószínű okot.

Összefoglalás

Az itt tárgyalt tíz kérdés lehetőséget teremt a nyelvtanároknak, hogy egyszerű eszközök segítségével megalapozottabb véleményt formáljanak nyelvvizsgákról. A szakértői szint összetett véleményalkotása redukálható egy viszonylag egyszerű kérdéssorrá, amely, a lehetőségesség és valószínűsíthetőség szintjén, választ adhat arra kérdésre, melyek a jobb nyelvvizsgák. Feltételezhető, hogy a válaszok szóródni fognak és nem biztos, hogy lesz olyan jó vizsga, amelyre mind a tíz kérdésre pozitív választ adhatunk. A maximum azonban nem is fontos. A cél szempontjából elég annyi is, hogy kiderül, melyik vizsgára lehet a legtöbb több pozitív választ adni a kérdések kapcsán.

Bár nem vizsgáltunk meg minden magyarországi vizsgát, az itt kapott eredmények nem tekinthetők elég jónak. Az elkülönült elemekből álló itemsorok általában nem nyújtanak elég pontszerzési esélyt és információt, a produktív integratív itemekből viszont, amelyekkel az elkülönült elemekből építkező vizsgarészek felváltandók-felválthatók lennének, nincs elég a vizsgákban, így nem is alkothatnak megfelelő mintát.

A cikk lektorálása során felmerült, célravezetőbb és etikusabb lenne-e, ha a vizsgák anonimizálva kerülnének elemzésre a cikkben. Véleményünk szerint ez azonban egyrészt épp a cikk célját, a nyelvtanári kar tájékoztatását (és a cikk érdekességét) ütné, másrészt pedig – ismét hangsúlyozzuk – a cikk semmilyen titkos, publikus helyen fel nem található, viszont az összehasonlításra alkalmas információt nem szellőztet meg.

Szinte hallom a kifogást, hogy az itt leírt számítások csak a vizsgák megbízhatóságával foglalkoznak, pedig a vizsga érvényessége (validitása) ennél előbbre való. Tény, hogy nyelvtanári körökben gyakran (?) hallani a nézetet, amely úgy helyezi előtérben az érvényességet, hogy elválasztja a megbízhatóságtól, és azt hangsúlyozza, hogy egy vizsga lehet megfelelő érvényességű anélkül, hogy megbízhatósága megalapozott lenne (Christopher Ryan, személyes közlés). Ez a különbségtétel és rangsorolás azonban figyelmen kívül hagyja, hogy az idegen nyelvtudás mérésének irodalmában, sőt a pszichológiai és pedagógiai mérés szakirodalmában is, egyetértés van a tekintetben, hogy a mérés megfelelő megbízhatósága nélkül nem képzelhető el annak érvényessége sem, néhány kulcsfontosságú mű tanúsága szerint is (*Messick*, 1989; *Kane*, 2006; *Bachman*, 1990; *Bárdos*, 2002). A megbízhatóság az érvényesség sine qua non típusú feltétele, és mint ilyen, nem engedi meg a voluntarista szétválasztást. Ha nem megfelelő a megbízhatóság, az érvényesség sem lehet megfelelő. Lehet bármilyen jó az egyetlen íráskészség vagy szóbeli feladat, ha nincs elég értékelő, vagy munkájukat nem végzik jól, az érvényesség biztosan csorbát szenved. Hozzátesszük, az érvényesség nem is magának a feladatoknak, itemeknek, vagyis a vizsgaanyagoknak a jellemzője, hanem az vizsgaeredmények adott értelmezésének (*Bachman*, 1990). Ha jól megnézzük e cikk néhány állítását, továbbá a néhány kérdést, amely látszólag a megbízhatóságot helyezi előtérbe, nem a megbízhatósággal foglalkozik a fogalom valódi értelmében. A megbízhatóság ugyanis a vizsgázó válaszadataira adott pontértékek – a számok – értékmérője (*Bachman*, 1990), evvel szemben e cikkben arról volt szó, hogy a tanár papír és ceruza vagy kalkulátor segítségével mit tudhat meg a vizsgákról. Lényegében tehát e cikk egy vélelmezett megbízhatóságot vizsgál, hiszen a tanár jellemzően nem fér hozzá válaszadatokhoz. A cikk azt (is) vizsgálja, hogy feltételezve, hogy a feladatok és az értékelők maguk kiválóak, a vizsgaközpont milyen szintű megbízhatóságot érhet el a legjobb esetben.

