

A KOMMUNIKATÍV TESZTELÉS ELMÉLETI ALAPJAI

Vígh Tibor

Szegedi Tudományegyetem, Neveléstudományi Doktori Iskola

A kommunikatív tesztelés célja a nyelvhasználó kommunikatív kompetenciájának mérése és értékelése. E tesztelési forma fejlődési szakaszában, a kilencvenes években vált az idegen nyelvi mérés és értékelés elmélete multidiszciplináris tudományággá, amely szoros kapcsolatban áll az alkalmazott nyelvészettel, valamint a statisztikai módszerekkel támogatott mérésmethodikával és a pedagógiai értékeléssel (*Alderson, Clapham és Wall, 1995; Bachman és Palmer, 1996; Bárdos, 2001, 2004*). Így a nyelvpedagógiába tartozó kommunikatív tesztelés, az idegen nyelvi mérés és értékelés egy dimenziójaként, alkalmazza és gazdagítja a tesztelmélet eljárásait. A tanulmány szakirodalmi áttekintésében egyrészt arra keresünk választ, hogy a kommunikatív tesztelés miként integrálódik a tesztelmélet egészébe, milyen fogalmakat, tesztelési módszereket alkalmaz; másrészt pedig arra, hogy a kommunikatív kompetencia méréséből és értékeléséből adódóan milyen problémákkal küzd.

A tanulmányban előforduló fontosabb szakkifejezések, alapfogalmak meghatározása után áttekintjük a kommunikatív tesztelés szempontjából releváns nyelvtudásmodellek fejlődését, illetve, hogy a modellek által feltárt nyelvtudásnak mely szintjeit lehet megkülönböztetni, és bemutatjuk a kommunikatív nyelvtudás egységes referenciaszintjeit. A kommunikatív tesztelés módszereinél meghatározzuk, hogy a kommunikatív teszteknek egyrészt milyen minőségi kritériumoknak kell megfelelniük, másrészt, hogy a tesztek mely típusai alkalmasak a különböző mérési célok elérésére, és milyen módon tudunk a teszteredményekből bizonyos döntéseket hozni. Ezen túl célunk bemutatni, hogy a kommunikatív nyelvtudás fogalmát leíró nyelvtudásmodellek, a tesztek jószágmutatói, eltérő funkciói és típusai milyen kapcsolatban állnak egymással. A tanulmányt lezáró részben a kommunikatív tesztelés alkalmazásaival foglalkozunk, bemutatjuk a nyelv- vizsgaszintek egységes meghatározásának szükségességét és folyamatát, és néhány kutatási és fejlesztési irányt.

A nyelvi mérés és értékelés fogalma

Az idegen nyelvi mérés és értékelés két különböző fogalmat jelöl, amelynek használata a szakirodalomban nem teljesen egységes (Figueras, 2005). A nyelvi mérés (*assessment/measurement/testing*) során az adatgyűjtést követően számokat rendelünk valamilyen tulajdonsághoz előre meghatározott szabályok szerint. A mérés tehát valamilyen számszerűsítést, „mennyiségi” folyamatot jelöl. Az értékelés (*evaluation*) szerepe, hogy a mérés számszerűsített információja alapján valamilyen döntést hozzunk, azaz itt valamilyen „minőségi” folyamatról van szó (Bachman, 1990; Bárdos, 2002a; Grotjahn, 2006; Közös Európai Referenciakeret, 2002).

Az idegen nyelvi mérés eszközei a különböző típusú és funkciójú nyelvi tesztek. A teszt fogalma *tágabb értelemben véve* olyan eljárást jelöl, amely az egyéneket kontrollált feltételek mellett bizonyos cselekvési és viselkedési módokra ösztönzi, és amelyekből következtetéseket lehet levonni az egyén többnyire közvetlenül nem megfigyelhető tulajdonságaira, képességeire és készségeire. A nyelvi tesztek pedig az idegen nyelvi készségek megállapításának eszközeit jelentik. Ebben az értelemben a teszt és a vizsga kifejezés között nincs különbség, egymás szinonimái. *Szűkebb értelemben véve* a teszt kevésbé, míg a vizsga erőteljesebben formalizált és standardizált eljárást jelent. Ebben az esetben az informális méréseket tesztnek, a formális méréseket pedig vizsgának nevezzük (Bárdos, 2002a; Krekeler, 2005; Lienert és Raatz, 1994; Perlmann-Balme, 2001). A nyelvi tesztek által mérendő tulajdonságokat, képességeket és készségeket konstruktnak (*construct*) nevezzük. Ez a kifejezés a pszichológiából származik, amelyet elsősorban a képességek, esetleg készségek, sajátos tudások jelölésére használnak, és elméleti célokra bevezetett fogalmat jelöl (Bárdos, 2002a). A konstruktnak általában látnak, közvetlenül nem megfigyelhető. A nyelvi tesztek alkalmazásának célja, hogy a konstruktnak mozgósítsa a tesztfeladatok és az egyes tesztelemelek, az itemek segítségével (Davies és mtsai, 1999; Grotjahn, 2000a, 2000b, 2006). Így egyrészt lehetővé válik, hogy a vizsgázók nyelvtudását mérjük, másrészt pedig, hogy a teszten elért eredmények alapján különböző döntéseket hozzunk, amelyek egyes személyekre, személyek csoportjára vagy egy teljes képzési programra is irányulhatnak. A döntések azonban csak akkor tekinthetők érvényesnek és megbízhatónak, ha meg tudjuk határozni, hogy milyen mértékben egyezik meg a teszten nyújtott teljesítmény más valós nyelvi szituációban, élethelyzetben nyújtott teljesítménnyel (Bachman, 1990; Grotjahn, 2002; Takala, 1998). Ehhez szükséges annak meghatározása, hogy mit értünk nyelvtudáson, milyen a vizsgálandó készségek konstruktnak (Bárdos, 2003, 2004). Olyan modellt kell tehát alkotni, amellyel lehetőség van a nyelvtudás feltárására, azért hogy pontosan meg tudjuk határozni, hogy mit kívánunk mérni és értékelni.

A kommunikatív nyelvtudás szerkezetének modelljei

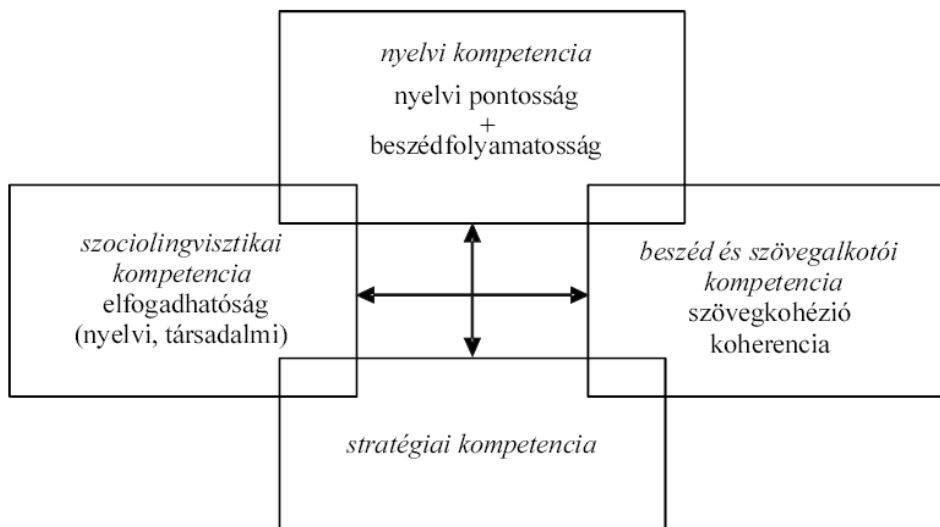
A nyelvtudást mérő tesztfeladatokból minden esetben kikövetkeztethető, hogy mi a tesztkészítők felfogása a nyelvtudás fogalmáról és szerkezetéről. A nyelvtudás értelme-

zése azonban a különböző nyelvtanítási módszerekben eltérő (Alderson, 2006; Bárdos, 2005, 2006; Einhorn, 1998), a nyelvi mérés és értékelés módszertana pedig jelentős fejlődésen ment keresztül (Bárdos, 2001, 2002a, 2004; Bolton, 1996; Neuner és Hunfeld, 1993), a szubjektív tesztelési technikáktól kezdve a strukturalista-pszichometrikus szemléleten (Lado, 1961) és az egységes kompetencia hipotézisén (Oller, 1979) keresztül a kommunikatív tesztelésig, amely autentikus szövegekből és feladatokból kiindulva, objektív és szubjektív tesztelési technikákkal méri a nyelvtudást. Ezt a fejlődésmenetet részletesen mutatja be Bárdos (2001, 2002a, 2004) és Krekeler (2005), az egyes fázisokról Dávid (1992), Katona (1995), Katona és Dörnyei (1991) ad átfogó képet.

A továbbiakban meghatározzuk a kommunikatív tesztelés nyelvtudásról alkotott fogalmát, amellyel kapcsolatban szoktak beszélni kommunikatív kompetenciáról (*communicative competence*), és kommunikatív nyelvi képességről (*communicative language ability*). Ezen túl azzal is foglalkozunk, hogy milyen szerepe van a nyelvi készségekről és képességekről alkotott elméleti modelleknek a kommunikatív tesztek szerkesztésekor és értékelésekor.

A kommunikatív nyelvtudást Morrow (1981) a nyelvhasználat öt tulajdonságának megfogalmazásával határozta meg. Eszerint minden nyelvhasználat (1) interakción alapul, egy bizonyos (2) kontextusban valósul meg, valamilyen (3) cél és (4) hatás elérése érdekében (5) autentikus nyelvi szituációban történik (lásd részletesebben: Albers és Bolton, 1995; Alderson, Clapham és Wall, 1995; Bárdos, 1984, 2004; Bolton, 1996; Miyata-Boddy és Langham, 2000). E tulajdonságokból levezethető a kommunikatív nyelvoktatás célja, azaz annak a képességnek a kialakítása, hogy a nyelvhasználó interakciókat tudjon létrehozni, a kommunikációs szituációnak és saját közlési szándékainak megfelelően, és így sikeresen tudjon nyelvileg cselekedni. Ezt a komplex képességet nevezzük kommunikatív kompetenciának. E kompetencia összetevőinek meghatározására és a kommunikatív nyelvi képességek leírására több modell is született, amelyek közös jellemzője, hogy a nyelvi kompetencia többdimenziós jellegéből indulnak ki, és amelyekben az egyes komponensek többé-kevésbé szoros összefüggésben, egymással kölcsönhatásban állnak. A bemutatandó négy modell célja a tesztben elért és meghatározott nyelvi teljesítmények és a nyelvi képességek közötti összefüggések meghatározása, azaz, hogy a teszten elért eredmények, hogyan interpretálhatók.

Canale és Swain (1980) modellje a kommunikatív kompetencia szerkezetét mutatja be, amelyet többdimenziósnak, és dinamikusnak írnak le. A Canale (1983) által kibővített modell négy összetevőt tartalmaz (1. ábra).



1. ábra

A kommunikatív kompetencia modellje
(Canale, 1983 alapján: Bárdos, 2004. 150. o.)

A modellben szerelő nyelvi kompetencia egyesíti a *Chomsky* (1965) és *Hymes* (1972) által meghatározott nyelvi kompetencia és performancia fogalmát. A szövegalkotói kompetencia a szövegkohézióknak és a koherenciának, míg a szociolingvisztikai kompetencia az elfogadhatóságnak és a társadalmi elvárásnak megfelelő kifejezésmódot jelenti. A stratégiai kompetencia olyan verbális és nem verbális stratégiák használatát jelenti, amelyek elsősorban a nyelvi hiányosságok kompenzálására szolgálnak. A modell, bár nem határozta meg pontosan a négy összetevő közötti kapcsolatot, mégis a szakirodalomban a leggyakrabban hivatkozott modell (lásd pl.: *Bachman*, 1990; *Bárdos*, 2002a, 2002b, 2004, 2006; *Grotjahn*, 2000a; *Krekeler*, 2005; *Miyata-Boddy* és *Langham*, 2000; *Sigott*, 1996; *Tschirner*, 2001; *Young*, 2000).

Bachman (1990) modellje a kommunikatív nyelvi képességekről az előző modell közvetlen továbbfejlesztésének tekinthető, és sajátossága a nyelvi tesztekkel való kapcsolata. A modell három fő komponenst tartalmaz: a (I) nyelvi és (II) stratégiai kompetenciát, valamint a (III) pszichofiziológiai mechanizmusokat, amelyek a produktív és a receptív készségeket jelentik. A korábbi modellel ellentétben a (I) nyelvi kompetencia szerkezetét két részre bontva hierarchikus rendszerben írja le:

- (1) szervezési/ elrendezési kompetencia (*organizational competence*), amely magába foglalja (a) a nyelvtani és (b) a szövegalkotói kompetenciát, valamint
- (2) pragmatikai kompetencia, amely (a) az illokúciós (beszédfunkciók kifejezésének képessége) és (b) a szociolingvisztikai kompetenciát tartalmazza.

Ezzel a rendszerezéssel, a korábbi modellel ellentétben, differenciáltabb képet kapunk a nyelvi kompetencia összetevőiről és a köztük levő kapcsolatról. A nyelvi kompetencia

ilyen felosztásához kapcsolódik azonban a modell kritikája is. *Bárdos* (2002a. 98. o.) szerint a modell egyrészt „zavaró kiemeléseket tartalmaz”, másrészt pedig az „ábrázolásból nem derül ki, hogy a nyelvi kompetencia alcsoportjai nemcsak egymásmellettekben, hanem egymás részeiként is léteznek.” Ezen kívül „*Bachman* kategorizálása [...] megismétli a *Canale* és *Swain* által létrehozott konstrukciót, illetve visszanyúl korábbi sémákhoz és visszahozza a készségeket”.

Bachman és *Palmer* (1996) az előző modellre épülve, a nyelvi képességek komponenseit az alábbiak szerint határozzák meg (1. táblázat).

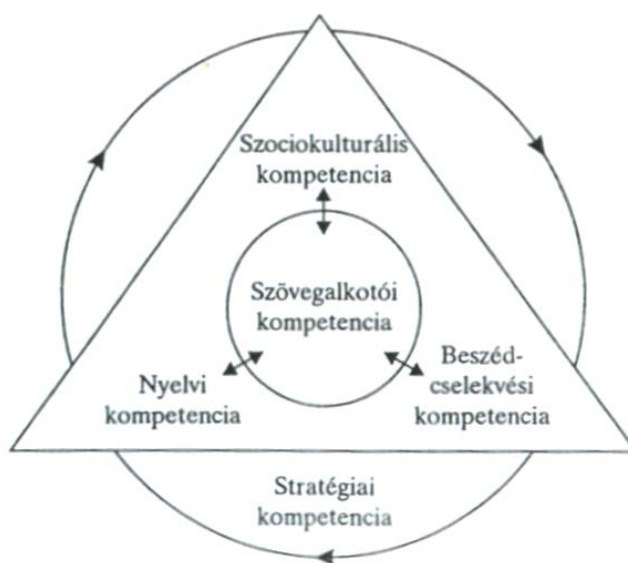
1. táblázat. *Bachman* és *Palmer* kommunikatív nyelvi képességek modellje (1996. 66–75. o.)

Nyelvi tudás				Stratégiai kompetencia
Nyelvszervezői tudás		Pragmatikai tudás		
Nyelvtani tudás	Szövegalkotói tudás	Funkcionális tudás	Szocilingvisztikai tudás	
fonológia	kohézió	kifejezési funkció	dialektusok / variánsok	<i>célok meghatározása</i>
morfológia	retorika	manipulatív funkció a) instrumentális; b) reguláris; c) interperszonális	regiszter	<i>célok tervezése</i>
szintaxis		heurisztikus funkció	idiómák	<i>célok értékelése</i>
lexika		imaginatív funkció	kulturális referenciák és nyelvi alakzatok	

Az egyik összetevő a nyelvi tudás, a másik pedig a stratégiai kompetencia. A nyelvi tudás megegyezik *Bachman* (1990) modelljének nyelvi kompetencia fogalmával, mivel az előzőekhez hasonlóan a szervezési és a pragmatikai tudást tartalmazza. A két modell közti különbség, hogy *Bachman* és *Palmer* (1996) nem említik a korábbi modell harmadik komponensét, a pszichofiziológiai mechanizmusokat. Ezen túl elsősorban a funkcionális tudás összetevőit részletezik, amely az előző modellben az illokúciós kompetenciával egyezik meg. A beszédfunkciók kifejezésének képességéhez tartozik az információk közlésének (kifejezési funkció), a nyelvi tartalmak kitalálásának (imaginatív funkció), az információk megszerzésének (heurisztikus funkció) és az emberek befolyásolásának (manipulatív funkció) a képessége. A stratégiai kompetencia felfogása a két modellben ugyanaz. Ennek az összetevőnek azonban jóval nagyobb szerepe van, mint *Canale* és *Swain* (1980) modelljében. Míg az első modellben inkább kompenzáló funkciója volt, addig a másik két modellben olyan metakognitív folyamatokról van szó, amelyek a beszédfunkciók tervezését, megvalósítását és kiértékelését is tartalmazzák. A modell a két komponens közti kölcsönhatást úgy emeli ki, hogy a nyelvi teljesítmény és a nyelvtudás közti kapcsolatot dinamikusnak tekinti. *Bachman* és *Palmer* (1996) modelljét különbö-

zöképpen ítéli meg a szakirodalom. *Skehan* (1998) elsősorban azt hiányolja, hogy a modell nem írja le a nyelvi feldolgozó mechanizmusokat, hanem olyan konceptuális sémákat ad, amelyek a tesztek szerkesztésére és a teszteredmények interpretálására szolgálnak. A modell a tesztfejlesztés és elemzés gyakorlatában jól alkalmazható (*Grotjahn*, 2000a; *Krekeler*, 2005; *Sigott*, 1996), amit az is tükröz, hogy a nyelvvizsgák leírásában ez az egyik leggyakrabban hivatkozott modell (*Dávid*, 2002).

Az idegennyelv-tanítás és vizsgáztatás folyamatának megtervezésére és lebonyolításának megkönnyítésére alkotta meg *Celce-Murcia*, *Dörnyei* és *Thurrell* (1995, idézi: *Bárdos*, 2002a) a nyelvtudás funkcionális-dinamikus modelljét (2. ábra).



2. ábra

A nyelvtudás funkcionális-dinamikus modellje

(*Celce-Murcia, Dörnyei és Thurrell, 1995, idézi: Bárdos, 2002a. 100. o.*)

A modell öt összetevőt tartalmaz: (1) a szövegalkotói kompetencia a szövegteremtés központi szerepét hangsúlyozza, amely közvetlen kapcsolatban van (2) a beszéd-cselekvési („akcionális”) kompetenciával, (amely a kommunikációs szándék megértésének és kifejezésének képességét jelenti), (3) a nyelvi és (4) a szociolingvisztikai kompetenciával. Ezeket veszi körül (5) a stratégiai kompetencia, amely a kommunikációs stratégiák helyes használatát jelenti, és a többi kompetencia kíségetésére szolgál. A modell előnye, hogy a szerzők felsorolják az összes alkompetencia összetevőit, és nagy hangsúlyt fektetnek a szociolingvisztikai és a beszéd-cselekvési kompetencia részletes bemutatására is.

A kommunikatív nyelvtudás szintjei

A kommunikatív nyelvtudás fontosabb modelljeinek áttekintése után azzal foglalkozunk, hogy a modellek által feltárt nyelvtudásnak mely szintjeit lehet megkülönböztetni. Ehhez a *Közös Európai Referenciakeret*¹ szintleírásai kerülnek bemutatásra. A dokumentum először angolul (*A Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, 2001), majd 2002-ben magyar nyelven is megjelent, és eddig 23 nyelvre fordították le. A Referenciakeret jelentőségét mutatja, hogy szintleírása „közös alapot teremt Európa-szerte a nyelvi tantervek, tantervkészítési irányelvek, vizsgák, tankönyvek stb. kidolgozásához [...]”. A Referenciakeret meghatározza a nyelvtudás szintjeit, és ezáltal lehetővé teszi, hogy a tanulók haladása a nyelvtanulás minden szakaszában és az egész életük során mérhető legyen” (KER, 2002. 1. o.).

A Referenciakeret a kommunikatív nyelvtudás modelljeire épülve és azokat valamilyen szinten ötvözve, definiálja a kommunikatív nyelvi kompetenciákat, amelyeknek alkotóelemei:

- (1) a nyelvi kompetenciák (amelyek a lexikális, grammatikai, szemantikai, fonológiai, helyesírási kompetenciából állnak),
- (2) a szociolingvisztikai kompetenciák (beleértve a társadalmi viszonyok nyelvi jelölését, az udvariassági szokások, népi bölcsesség és stílusbeli különbségek ismeretét, valamint a dialektusokat és az akcentust), és
- (3) a pragmatikai kompetenciák (ideértve a beszédalkotási, funkcionális kompetenciákat) (lásd részletesebben: KER, 2002. 5.2. fejezet).