A cikk megírásának időpontjában kurrens és széles elfogadottsággal bíró érvényesség nézetek (különösen *Messick*, 1989 és *Kane*, 2006) alapján minden itt feltett kérdés végső soron a vizsga érvényességét veszi célba. A legfontosabbat, azt, amelyik arról szól, hogy amit a vizsga a maga keretei között megmér, nyelvtudásnak tekinthetjük-e, tárgyalásunk során az első helyre tettük. A jó vizsga e kritériumának pontosabb meghatározása azonban nehézségbe ütközik, mert a nyelvtudás mibenléte, tagolása és alkotóelemei nem

tekinthetők állandónak, történetileg változó fogalomról van ugyanis szó (Bárdos, 2003). Az érvényesség tárgyalásához tartozik a folyamatszerű követelmények betartásáról szóló kérdés is csakúgy, mint a cikkben feltett többi jelentősebb és kevésbé jelentősnek tekinthető kérdés. Az érvényesség összetett fogalom. Különféle, a célzott (elvárt, remélt) értelmezést megerősítő bizonyítékot, adatot és információt integrál, de az érvényesség megalapozásához továbbá a célzott értelmezéstől eltérő vagy avval ellentétes értelmezéseket ('rival hypotheses') is meg kell tudni dönteni (Messick, 1989, 35–36. o).

Irodalom

- Alderson, J. C., Clapham, C. és Wall, D. (1995): *Language Test Construction and Evaluation*. Cambridge University Press, Cambridge.
- Andor József, Berényi Sarolta, Borgulya Ágnes, Dávid Gergely, Fekete Hajnal, Heltai Pál, Lengyel Zsolt, Németh Zsuzsanna és Völgyes Gyöngyvér (1999): *Az államilag elismert nyelvvizsgák akkreditációjának kézikönyve*. Professzorok Háza, Budapest.
- Bachman, L. F. (1990): *Fundamental Considerations in Language Testing*. Oxford University Press, Oxford.
- Bárdos Jenő (2002): *Az idegen nyelvi mérés és értékelés elmélete és gyakorlata*. Nemzeti Tankönyvkiadó, Budapest.
- Bárdos Jenő (2003): A nyelvtudás megítélésének korlátai. *Iskolakultúra*, 13. 8. sz. 28–39.
- Bárdos Jenő (2005): *Élő nyelvtanítás-történet*. Nemzeti Tankönyvkiadó, Budapest.
- Buck, G. (2001): *Assessing Listening*. Cambridge University Press, Cambridge.
- Cambridge English Language Assessment (2013): *Principles of Good Practice. Quality Management and Validation in Language Assessment*. 2013. 11. 17-i megtekintés, <http://www.cambridgeenglish.org/images/22695-principles-of-good-practice.pdf>
- Carroll, J. B. (1961): Fundamental considerations in testing for English language proficiency of foreign students. In: *Testing the English Proficiency of Foreign Students*. Center for Applied Linguistics, Washington, DC.
- Carroll, J. B. (1980): *Testing Communicative Performance: An Interim Study*. Pergamon Press, Oxford.
- Carroll, J. B. (1981): Specifications for an English language testing service. In: Alderson, J. C. és Hughes, A. (szerk.): *Issues in Language Testing*. ELT Documents 111. The British Council, London.
- Chomsky, N. (1965): *Aspects of the Theory of Syntax*. M.I.T. Press, Cambridge, Mass.
- Csepes Ildikó (2012): Államilag elismert nyelvvizsgáztatás: Kik és hogyan látják el a szakmai felügyeletet? *Modern Nyelvoktatás*, 18. 4. sz. 57–60.
- Davies, A. (1978): Language Testing. Survey Article Parts I and II. *Language Teaching and Linguistics Abstracts*, 2. 3–4 sz. 145–159. és 215–231.
- Dávid Gergely (2001): *A Nyelvvizsgát Akkreditáló Testület bírálati eszközeinek próbája*. Előadás: I. Neveléstudományi Konferencia. Budapest, 2001. 10. 25–27.
- European Association for Language Testing and Assessment. (é. n.): *EALTA guidelines for Good Practice in language Testing and Assessment*. <http://www.ealta.eu.org/documents/archive/guidelines/English.pdf>
- Európa Tanács (2007): *Szintillesztési módszertani segédlet a Közös Európai Referenciakerethez című Kézikönyvhöz*. Nyelvvizsgát Akkreditáló Testület, Nyelvvizsgáztatási Akkreditációs Központ, Budapest.
- Fazekas M. (2004, szerk.): *A (nyelv)tudás hatalom! Avagy minden, amit az államilag elismert nyelvvizsgákról és az emelt szintű idegennyelvi érettségiről tudni lehet*. Nyelvvizsgáztatási Akkreditációs Központ, Budapest.
- Fekete Hajnal és Óvári Valéria (szerk.) (2004): *Nyelvtudásmérési terminológiai szótár. Magyar nyelvű kiadás*. Idegennyelvi Továbbképző Központ, Budapest.
- Hymes, D. H. (1972): On communicative competence. In: Pride, J. B. és Holmes, J. (szerk.): *Sociolinguistics: selected readings*. Penguin Education, Harmondsworth, Middlesex. 269–293.
- Horváth György (1993): *Bevezetés a tesztelméletbe*. Keraban Kiadó, Budapest.
- Hughes, A. (2003): *Testing for Language Teachers. Second Edition*. Cambridge University Press, Cambridge.
- Kane, M. T. (2006): Validation. In: Brennan, R. (szerk.): *Educational measurement*. 4. kiadás. American Council on Education and Praeger, Westport, CT. 17–64.
- Király Zsolt (2012): Az angol érettségi „hallott szöveg értéke” feladatának gyártási technológiája. *Modern nyelvoktatás*, 18. 3. sz. 55–66.