„A stratégiai komponens szerepet játszik a kommunikatív folyamatok menet közben történő megfigyelésében, valamint a folyamat ennek megfelelő irányításában is” (KER, 2002. 110. o.). A Referenciakeret (1) a produktív, (2) a receptív, (3) az interaktív és (4) a közvetítő stratégiákat (KER, 2002. 4.4. fejezet) határozza meg.

A Közös Európai Referenciakeret szintleírásait függőleges és vízszintes dimenziókra lehet bontani. A függőleges dimenzió egy olyan fokozatosan emelkedő referenciaszintrendszer vázolat, amelynek segítségével a nyelvtanulók nyelvi kompetenciáinak szintje leírható. Ezek a szintek a következők: (1) Alapszintű nyelvhasználó, amelyhez az A1: Minimumszint (*Breakthrough*) és az A2: Alapszint (*Waystage*) tartozik. (2) Az Önálló nyelvhasználó szinthez a B1: Küszöbszint (*Threshold*) és a B2: Középszint (*Vantage*) rendelhető, míg (3) a Mesterfokú nyelvhasználó szintet a C1: Haladó szint (*Effective Operational Proficiency*) és a C2: Mesterszint (*Mastery*) alkotja (lásd részletesebben: KER, 2002. 3. fejezet). A Referenciakeret szintrendszerében egyrészt „az Alapszintet (A2) a skálán hozzávetőleg ugyanakkora távolság választja el a Küszöbszinttől (B1), mint amekkora az utóbbit a Középszinttől (B2)” (KER, 2002. 23. o.). Másrészt a tanítási gyakorlat azt mutatja, „hogy sok diáknak a Küszöbszint elérése az Alapszintről kétszer annyi idejébe került, mint amennyit az Alapszint elérése igényelt. Feltételezhető, hogy több mint kétszer annyi ideig fog tartani, amíg a Küszöbszintről [...] eléri a Középszin-

¹ A továbbiakban a Referenciakeret fogalmat használjuk, ha a dokumentumról általában van szó. A rövidített hivatkozás: KER, 2002.

tet [...]. Ennek oka, hogy a nyelvi tevékenységek, a készségek és a nyelvi eszközök köre fölfelé haladva szükségképpen bővül” (KER, 2002. 23. o.), mivel a szintrendszerben minden szint magába foglalja a skálán alatta levőt is. Ebből is következik, hogy a függőleges dimenzió szintjeit a Referenciakeret további szintekre osztja egy svájci vizsgálat eredményei alapján (lásd részletesebben a KER, 2002. „B” függelékét), és így kilenc szintet határoz meg. A Referenciakeret megkülönbözteti a már ismertetett hat „kritériumszintet” (A2; B1; B2 stb.) és három ún. „plusz szintet” (A2+; B1+; B2+). A kritériumszint és a plusz szint megkülönböztetése szubjektív, azonban így lehetőség nyílik arra, „hogy a szintek és/ vagy szintleírások közös rendszerét különböző felhasználók helyi szükségletei kielégítése érdekében eltérő gyakorlati szintekre oszthatják, és mégis a közös rendszerhez kapcsolódnak” (KER, 2002. 41. o.). A plusz szint olyan „ismeretet határoz meg, amely magasabb a kritériumszint által képviselnél, ám nem éri el a következő szintet” (KER, 2002. 47. o.).

A Referenciakeret vízszintes dimenziói ismertetik azokat a kategóriákat, amelyek a kommunikatív tevékenység és a kommunikatív nyelvi kompetencia paramétereit (deskriptorait) tartalmazzák. Ezek meghatározzák, hogy mi várható el a nyelvtanulótól a különböző szinteken, így lehetővé teszi annak eldöntését, hogy az adott területeken a tanulók teljesítménye megfelel-e a nyelvtudás adott szintjéhez rendelhető színvonalnak (KER, 2002). A Referenciakeret szintleírásai pozitív irányultságúak, cselekvés-orientált megközelítésűek, azaz arra koncentrálnak, hogy a tanuló mit tud a nyelvvel kezdeni. A szintek leírásának alapelve, hogy ne legyen pontatlan, túl általános, hanem a szintek között konkrét különbséget lehessen tenni, azaz a szintrendszer legyen átlátható, felépítése világos, logikus, megfogalmazása rövid, így a szintleírások kontextus nélkül is megállnak, ezáltal téve lehetővé, hogy kritérium-listává alakítva az értékelés során is használhatóak legyenek (Glaboniat és Müller, 2006; KER, 2002).

A nyelvi tesztelés szempontjából a Közös Európai Referenciakeretet egyre több kritika is éri. A Referenciakeret szintleírásait a *Dutch CEF Construct Project* (Alderson és mtsai, 2004) keretében elemezték, illetve *Quetz* (2001, 2003) is, aki a dokumentum német nyelvi fordításában vett részt. Ezek az elemzések rámutattak arra, hogy a szintek leírása nyelviileg nem mindenhol pontosak, hiszen a vízszintes dimenzió leírásához alkalmazott fogalmak és kategóriák különböző szintekbe is besorolhatók. Ezen kívül a szintek túldimenzionáltak, így a Referenciakeret szintleírásai nehezen használhatóak például a tesztfeladatok készítéséhez, a produktív nyelvi készségek mérésénél javítási- és értékelési eljárások kialakításához (Altmayer, 2004). A Referenciakeret azonban fontos és hasznos dokumentum, alkalmazásához viszont szükséges a nyelveken átívelő szintleírások konkretizálása, így megjelennek olyan kiadványok is, amelyek a Közös Európai Referenciakeret alapján egy konkrét nyelvre kidolgozva tartalmazzák többek között a kommunikációs szándékokat, a nyelvtani szerkezeteket és tematikus szóincset, amelyeket a vizsgán vagy receptív, vagy produktív szinten kell tudni alkalmazni. Ezeket az információkat tartalmazza például a *Profile Deutsch* (Glaboniat és mtsai, 2002, 2005) a Referenciakeret összes szintjére. A program a szintleírások között úgy differenciál, hogy definiálja az általános (*Globale Kann-Beschreibungen*), és a részletes szintleírásokat (*Detaillierte Kann-Beschreibungen*). Az általános szintleírás szerepe, hogy meghatározza, hogy a vizsgázó milyen jól tudja használni a nyelvet, a részletes pedig megadja, hogy

konkrétan mit tud a vizsgázó (Altmayer, 2004; Boócz-Barna, 2004; Glaboniat és Müller, 2006; Perlmann-Balme, 2006; Seiffert, 2003).

A kommunikatív tesztek jóságmutatói

A kommunikatív teszteknek különböző minőségi kritériumoknak kell megfelelniük, amelyek meghatározásában a klasszikus tesztelmélet fontos szerepet tölt be. Ez a tesztelmélet a tesztben mért értéket a valódi érték és valamilyen hiba összegeként határozza meg, és elsősorban a valódi érték, illetve a hiba becslésével foglalkozik. Ez alapján határozza meg egy teszt minőségét, azaz a teszt objektivitását, reliabilitását és validitását. A klasszikus tesztelmélet elsősorban nem a nyelvi tesztekre koncentrált, hanem általánosan a pszichológiai tesztekre, így a szakirodalomban (pl. Bachman és Palmer, 1996; Lienert és Raatz, 1994) a jóságmutatóit az idegen nyelvi tesztek esetén nem tartják elég átfogónak, ezért ezeket tovább differenciálják és más kritériumokat is megneveznek. A továbbiakban a Bachman és Palmer (1996) által a kommunikatív tesztek jellemző kritériumokat mutatjuk be, amelyek összhangban állnak a klasszikus tesztelmélet jóságmutatóival, de szükségképpen más aspektusra helyezik a hangsúlyt.

Bachman és Palmer (1996) a *usefulness*-t (hasznosság, használhatóság) tartják a legfőbb kritériumnak: „The most important consideration in designing and developing a language test is the use for which it is intended, so that the most important quality of a test is its usefulness.” (Bachman és Palmer, 1996. 17. o.) Ezt tükrözi a szerzőpáros könyvének címe *Language Testing in Practice: Designing and Developing Useful Language Tests*. Ezt a kritériumot mintegy gyűjtőfogalomként kívánták bevezetni a tesztelés jóságmutatói fölé. Ez a törekvésük azonban nem sikerült, a szakma nem vette át ezt a terminust, így a magyar nyelvű szakirodalom nem ad fordítást a *usefulness* fogalomra, a német nyelvűben a *Nützlichkeit* kifejezés jelenik meg (Grotjahn, 2000a, 2006; Krekeler, 2005; Tschirner, 2001). A továbbiakban a *használhatóság* kifejezést alkalmazzuk e gyűjtőfogalom jelölésére, amelyet hat egymást kiegészítő tulajdonság (1) a reliabilitás (*reliability*), (2) a szerkezeti validitás (*construct validity*), (3) az autentikusság (*authenticity*), (4) az interaktivitás, (*interactiveness*) (5) a teszthatás (*impact*) és (6) a hatékonyság/ gazdaságosság/ praktikusság² (*practicality*) ötvözeteként lehet meghatározni (Bachman és Palmer, 1996). Az angol és a német szakirodalomban gyakran sorolják fel, jellemzik az összetevőket (lásd pl. Alderson, 2006; Alderson és mtsai, 2004; Bachman és Palmer 2005; Grotjahn 2000a, 2006; Krekeler, 2005; Tschirner, 2001). A továbbiakban ezek felhasználásával mutatjuk be a használhatósági kritérium összetevőit.

(1) A nyelvi tesztelésben a *reliabilitás* ugyanazt jelenti, mint a klasszikus tesztelméletben. Jelentőségét az is mutatja, hogy a fent hivatkozott szakirodalmakon túl is, sokan foglalkoznak a reliabilitás kérdésével (lásd pl. Bachman, 1990; Bárdos, 2002a; Ennis, 1999; Morrow, 1981). Egy teszt akkor megbízható, ha a teszteredmények reprodukálha-

² A *practicality* fordításában a német szakirodalom a *Praktikabilität* (Tschirner, 2001), és a *(Test)ökonomie* (Krekeler, 2005) kifejezést használja.

tóak, azaz ha a teszteredmény csak a mérendő tulajdonságtól függ. A reliabilitás rendkívül fontos a teszt funkciója szerinti alkalmazásához.

A reliabilitásnak három aspektusát különböztetik meg: Egyrészt, a teszt belső következetessége (*internal consistency*) az itemek közötti különbségekre vonatkozik. Számításakor különböző felezési módszerekkel (*split-half-method*) a teszt két felét egymással korreláltatják. A reliabilitás számításának módszerei között találjuk a Cronbach alfa, a Spearman-Brown és a dichotóm itemekre vonatkozó Kuder-Richardson 20 formulát, amelyek a reliabilitást alulról becslik. Másrészt, a teszt ismétlése (*test-retest reliability*) a teszteredmények összehasonlíthatóságára vonatkozik, amennyiben a tesztet megismétlik. A reliabilitás a két teszten elért pontszámok korrelációjával határozható meg, ezen kívül a középértékeket is össze lehet hasonlítani. Harmadrészt pedig, a párhuzamos tesztelés egy teszt különböző verzióira vonatkozik és a tesztek különböző variációinak korrelációjával határozható meg.

A tesztek reliabilitása nagymértékben függ a kiértékelés objektivitásától, így főként a szubjektív elbírálású teszteknel fontos a vizsgáztatók képzése. Itt térünk ki arra, hogy a használhatósági kritérium összetevői közül az objektivitás, mint a klasszikus tesztelmélet fontos jóságmutatója hiányzik, mivel *Bachman* és *Palmer* (1996) az adatfelvételi és kiértékelési objektivitást a reliabilitáshoz számítják. Mivel a reliabilitás populációfüggő, így a teszt megbízhatóságát jelentősen csökkentik olyan itemek, amelyek a vizsgázók között kevésbé differenciálnak. E problémát vagy a teszt itemszámának növelésével, vagy a teszt bemérése után, a gyengén differenciáló itemek megváltoztatásával lehet megoldani.

(2) A validitástípusok közül a kommunikatív tesztelésben fontos szerepe van a *szerkezeti validitás*nak, amely azt határozza meg, hogy az adott teszt eredményei igazolják-e a vizsgált nyelvi jelenségek konstruktumának meglétét. A teszten elért eredmény egyfelől arról ad információkat, hogy a vizsgált személy hogyan teljesített, másrészt arról kellene információt adnia, hogy a jelölt egy valós nyelvi szituációban hogyan viselkedne. Így a szerkezeti validitás a kommunikatív tesztek esetén elsősorban az eredmények interpretálásának alapját képezi, és azt nem a teszt tulajdonságaként fogják fel (*Bárdos*, 2002a; *Messick*, 1996; *Young*, 2000).

A konstruktumot kétféle módon definiálhatjuk. Egyrészt a nyelvtudás szerkezetének elméleti modellezésével, másrészt empirikusan a vizsgázók nyelvi igényeinek (*Bedarfsdiagnose*) felmérése (*Tschirner*, 2001) alapján, úgy, hogy pontosan definiáljuk azokat a kontextusokat és kommunikációs helyzeteket, amelyekben a vizsgázónak nyelvi cselekedniük kell, majd ezekből a vizsga készítésekor reprezentatív választást teszünk lehetővé. A vizsga annál érvényesebb, minél több kommunikációs helyzetet reprezentálnak a tesztfeladatok. Így tehát fontos, hogy egy vizsga a célcsoportra minél precízebben legyen illesztve (*Tschirner*, 2001).

Messick (1996) koncepciójában a klasszikus tesztelmélet validitástípusai nem egymás mellett állnak, hanem a szerkezeti validitás a többi fölött áll, és a többit is belefoglalja. *Messick* (1996) a szerkezeti validitás hat aspektusát határozza meg:

- tartalmi aspektus (*content aspect*; a tartalom reprezentativitása),
- lényegi aspektus (*substantive aspect*; az elméleti konstruktummal való egyezés),
- szerkezeti aspektus (*structural aspect*; a mérőeszközök egyezése a konstruktum szerkezetével),

- általánosíthatóság aspektusa (*generalizability aspect*; a mérési eljárás reprezentativitása a konstruktumra vetítve),
- külső aspektus (*external aspect*; a mérési eredmények egyezése az elméleti konstruktummal),
- hatás aspektusa (*consequential aspect*; a teszteredmények interpretációjának hatása).

A szerkezeti validitás egyrészt akkor teljesül, ha a teszt a mérni kívánt nyelvi készség konstruktumának kevés olyan aspektusát tartalmazza, amelyet a teszt nem mér (*construct underrepresentation*), másrészt, ha a teszt a vizsgázó tulajdonságainak kevés olyan aspektusát méri és értékeli, amely a konstruktum számára irreleváns (*construct irrelevance*) (Alderson, 2002, idézi: Krekeler, 2005; Brualdi, 1999; Grotjahn, 2000a).

Ezek a szempontok Bachman és Palmer (1996) használhatósági kritériumának leírásában úgy jelennek meg, hogy a validitás különböző aspektusaiként határozzák meg az autentikusságot, az interaktivitást és a teszthatásokat.

(3) Az *autentikusság* a tesztfeladatok fontos minőségi jellemzőjét adja meg. Annak a módnak, ahogy az idegen nyelvet a tesztfeladatban használjuk, meg kell egyeznie azzal, ahogy azt valós nyelvi szituációban használnánk. Ez az alapelv nemcsak az autentikus (a célnyelven valóban előforduló, és nem pedagógiai céllal a nyelvtanulóknak íródott) szövegek alkalmazásával teljesül, hanem akkor is, ha az autentikus tesztfeladatokban egyrészt egyértelművé tesszük, hogy milyen szituációban, milyen céllal kell elvégezni a nyelvi feladatot, és az mikor tekinthető sikeresnek (*interakciós autentikusság*). Másrészt pedig, ha a szöveg nyelvi kihívásai, valamint a feladat jellemző vonásai megegyeznek azokkal a tipikus tulajdonságokkal, amelyek a célnyelvi élethelyzetekben elvárhatók (*szituációs autentikusság*) (Bárdos, 2003, 2004; Sigott, 1996). Egy teszt autentikusságát nem lehet statisztikai úton meghatározni. A tesztfeladatban megadott nyelvi szituáció tartalmi elemzésével azonban meg lehet adni az autentikusság bizonyos fokát (McNamara, 2000; Messick, 1996; Takala, 1998). A magas autentikussággal rendelkező kommunikatív teszt nagymértékben járul hozzá a teszt használhatóságához. Azok az autentikus nyelvi szituációk, amelyek a konstruktummal összefüggésben állnak, megkönnyítik a teszteredmények megfelelő értelmezését is, és az autentikus feladatokat a vizsgázók is pozitívabban ítélik meg (Bachman és Palmer, 1996).

(4) Az *interaktivitás* arra vonatkozik, hogy a vizsgázók milyen mértékben vonják be a tesztfeladat megoldásába a nyelvi készségeiket, és egy adott témáról megszerzett előzetes háttértudásukat és ismereteiket (Bachman és Palmer, 1996). Ez a kritérium egyrészt hatással van a reliabilitásra, másrészt a validitásra is. A tesztnek annál nagyobb a megbízhatósága, minél kevésbé befolyásolják a tesztben elért eredményt a vizsgázók emocionális reakciói, illetve azok az ismeretek és képességek, amelyeket a teszt nem kíván mérni. A kommunikatív teszt validitása annál nagyobb, minél inkább bevonja a jelölt a nyelvtudását a feladat megoldásába. Egy teszt interaktivitását tartalmi elemzéssel, a vizsgázók viselkedésének megfigyelésével és a teszteredmények elemzésével határozhatjuk meg. Az autentikussághoz hasonlóan csak leíró eredményhez lehet jutni, azaz az interaktivitás bizonyos szintjét lehet csak meghatározni.

(5) *Bachman és Palmer* (1996) a teszt hatásait (*impact*) két szinten különböztetik meg. A mikroszintű teszthatás³ a tanítási-tanulási folyamatban résztvevőkre és a tanulási módszerekre, a makroszintű hatás pedig a társadalomra és az iskolarendszerre vonatkozik. A visszahatás kétféle irányú lehet: pozitív vagy negatív. Abban az esetben, ha a bevezetett vizsga modernizálja az oktatási folyamatot, az alkalmazott módszereket, illetve az oktatás tartalmát, akkor a vizsgának pozitív a hatása. Ha viszont a nyelvtanítás középpontjában a vizsgán előforduló feladattípusok és vizsgázási stratégiák tanítása áll (*Krumm, 2006; Perlmann-Balme, 2006; Prodromou, 1995*), tehát a vizsgára való felkészítés „titkos” tantervvé („heimlicher” Curriculum) válik *Glaboniat és Müller, 2006*), akkor negatív hatásról van szó. Ebben az esetben az a kérdés is felmerül, hogy valójában mit mérünk, a feladat-megoldási képességet vagy a nyelvtudást, tehát a validitás is sérülhet. Ez is jelzi, hogy a visszahatás a validitás egyik aspektusa, ezért ezt szokták *következményes validitásnak (consequential validity)* is nevezni (*Alderson, 2004; Dévény, 2003*). E kritérium jelentőségét mutatja, hogy a vizsgarendszerek hatásának vizsgálata egyre nagyobb hangsúlyt kap (pl. *Horváth és Lukács, 2006*).

(6) A *hatékonyság* kritériuma egy vizsga lebonyolításához a szükséges, és az arra fordított fejlesztés között meglévő helyes arányra vonatkozik, amelyet számolással úgy lehet meghatározni, hogy elosztják a rendelkezésre álló forrásokat, a szükséges forrással. Abban az esetben, ha az eredmény egy, vagy annál nagyobb, a teszt hatékony, ha egynél kisebb, akkor nem hatékony (*Bachman és Palmer, 1996, 2005*). A gyakorlatban így nehezen lehet számszerűsíteni egy teszt hatékonyságát, mivel a szükséges forrás mennyiségéről sokszor nincs pontos információnk. E kritérium megadásának inkább az a célja, hogy egy tesztet olyan hatékonyra tegyünk, amennyire csak lehet, a kritérium tágabb értelemben véve arra hívja fel a figyelmet, hogy egy vizsgarendszer működtetésekor az optimálisnak megfelelő, legszükségesebb erőforrásokat használjuk fel (*Krekeler, 2005*), mivel a vizsgafeladatok előkészítésének és a vizsga lebonyolításának lépései jelentősen befolyásolják a vizsga minőségét, megbízhatóságát (*Einhorn, 2006a*).

A kommunikatív teszt használhatóságának (*usefulness*) biztosítása, kimutatása és elemzése kapcsán *Bachman és Palmer* (1996) három alapelvet nevez meg: Egyrészt, egy teszt használhatóságát kell maximálni, és nem a kritérium egyes elemeit (pl. reliabilitás). Másrészt, a használhatósági kritérium komponenseit nem lehet egymástól függetlenül megítélni, hanem kombinált hatásukat kell vizsgálni. Harmadrészt, a használhatóság fokát és az egyes komponensek közti megfelelő arányt csak az aktuális tesztszituációra vonatkoztatva lehet meghatározni.