- Lado, R. (1961): *Language Testing*. Longman, London.
- Linacre, J. M. (2013): *Facets: Rasch Measurement Computer Program*. Version 3.71 [Computer software] Mesa Press, Chicago.
- McNamara, T. (1996): *Measuring Second Language Performance*. Longman, Harlow.
- McNamara, T. (2000): *Language Testing*. Oxford University Press, Oxford.
- 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.
- Morrow, K. (1977): *Techniques of evaluation for a notional syllabus*. Royal Society of Arts, Reading.
- Morrow, K. (1979): Communicative Language Testing: Revolution or Evolution? In: Brumfit, C. J. és Johnson, K. (szerk.): *The Communicative Approach to Language Teaching*. Oxford University Press, Oxford. 143–159.
- Morrow, K. (1983): Some comments on Issues. In: Hughes, A. és Porter, D. (szerk.): *Current Developments in Language Testing*. Academic Press, London.
- Morrow, K. (1986): The Evaluation of Tests of Communicative Performance. In: Portal, M. (szerk.): *Innovations in Language Testing*. NFER-Nelson, Windsor, Berks. 1–13.
- Nyelvvizsgát Akkreditáló Testület (2003): *Részletes jelentés a Budapesti Műszaki és Gazdaságtudományi Egyetem nyelvviszsgarendszerének felülvizsgálatáról*. Nyelvvizsgáztatási Akkreditációs Központ, Budapest.
- PTMIK (2002): *Közös Európai Referenciakeret. Nyelvtanulás, nyelvtanítás, értékelés*. Pedagógus-továbbképzési Módszertani és Információs Központ Kht.
- Somlai Péter (1997): *Szocializáció: A kulturális átöröklés és a társadalmi beilleszkedés folyamata*. Corvina, Budapest.
- Szokolszky Ágnes (2004): Kutatómunka a pszichológiában. Osiris Kiadó, Budapest.
- Weir, C. J. (1990): *Communicative Language Testing*. Prentice Hall, Hemel Hempstead, Herts.
- Weir, C. J. (1993): *Understanding and Developing Language Tests*. Prentice Hall, Hemel Hempstead, Herts.

Jegyzetek

¹ <http://www.ealta.eu.org/documents/archive/guidelines/Hungarian.pdf>

² <http://www.oktatas.hu/nyelvvizsga/nyelvvizsgakozpontoknak/fogalomtar/>

³ Bár az alapkészségek leírását Bárdos (2005) a 19. századba vezeti vissza (*Marcel*, 1869), e cikk szempontjából fontosabb, hogy a készségek elkülönített mérését az idegen nyelvek kommunikatív mérése is hagyományként vitte tovább (*Carroll*, 1961).

⁴ Nincs szó arról sem, hogy a készségek használata nem a nyelvtudás megnyilvánulása lenne.

⁵ Az, hogy mennyi információ tekinthető elegendő információnak, nyilvánvalóan nem könnyen megválaszolható kérdés. A mérési hiba szerepének tár-

gyalásakor teszünk kísérletet arra, hogy e tekintetben használható támpontot nyújtsunk.

⁶ Ugyanezt a gondolkodást követi számos szoftver, köztük a Facets (*Linacre*, 2013) is, amikor alapbeállítás szerint számításait akkor tekinti befejezettnek, ha a nyerspont-átlagok és a mért értékek átlagai között már csak fél pont különbség van.

⁷ http://www.nyak.hu/nyat/doc/ak2012/Folyamatszeru_kovetelmenyek.htm

⁸ Azon skálák esetében, melyek végesek (például a nyerspont skála), a szélső pontértékeknél lecsökken a mérési hiba (Feliánka Kaftandjieva, személyes közlés), míg a végtelen skálák esetében (például a valószínűségi mérésnél alkalmazott logit skálánál) a szélső értékeken hirtelen megnő a mérési hiba.