A kommunikatív tesztek típusai

A megfelelő használhatósági mutatókkal rendelkező kommunikatív tesztek többféle szempont szerint lehet osztályozni. *Bachman* (1990) (1) a teszt célja, funkciója (*intended*

³ Az angol nyelvű szakirodalomban a *washback* (pl. *Messick, 1996*) és a *backwash effect* (pl. *Prodromou, 1995*) fogalma egyaránt megtalálható.

use); (2) tartalma (*content*); (3) viszonyítási rendszere (*frame of reference*); (4) értékelésének típusa (*scoring procedure*); és (5) a bevezetett teszttechnikák és tesztműszerek (*testing method*) szerint határozza meg a nyelvi tesztek.

A továbbiakban a kommunikatív teszteket először a különböző funkciók és a tartalom szerint osztályozzuk, és meghatározzuk, hogy a teszt típusa mire vonatkozik, majd foglalkozunk a standardizált és nem standardizált, a norma- és kritériumorientált, a performancia és a kompetencia tesztek készítésének problematikájával, alkalmazásuk dilemmájával.

A kommunikatív tesztek funkciói

A kommunikatív tesztek szerkesztésének fő szempontja a teszt funkciója (célja), amely befolyásolja azt, hogy melyek azok a tartalmak, illetve készségek, amelyeket a tesztben mérni kívánunk. A kommunikatív tesztek funkciójának meghatározásakor egyrészt meg kell állapítani, hogy mely feltételek alapján alkalmazzuk a tesztek bizonyos célok mérésére, másrészt pedig, hogy a tesztek alkalmazásának milyen következményei, hatásai vannak (Krekeler, 2005; Perlmann-Balme, 2001).

A *jövőbeli bevételre vonatkozó nyelvi tesztek* közé tartoznak a nyelvtanulási képességek tesztjei (*language aptitude tests/ Eignungstests*). Ezek olyan prognosztikai tesztek, amelyek a nyelvtanulás sebességét jósolják meg, így a nyelvtanulók sikerét jelezi előre (Bárdos, 2002a). Az ebbe a csoportba tartozó besoroló tesztekre (*placement tests/ Einstufungstests*) jellemző, hogy a tanulói csoportokat változatos, többnyire objektív tesztelési technikával alakítják ki (Albers és Bolton, 1995; Bárdos, 2002a). A szelekciós teszt/ felvételi vizsga (*selection tests*, illetve *gatekeeping tests/ Zulassungsprüfungen*) célja a jelentkezők szelektálása, így, mivel itt a diszkriminatívitas a legfontosabb szempont, olyan tesztfeladatokat használnak, amelyek jól differenciálnak a jelöltek között. Egy besoroló teszt is lehet szelekciós teszt, azonban ez a teszt típus sokkal átfogóbb, mint a besoroló teszt. Abban az esetben, ha a teszteredmények között tovább differenciálnak, és ez arról is dönt, hogy a jelöltet melyik kurzusra veszik fel, és mely készségeket kell tovább fejleszteni, akkor inkább besoroló tesztről van szó (Bachman és Palmer, 1996; Krekeler, 2005; Perlmann-Balme, 2001).

A *tantervre vonatkozó nyelvi tesztek* csoportjába tartozó teljesítménytesztek (*achievement tests*) azt vizsgálják, hogy a tananyagot milyen mértékben sikerült elsajátítani, és az eredményeket a tantervre vonatkoztatva értelmezik. Ezek a tesztek a tanítási-tanulási folyamat eredményességére utalnak. A tesztfeladatokat nagyon gyakran a tanárok készítik. A teljesítményteszteknek két csoportját lehet elkülöníteni. (1) A haladási tesztek (*progress tests/ Lernfortschritstests*) célja a további tennivalók feltérképezése. Az elért eredmény következménye nem túl jelentős. (2) Egy tanítási-tanulási folyamatot lezáró tesztek (*exit tests/ Kursabschlussstests*) nagy részletességgel, átfogó módon azt vizsgálják, hogy a tanulóknak milyen mértékben sikerült elsajátítani egy kurzus teljes anyagát. A teszt eredményének következménye messzemenőbb, mivel azt is meghatározzák, hogy egy adott kurzust meg kell-e ismételni, illetve sikeressége belépőül szolgálhat egy magasabb szintű kurzusba (Albers és Bolton, 1995; Bárdos, 2002a; Davies és mtsai, 1999; KER, 2002; Krekeler, 2005; Perlmann-Balme, 2001).

A *konstruktumra vonatkozó nyelvi tesztek* csoportjába sorolhatók a készségszinteket mérő tesztek (*proficiency tests/ Feststellungsprüfungen*), amelyek a tanítási-tanulási folyamattól függetlenül mérik a vizsgázó nyelvi készségeit. A tesztek többnyire standardizáltak, a feladatokat központilag készítik, és többnyire képzett, külső értékelők minősítenek (Einhorn, 1999). Ezeknek a teszteknek a célja az elért tudásszint mérése a viszonylagos teljesség igényével, és így a nyelvtudásról többé-kevésbé átfogó képet adnak. Az eredmények kívülállóknak számára is értelmezhetők (Krekeler, 2005; Perlmann-Balme, 2001). A készségszinteket vizsgáló tesztek két csoportra bonthatók: (1) az általános, azaz a nyelvi tartalom közvetítését és a készségeket, valamint (2) egy specifikus terület szaknyelvét vizsgáló tesztek (Bárdos, 2002a).

A nyelvi tesztek ilyen típusú felosztásának célja, hogy a teszt tulajdonságát egy bizonyos funkcióval együtt mutassa be. Ez még akkor is hasznos, ha egy teszt funkcióját nem lehet mindig egyértelműen meghatározni. Ennek oka egyrészt az egyes kategóriák közti átfedések, másrészt a tesztkészítők, -alkalmazók és a vizsgázók eltérő érdekei (Davies és mtsai, 1999).

A különböző tesztfunkciók a validitás eltérő aspektusait jelenítik meg (Bachman, 1990; Brualdi, 1999; Krekeler, 2005). A szelekciós tesztek fő kérdése, hogy hogyan határozzák meg a bekerülés határát. A küszöbértékek meghatározásánál a megfelelési validitás (*criterion-related validity*) játszik központi szerepet, amely megmutatja, hogy a vizsgázó elérte-e az előre megadott megfelelési (*threshold*) szintet, amely a nyelvtudás bizonyos minőségét jelöli. A megfelelési validitás vagy úgy határozható meg, hogy a teszteredményt a teszt megfelelési szintjével korreláltatjuk, vagy egy olyan teszttel, amelynek megfelelési validitása ismert. Ezt a fajta validitástípust a szakértők véleménye alapján is meghatározhatjuk (Bárdos, 2002a; Brualdi, 1999; Davies és mtsai, 1999; Grotjahn, 2000a, 2006). A teljesítménytesztek esetén a tartalmi validitás (*content validity*) teljesülése az elsődleges szempont, amely annak mértékét adja meg, hogy a teszt tartalma mennyire méri és reprezentálja a kurzus teljes tartalmát (Grotjahn, 2000a). A tartalmi validitást a szakértők véleménye vagy más tesztek korrelációja alapján lehet meghatározni (Alderson és mtsai, 1995; Bárdos, 2002a; KER, 2002). A készségszinteket vizsgáló tesztek esetén a megegyező/ egyidejű validitás (*concurrent validity*) és az előrejelző validitás (*predictive validity*) játszik központi szerepet. A megegyező validitás a teszten elért eredmény megegyezése más tesztek eredményével, amit korrelációval határozhatunk meg. A prediktív validitás pedig a teszteredmény és annak a kritériumnak az egyezése, amely egy a jövőbeli valós nyelvhasználati szituációra vonatkozik (Alderson és mtsai, 1995). A korábban részletesen ismertetett szerkezeti validitás és a triviális-/látszatvaliditás (*face validity*) minden tesztípus esetén fontos szerepet tölt be. E validitás a laikusok véleményét tartalmazza egy teszt érvényességéről. Ennek a validitástípusnak a megítélése a szakirodalomban nem egységes, mivel ezt egyrészt populáris, de a teszt irreleváns aspektusának tartják (pl. Stevenson, 1985 „*pop validity*”-nek nevezi; idézi Krekeler, 2005), másrészt viszont, ha a vizsgázók egy tesztet értelmesnek tartanak, jobb eredményt akarnak elérni (Bachman és Palmer, 1996; Grotjahn, 2000a).

Standardizált, informális és nem standardizált formális tesztek

A *standardizált tesztek*et úgy fejlesztik, hogy a legfontosabb jóságmutatókat elemelik és az adatfelvételt, a kiértékelést és az eredmények értelmezését is standardizálják, szabványosítják, és normát állapítanak meg. A szabványosítás csak abban az esetben alkalmazható, amennyiben a teszt jóságmutatóit empirikus elemzések alapján is biztosították.

Az *informális tesztek*et a tanárok készítik, és saját célokra használják fel, ezeknél a teszteknel a szabványosítás nem jelenik meg kritériumként, azonban az informális tesztek legfontosabb jóságmutatója a szituáció érvényessége, amelyben a tesztet alkalmazzák, valamint a tartalmi validitás (Csapó, 1996; Dávid, 1992; Grotjahn, 2000a; Lienert és Raatz, 1994).

A standardizált és az informális tesztek között helyezhetők el a *nem standardizált formális tesztek*, amelyek nem felelnek meg a standardizáció követelményének, de céljuk, hogy objektív döntéseket hozzanak a vizsgázók nyelvtudásáról, így az ilyen tesztek elkészítésénél ügyelnek a használhatósági kritériumok betartására. Erre a teszt típusra példa a német *Deutsche Sprachprüfung für den Hochschulzugang (DSH)* vizsgarendszer (Krekeler, 2005).

Norma- és kritériumorientált tesztek

A *norma- és kritériumorientált tesztek* közti megkülönböztetés elsődleges szempontja a teszteredmények értelmezésének módja. Ezen kívül a két tesztforma a tesztfejlesztés folyamatának, struktúrájának, tartalmának és a célok meghatározásának módjában is különbözik (Bond, 1996; Csapó, 1987, 2004; Grotjahn, 2000a, 2006).

A normaorientált tesztek fő jellemzője, hogy az egyén eredményeit vagy egy viszonyítási csoport eredményeihez vagy más vizsgázók teljesítményéhez képest értelmezik (Glaboniat és Müller, 2006; Grotjahn, 2000a, 2006). Mivel ezeknek a teszteknek az a fő célja, hogy a vizsgázók teljesítménye között különbséget tegyenek, így erősen differenciáló itemekre van szükség (Gronlund, 1988; Krekeler, 2005).

A kritériumorientált tesztek célja az egyéni készségek leírása. A teszt a kritériumot reprezentálja, amelynek eredményével az egyéni készségszintet az elvárt készségszinttel hasonlítják össze. A kritériumorientált tesztelés esetén az elérendő célokat szintekre bontják, minden egyes szintnek részletes leírása van, amely alapján a jelölt valamelyikbe besorolhatóvá válik. Ebben az esetben meghatározható, hogy a nyelvhasználó az adott szinten mire képes (Bárdos, 2002a; Glaboniat és Müller, 2006; Perlmann-Balme, 2006). A kritériumorientált tesztelés magába foglalja a készségszint folyamatosságának (a Referenciakeret függőleges dimenziójának) és megfelelő tartományok körének (vízszintes dimenziójának) feltérképezését is annak érdekében, hogy a tesztben elért egyéni eredmények a teljes kritériumsorban elhelyezhetők legyenek (KER, 2002). A kritériumorientált teszteknel előfordulhat, hogy minden vizsgázó azonos eredményt ér el. Mivel a kritériumorientált teszteknel más vizsgázók eredményének nincs szerepe, így az értékelés a csoportra viszonyítva abszolút. A kritériumorientált teszteknek gyakran van kapcsolata egy bizonyos tantervvel és a tanítási-tanulási folyamattal is. Ezekben a tesztekben a

vizsgálók pontosan tudják, hogy mely tartalmat és készségeket mérik, és mivel a teszt az oktatási folyamatra irányul, magas ezeknek a teszteknek a látszatvaliditása és a visszahatása (Grotjahn, 2000a; Krekeler, 2005).

Performancia és kompetencia tesztek

A *kompetencia és performancia tesztek* megkülönböztetése Chomsky (1965) nyelvi kompetencia és performancia fogalmára vezethető vissza.

A kompetencia tesztek azt mérik, hogy a vizsgázó egy konkrét nyelvi szituációtól függetlenül hogyan tudja a nyelvet általánosan használni. Ezek a tesztek abból a feltételezésből indulnak ki, hogy a nyelv különböző, egymással kapcsolatban levő komponensekből áll, amelyeket egymástól függetlenül lehet mérni és értékelni. A kompetencia tesztek szerkesztésének alapja egy olyan átfogó és részletes nyelvi modell, amely a nyelvi készségek és képességek szerkezetét és a köztük levő kapcsolatot is meghatározza.

A kompetencia tesztek sokoldalúan alkalmazhatók, az eredményekből jól lehet általánosítani más teszten kívüli nyelvi szituációra. Ehhez azonban szükséges, hogy a teszt a nyelvi készségeket és képességeket leíró modellt minél jobban reprezentálja, mivel így nagyobb azoknak a nyelvi szituációknak a száma, amelyeket a teszt előre tud jelezni. A teszteknek a hátránya, hogy nem sok közülük van a valós nyelvi tevékenységhez (például a tesztben a szavak aláhúzása, néhány szó beírása stb.), így a látszatvaliditásuk is alacsony. Ennek ellenére a zárt feladatok alkalmazása objektív értékelést és magas reliabilitást biztosít (Grotjahn, 2000a; Krekeler, 2005; Niska, 1997; Robinson, 1996; Skehan, 1998).

A performancia tesztek mérési és értékelési eljárása abban különbözik a kompetencia tesztekétől, hogy a tesztben egy realitáshoz nagyon közeli nyelvi szituációt képezünk le, és a vizsgázókat egy ilyen szituációval szembesítjük. Így lehetőség nyílik arra, hogy a valós nyelvi teljesítményre többé-kevésbé közvetlenül következtessünk. Egy nyelvi performancia tesztben azt a nyelvi képességet tudjuk mérni, hogy a vizsgázó hogyan tud egy bizonyos nyelvi szituációban viselkedni. (Bárdos, 2002a; McNamara, 2000; Robinson, 1996; Young, 2000). McNamara (1996; idézi: Grotjahn, 2000a) az alábbi módon tovább osztja fel a nyelvi performancia tesztek: azokat a tesztek, amelyek kizárólag csak a nyelvi teljesítményre vonatkoznak, szűkebb értelemben vett performancia teszteknek (*weak language performance tests*) nevezzük. Ezzel szemben azokat a tesztek, amelyek a feladat megoldását, és így a kommunikációs szituáció megvalósítását is értékelik, tágabb értelemben vett performancia teszteknek (*strong language performance tests*) nevezzük.

A performancia teszteknek magas a validitásuk. A valós nyelvi szituációk alkalmazása biztosítja az eredmények értelmezését a konstruktummal való összevetés alapján, ami ideális esetben egymással megegyezik. Az ilyen teszteknek viszonylag magas az elfogadottsága (látszatvaliditása), illetve pozitív a visszahatása a tanítási-tanulási folyamatra. A performancia tesztek eredményeiből viszont nem tudunk pontos következtetéseket levonni, hogy egy másik nyelvi szituációban hogyan teljesítene a vizsgázó. Ezeknek a teszteknek a reliabilitása is problémás, mivel az értékelés szubjektivitása is megnő. Néhány szerző a közvetlen performancia teszteknel már eleve azt is megkérdőjelezi, hogy egyáltalán le lehet-e pontosan képezni egy reális nyelvi szituációt? A tesztfeladat auten-

tikussága azért fontos, mert ha hiányzik, akkor a teszt nem közvetlen, és az eredmények értelmezésének validitása is megkérdőjeleződik. Ezen kívül a tesztfeladatban megjelenő nyelvi szituációk kiválasztásának reprezentatívnak kell lennie a konstruktumra nézve (Davies és mtsai, 1999; Grotjahn, 2000a; Krekeler, 2005; McNamara, 2000; Skehan, 1998; Takala, 1998).

Annak meghatározása, hogy egy performancia vagy kompetencia tesztet alkalmazunk-e, elsősorban a tesztek funkciójától függ. Abban az esetben, ha egy tesztnek a vizsgázó teljesítményéről konkrét nyelvi szituációban kell információt adnia, akkor a performancia teszt alkalmazása lehetséges (például a produktív készségeket mérő feladatoknál ilyen tesztet lehet alkalmazni). Ha az eredményeket, mint különböző nyelvi szituációkra való utalásokat szeretnénk értelmezni, akkor a kompetenciatesztek a megfelelőbbek (Niska, 1997; Robinson, 1996).

Közvetlen, közvetett és félig közvetett tesztek

A *közvetlen* és a *közvetett tesztek* megkülönböztetése az eredmények eltérő értelmezésére vonatkozik, és a teszt tartalomnak a konstruktumhoz való viszonyát tükrözi. A közvetlen teszteljárásnál a teszteredményből egyből lehet következtetni a konstruktumra, mivel a tesztben és a valóságban nyújtott teljesítmény egymással megegyezik. Minél közvetettebb a tesztelési eljárás, annál nagyobb szükség van az eredmények interpretálásához a konstruktum modellezésére. A közvetett tesztek esetén nem magát a mérni kívánt képességet vizsgáljuk, hanem annak csak egy fontos elemét, és aztán ezt az elemet a teljes képességgel korreláltatjuk (Bárdos, 2002a; KER, 2002; Robinson, 1996). Ilyen teszteknel például kimutatható, hogy a szókincs gazdagsága nagyon jól korrelál az olvasott szöveg értésével (Pike, 1979, idézi Tschirner, 2001). Ez ahhoz vezetett, hogy a diszkrét pontos tesztelés esetén az olvasásértést sokáig csak a szókincs tesztelésével mérték.

A közvetlen és a közvetett tesztek mellett megkülönböztetik még a *félig közvetett tesztek*et (Tschirner, 2001) is, amelyeknél azt a képességet, amelynek mérését célul tűzzük ki magunknak, integrálva teszteljük, tehát a képességet nem bontjuk elemeire, és nem mérjük külön. Ezeknél a teszteknel a mérés sem teljesen autentikus módon történik, mivel az autentikusságot csak szimulálják. A *TestDaF* vizsga szóbeli részén alkalmazott szimulált szóbeli interjú (*Simulated Oral Proficiency Interview; SOPI*) példa erre a teszt-típusra, amelyben a megadott input feladat standardizálásával és a felvett hanganyag központi értékelésével lehetőség nyílik az eljárás objektívitasának, reliabilitásának, és a validitásának a növelésére is (Grotjahn, 2000a; Malone, 2000).

A gyakorlat azt mutatja, hogy a performancia tesztek közvetlen tesztek, míg a kompetenciatesztek közvetett tesztek. Ez a megkülönböztetés azonban azért problémás, mert egy bizonyos szinten minden teszt közvetett. A tesztfeladatban megadott nyelvi szituációk (az autentikusságra való törekvés ellenére is) mesterségesek, a mérés nem kívánt beavatkozás; a vizsgahelyzet irreális és nyelvezete természetellenes (Bárdos, 2003, 2004). A közvetlen nyelvi tesztek gyakran nem tudják a hozzájuk fűzött elvárásokat teljesíteni. Az azonban megállapítható, hogy a nyelvi tesztek különbözőképpen közvetettek lehet-

nek, így a konstruktum mérése és értékelése vagy közelebb, vagy távolabb kerül (Bachman, 1990; Messick, 1996).

A kommunikatív tesztelés alkalmazásai

Egy kommunikatív teszt elkészítésekor, illetve tágabb értelemben véve egy vizsga működtetésekor folyamatos mérlegelésre, fejlesztő munkára van szükség. Ehhez vizsgálni kell a használhatósági kritérium összetevőit, azok egymásra gyakorolt hatását, a különböző vizsgarészekben alkalmazott tesztfeladatokat, értékelési eljárásokat, valamint a teljes vizsgarendszer validitását, amely „kizárólag a vizsgához kapcsolódó döntések, az eredmények adott értelmezése és a vizsga következményei fényében ítéltető meg” (Dávid, 2005. 279. o.). Ezt nevezzük a vizsga belső validitásának, amely kérdéskörhöz tartozik, hogy a vizsgarendszer mérési modellje idomuljon az alkalmazott nyelvtudásmodellhez, ehhez azonban szükséges, hogy megvizsgálják érvényességét a vizsgarendszer működése során (Dávid, 2002).

2. táblázat. Különböző vizsgarendszerek szintjei (Barabás, 2004. 25. o.⁴; Bárdos, 2005. 10. o.; University of Cambridge ESOL Exams.; Zertifikate und Prüfungen für DaF, 2006. 45. o. alapján.)

KER szintjei	University of Cambridge	Goethe Institut	Magyarországi államilag elismert nyelvvizsgák
C2	CPE (Certificate of Proficiency in English)	ZOP (Zentrale Oberstufenprüfung) KDS (Kleines Deutsches Sprachdiplom)	
C1	CAE (Certificate in Advanced English)	ZMP (Zentrale Mittelstufenprüfung)	Felsőfok
B2	FCE (First Certificate in English)	ZDfB (Zertifikat Deutsch für den Beruf)	Középfok
B1	PET (Preliminary English Test)	ZD (Zertifikat Deutsch)	Alapfok
A2	KET (Key English Test)	Fit in Deutsch 2 ⁵ Start Deutsch 2	
A1		Fit in Deutsch 1 Start Deutsch 1	

A különböző nyelvtudást mérő vizsgákat összekapcsolja a már bemutatott Közös Európai Referenciakeret, amely megteremti az egyes nyelvvizsgáztató központok szintjei-

⁴ Az Akkreditációs Kézikönyv első kiadásában a szintek nem estek egybe a KER szintjeivel.

⁵ A Fit in Deutsch a fiatalok, míg a Start Deutsch a felnőttek számára készített vizsga.

nek átláthatóságát, a nyelvvizsgák kölcsönös elismerését, és lehetővé teszi a minősítési rendszerek közötti összehasonlításokat (Bárdos, 2006; Boócz-Barna, 2004; Boócz-Barna és Majorosi, 2002). Ha a különböző vizsgákat közös nyelvismereti skálán helyezük el (2. táblázat), akkor lehetségesnek kell lennie, hogy a vizsgarendszer egyik vizsgáján elért teljesítmény egy másik vizsgán elért eredményhez viszonyítható legyen (KER, 2002).

A Referenciakeret szintleírása tehát egy olyan külső kritérium, amely alapján egy adott vizsgarendszer validitását is értékelhetjük. Ezt nevezzük külső validitásnak (Dávid, 2005; Glaboniat és Müller, 2006), amit úgy határozhatunk meg, hogy a vizsga valóban azt a szintet méri-e, amit mérni szándékozik. A továbbiakban a külső validitás kérdését tárgyaljuk, és bemutatjuk az ennek eléréséhez szükséges folyamat fázisait.⁶

A Referenciakeret nem ad módszertani ajánlásokat arra vonatkozóan, hogy egy vizsga hogyan tudja szintjeit a referenciaszintekhez validálni. Ezt tükrözi Alderson mára már klasszikussá vált kérdése: „How do I know that my Level B1 is your Level B1?”. A Referenciakeretet kiegészítve 2003-ban adták ki a *Relating Examinations to the Common European Framework of Reference for Languages* (North és mtsai, 2003) c. kézikönyv kísérleti verzióját, amely 2005-től magyar nyelven is hozzáférhető (Barabás, 2005). A Kézikönyvben megadott illesztő eljárás négy szakaszból áll. (1) Az ismerkedési szakaszban a Referenciakeretet alaposan meg kell ismerni, és a szintek leírásait megbízhatóan elsajátítani. (2) A specifikációs szakaszban a vizsgaközpont tartalmi elemzést készít, amelyben már az esetlegesen meglévő nyelvvizsga-leírást kell a KER rendszerével összevetni, majd a specifikáció során felülvizsgálni, átdolgozni és igazítani. (3) A külső validálás eljárásának standardizációs szakaszában a feladatírók, vizsgáztatók KER-re vonatkozó „megítéléseinek” standardizálása a fő feladat, illetve az eddigi vizsgafeladatok és értékelési rendszer standardításának felülvizsgálata a KER eddig kalibrált példái alapján. Itt már megindul az új folyamatok tervezése, beiktatása a korábbi rendszer elemei közé a vizsgafejlesztők és feladatírók bevonásával. Elindul a feladatkészítési folyamat, amely a KER által definiált folyamatszerű követelményeknek megfelelően történik. (4) Az empirikus validálás szakaszában történik meg a feladatok kipróbálása, javítása előteszt szervezésével és lebonyolításával. A szakasz célja az értékelési rendszer standardizáltságának és megbízhatóságának, és a feladatok nehézségi szintjének meghatározása. Ezután javaslatok megfogalmazása történik a szintek nehézségének és a feladattípusok működésének tanulságai alapján az esetleges, a vizsgaleírást is érintő korrekcióira, változtatásaira (az egyes fázisokról lásd részletesebben: Dávid, 2005; Figueras, 2006; Martyniuk, 2006; Zeidler, 2006).

A nyelvvizsgaszintek pontos illesztése a Közös Európai Referenciakeret szintjeihez még sehol sem történt meg. A probléma többrétű: egyrészt nincsenek rutinszerűen és olcsón alkalmazható eljárások, másrészt pedig még nincsenek referencia feladatbankok (Einhorn és Major, 2006). Ezt a hiányt pótolja majd a 2004. októberben kezdődött és 2007. szeptemberig tartó *EBAFLS projekt* (*Building a European Bank of Anchor Items for Foreign Language Skills*), amelyet a Cito szervezet koordinál és a feladatbank létre-

⁶ A probléma aktualitását mutatja, hogy a folyamatot több konferencián is bemutatják. Lásd pl. *INTO EUROPE. European Standards in Language Assessment*, Budapest, 2006. február 9-10, és *34. Jahrestagung des Fachverbandes Deutsch als Fremdsprache*, Themenschwerpunkt 2: *Testen und Prüfen in DaF/DaZ*, Hannover, 2006. június 8-10.

hozásában hét európai ország, köztük Magyarország vesz részt. A projektnek három célja van: (1) angol, francia és német nyelvű olvasott és hallott szöveg értése feladatokhoz feladatbank létrehozása, (2) a feladatbank alkalmazásához és funkciójának leírásához kézikönyv megírása, (3) a mérőeszközök validitásának és reliabilitásának vizsgálata, illetve olyan alapelvek megfogalmazása, amelyek a feladatbank kibővítésére és új feladatbank létrehozására szolgálnak.

A nyelvvizsgaszintek validálása fontos kutatási feladatokat határoz meg, mivel egyrészt meg kell állapítani, hogy egy vizsgarendszer adott szintje hogyan illeszkedik a Referenciakeret szintjéhez, másrészt pedig, hogy a különböző vizsgák tesztfadatai, értékelési eljárásai mennyiben felelnek meg a Referenciakeret által meghatározott nyelvtudás szintjének. E probléma megoldásában a kommunikatív tesztelés támaszkodik a modern (valószínűségi) tesztelméletre (*Item Response Theory [IRT]*), amelynek leírásáról, modelljeinek bemutatásáról lásd pl.: Horváth, 1991, 1997 könyvét, Csapó, 2004 és Molnár, 2006, 2007 tanulmányait. A kommunikatív tesztelésben a Rasch-modelleket a DIALANG (*Diagnostic Language Assessment System for Learners*) skálák szintezésénél (KER, 2002. „C” függelék; Figueras, 2005; Perlmann-Balme, 2006), és az ALTE (1998) „Mit tud?”-állításainak a Referenciakeret szintleírásával történő megfeleltetése során már alkalmazták (KER, 2002. „D” függelék), és az egyes nyelvi feladatok szintjeinek meghatározásában több hazai empirikus kutatás eredményeit is ismerjük (pl. Alderson, 2000; Dávid, 2005, 2006; Nikolov, Pércsich és Szabó, 2000).

A tanulmány zárásaként meg kell említenünk, hogy a magyar közoktatási rendszerben az idegen nyelvi mérés és értékelés területén jelentős fejlődés ment végbe a 2005-ben bevezetett érettségi vizsgával. A vizsgafejlesztés folyamatának munkálatai során (lásd Alderson, Nagy és Óveges, 2000; Einhorn, 2004) az idegen nyelvi érettségit úgy alakították ki, hogy mérési és értékelési eljárásaiban kövesse a kommunikatív tesztelés alapelveit, valamint illeszkedjen a Közös Európai Referenciakeret szintleírásaihoz. Így e vizsga működtetésekor is felmerül és fontos a külső és belső validitás biztosítása. Az idegen nyelvi érettségi azonban több ponton eltér a nyelvtudást mérő vizsgáktól, hiszen elsődleges szempontja a magyar közoktatási rendszerbe való illeszkedése, így funkcióját tekintve, alapvetően a középfokú oktatást lezáró vizsga. E mellett még nyelvtudásszintet mérő vizsga is, hiszen a továbbtanuláshoz a szelekció e vizsga eredménye alapján történik. A vizsga működtetésének sok eleme, mint például a feladatok készítése, a vizsga lebonyolítása, értékelése azonban kevés minőségbiztosítási elemet tartalmaz. Ezen kívül még nem tisztázott, hogy a vizsga két szintje egymáshoz, illetve a Referenciakeret szintjeihez hogyan illeszkedik, így a vizsgaszintek tekintetében jelentős fejlesztő kutatásokra van szükség. Egyrészt vizsgálni kell az idegen nyelvi érettségi vizsgák eredményeit, hogy az érettségizők reális tudásszintjéhez igazíthassák a vizsgát, másrészt pedig azt, hogy a feladatok szintje hogyan viszonyul a követelményekben rögzített készségszintekhez (Einhorn, 2006b). Az idegen nyelvi érettségi vizsgát a nyelvvizsgáktól még az is alapvetően megkülönbözteti, hogy jelentős hatással van a közoktatásra, mivel iskolai követelménnyé teszi a Referenciakeret adott szintjének elérését (Fazekas, 2006), amely megjelenik az 1995-ben bevezetett és 2003-ban módosított Nemzeti Alaptantervben (Petmeki, 2002, 2006). Így kutatást igényel annak feltárása, hogy az idegen nyelvi érettséginek milyen a visszahatása a közoktatásra, azaz hogyan szolgálja az idegen nyelvek

tanításában bekövetkezett szemléleti és tartalmi modernizációt, és milyen mértékben játszik szerepet abban, hogy a kommunikatív tesztelés ne csak a minősítő értékelésben, hanem a mindennapi tanítási-tanulási folyamatban is elterjedjen.

Összegzés

A tanulmány kísérletet kívánt tenni arra, hogy bemutassa a kommunikatív tesztelés elméleti alapjait, általános kérdéseit és dilemmáit.

Az elméleti és gyakorlati kérdésekkel foglalkozó hazai és nemzetközi szakirodalom áttekintése után megállapíthatjuk, hogy a kommunikatív tesztelés mára igen gazdag szakterületté vált, amely a pedagógiai értékelés sajátos jellegű önálló területe. A kommunikatív tesztelés a klasszikus tesztelméletre építve határozza meg a kommunikatív tesztek használhatósági kritériumait, funkcióit és típusait. Ezen kívül sok az áthatás a normaorientált és kritériumorientált tesztelés elméleti fogalmaival, gyakorlati alkalmazásában pedig támaszkodik a modern tesztelméleti (IRT) eszközökre. Ezeket azonban nemcsak átveszi, hanem a kommunikatív kompetencia mérését és értékelését jellemző problémák, dilemmák miatt a tesztelmélet egészére vonatkozó mérésmethodikai fogalmakat, mérési és értékelési eljárásokat specifikus tartalmakkal tölti fel, kialakítja saját fogalomrendszerét, újabb sajátos fogalmakat, terminológiát használ. Gazdagítja a klasszikus tesztelmélet jóságmutatóit, kiemeli a validitás fontosságát, és újabb típusait, aspektusait is meghatározza. Ezek közé tartozik az autentikusság, az interaktivitás és a teszthatások, amelyek bár statisztikai úton nem interpretálhatók, de a kommunikatív tesztelésben kiemelten fontosak, és ezeknek a kritériumoknak a biztosítása a teszt érvényességének és a megbízhatóságának az előfeltételeként fogható fel. Ebből következik, hogy a kommunikatív tesztelésben sok publikáció foglalkozik a használhatósági kritériumok ezen összetevőinek a vizsgálatával. A kommunikatív tesztek típusainak alkalmazásánál is újabb dilemmákkal találkozunk, amelyek a kompetencia és performancia tesztek, és az ezzel szorosan összefüggő közvetlen és közvetett tesztek jellemzik. Ezek a tesztek lényegében egy-egy skála két végpontját jelenítik meg, és a tesztalkalmazóknak a feladata, hogy ezeknek a teszteknek valamilyen egyensúlyát megtalálják.

A nyelvtudást mérő vizsgák alkalmazzák a kommunikatív tesztelési eljárásokat, és ezt a folyamatot a Közös Európai Referenciakeret szintrendszere is elősegíti, amely esetén fontos kérdés, hogy miként tudják a vizsgafejlesztők a tesztfeladataikat, értékelési eljárásukat a Referenciakeret szintjeihez igazítani. Mivel a nyelvtudás igazolása elengedhetetlenül fontossá vált, így egy vizsgarendszer működtetésekor fontosak az alapos empirikus fejlesztő kutatások, amelyek hozzájárulnak ahhoz, hogy a vizsgázók nyelvtudásáról pontos adatokat szerezzünk, és ezek alapján megalapozott döntéseket tudjunk hozni.

Irodalom

- Albers, H. G. és Bolton, S. (1995): *Testen und Prüfen in der Grundstufe. Einstufungstests und Sprachstandsprüfungen*. Fernstudieneinheit 7. Langenscheidt, Berlin.
- Alderson, J. C. (2000): Teljesítménytesztek az angol nyelvi érettségi kipróbálásán. *Magyar Pedagógia*, **100**. 4. sz. 423–558.
- Alderson, J. C. (2004): The shape of things to come: will it be normal distribution? In: Milanovic, M. és Weir, C. J. (szerk.): *Studies in language testing 18: European language testing in a global context: Proceedings of the ALTE Barcelona Conference July 2001*. Cambridge University, Cambridge. 1–27.
- Alderson, J. C. (2006): *Principles and Practice in Language Testing: Compliance or Conflict?* (Az előadás elhangzott az INTO EUROPE. European Standards in Language Assessment konferencián, 2006. február 9-én, Budapesten.)
- Alderson, J. C., Clapham, C., és Wall, D. (1995): *Language Test Construction and Evaluation*. Cambridge University Press, Cambridge.
- Alderson, J. C. és mtsai (2004): *The Development of Specifications for Item Development and Classification within The Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Reading and Listening. Final Report of The Dutch CEF Construct Project*. Lancaster University, Lancaster. (www.ling.lancs.ac.uk/cefgrid, 2006. május 6.)
- Alderson, J. C., Nagy, E. és Öveges, E. (2000): *English Language Education in Hungary. Part II. Examining Hungarian Learner's Achievements in English*. The British Council Hungary, Budapest.
- ALTE Handbook of European Language Examinations and Examination Systems* (1998): Association of Language Testers in Europe, Cambridge (England).
- Altmayer, C. (2004): Sprachkultur und Mehrsprachigkeit: Neuerscheinungen zur europäischen Sprachenpolitik. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, **9**. 2. sz. 1–10. (<http://zif.spz.tu-darmstadt.de/jg-09-2/beitrag/Sprachenpolitik2.htm>, 2006. szeptember 16.)
- Bachman, L. F. (1990): *Fundamental Considerations in Language Testing*. Oxford University Press, Oxford.
- Bachman, L. F. és Palmer, A. (1996): *Language Testing in Practice*. Oxford University Press, Oxford.
- Bachman, L. F. és Palmer, A. (2005): *Language Assessment in Practice*. Oxford University Press, Oxford.
- Barabás László (2004, szerk.): *Akkreditációs Kézikönyv*. Nyelvizsgát Akkreditáló Testület, PH Nyelvizsgáztatási Akkreditációs Központ, Budapest.
- Barabás László (2005, szerk.): *Nyelvizsgák szintillesztése a Közös Európai Referenciakerethez. Kézikönyv*. Előkészítő, kísérleti változat. Nyelvizsgát Akkreditáló Testület, PH Nyelvizsgáztatási Akkreditációs Központ, Budapest.
- Bárdos Jenő (1984): Az idegen nyelvek tanítása a nyolcvanas években. *Pedagógiai Szemle*, **34**. 2. sz. 105–118.
- Bárdos Jenő (2001): Az idegen nyelvi mérés és értékelés fejlődése. In: Csapó Benő és Vidákovich Tibor (szerk.): *Neveléstudomány az ezredfordulón. Tanulmányok Nagy József tiszteletére*. Nemzeti Tankönyvkiadó, Budapest. 254–267.
- Bárdos Jenő (2002a): *Az idegen nyelvi mérés és értékelés elmélete és gyakorlata*. Nemzeti Tankönyvkiadó, Budapest.
- Bárdos Jenő (2002b): Kulturális kompetencia az idegen nyelvek tanításában. *Modern Nyelvtanítás*, **8**. 1. sz. 5–18.
- Bárdos Jenő (2003): A nyelvtudás megítélésének korlátai. *Iskolakultúra*, **13**. 8. sz. 28–39.
- Bárdos Jenő (2004): *Nyelvpedagógiai tanulmányok*. Iskolakultúra könyvek 24. Pécs.
- Bárdos Jenő (2005): A magyarországi nyelvizsgák és az európai mobilitás. *TEE Szemle*, 3–12.

A kommunikatív tesztelés elméleti alapjai

- Bárdos Jenő (2006): A nyelvtudás-fogalom metamorfózisai – kritikai elemzés. *PORTA LINGUA – 2006 Utak és perspektívák a hazai szaknyelvoktatásban és -kutatásban*, Debrecen. 15–22.
- Bolton, S. (1996): *Probleme der Leistungsmessung. Lernfortschrittstests in der Grundstufe*. Fernstudieneinheit 10. Langenscheidt, Berlin.
- Bond, L. A. (1996): Norm- and Criterion-referenced testing. *Practical Assessment, Research and Evaluation*, 5. 2. sz. (<http://pareonline.net/getvn.asp?v=5&n=2>, 2006. május 8.)
- Boócz-Barna Katalin (2004): Az Európa Tanács dokumentumai a nyelvoktatásban In: Einhorn Ágnes (szerk.): *Kézikönyv a német érettségi vizsga reformjához*. Országos Közoktatási Intézet, Budapest. 283–297.
- Boócz-Barna, K. és Majorosi, A. (2002): Ein Dokument für die Schublade oder doch etwas mehr? Gemeinsamer Europäischer Referenzrahmen für Sprachen: lehren, lernen, beurteilen. *Deutschunterricht für Ungarn*, 17. 1–2. sz. 21–32.
- Brualdi, A. (1999): *Traditional and Modern Concepts of Validity*. ERIC Document, ED435714. ERIC Clearinghouse on Assessment and Evaluation, Washington DC.
- Canale, M. (1983): On some dimensions of language proficiency. In: Oller, J. W. Jr.: *Issues in Language Testing Research*. Newbury House, Rowley, Mass. 333–342.
- Canale, M. és Swain, M. (1980): Theoretical bases of communicative approaches to language learning and testing. *Applied Linguistics*, 1. sz. 1–47.
- Chomsky, N. (1965): *Aspects of the theory of syntax*. Cambridge, Mass.: Massachusetts Institute of Technology (MIT) Press.
- Csapó Benő (1987): A kritériumorientált értékelés. *Magyar Pedagógia*, 87. 3. sz. 247–266.
- Csapó Benő (1996): Standard érettségi: lehetőségek és dilemmák. *Új Pedagógiai Szemle*, 9. sz. 17–26.
- Csapó Benő (2004): Tudásszintmérő tesztek. In: Falus Iván (szerk.): *Bevezetés a pedagógiai kutatás módszereibe*. Műszaki Könyvkiadó, Budapest. 277–316.
- Dávid Gergely (1992): Objektivitás, értékelés standardizáció. *Iskolakultúra*, 2. 1. sz. 26–29.
- Dávid Gergely (2002): A nyelvvizsgák akkreditációjának tanulságai. In: Kárpáti Eszter és Szűcs Tibor (szerk.): *Nyelvpedagógia*. Iskolakultúra könyvek 12. Pécs. 159–165.
- Dávid Gergely (2005): Nyelvvizsgaszintek validálása: lehetőségek és korlátok. *PORTA LINGUA – 2005 Szakmai nyelvtudás - szaknyelvi kommunikáció cikkek, tanulmányok a hazai szaknyelvoktatásról és -kutatásról*, Debrecen. 279–295.
- Dávid Gergely (2006): *Az emelt szintű idegen nyelvi érettségi és az államilag elismert nyelvvizsgák a vizsgázói teljesítmények tükrében*. Összegző tanulmány. Kézirat. Nyelvvizsgát Akkreditáló Testület, PH Nyelvvizsgáztatási Akkreditációs Központ, Budapest.
- Davies, A. és Mtsai (1999): *Dictionary of language testing* (Studies in Language Testing 7). Cambridge University Press, Cambridge.
- Dévény Ágnes (2003): Az idegen nyelvi tesztelés etikai problémái. *Új Pedagógiai Szemle*, 7–8. sz. 134–145.
- DIALANG Diagnostic Language Assessment System for Learners (www.dialang.com)
- EBAFLS Building a European Bank of Anchor Items for Foreign Language Skills (www.ebafls.cito.com)
- Einhorn Ágnes (1998, szerk.): Vizsgáztatási hagyományok Magyarországon. In: Einhorn Ágnes (szerk.): *Vizsgatárgyak, vizsgamodellék I. Német nyelv*. Országos Közoktatási Intézet Értékelési és Érettségi Vizsgaközpont, Budapest. 97–191.
- Einhorn Ágnes (1999): Nyelvtudás és vizsga. *Educatio*, 3. sz. 543–556.
- Einhorn Ágnes (2004, szerk.): *Kézikönyv a német érettségi reformjához*. Országos Közoktatási Intézet, Budapest.
- Einhorn Ágnes (2006a): A vizsgafeladat fejlesztésének folyamata és kritériumai. *Új Pedagógiai Szemle*, 1. sz. 67–74.

- Einhorn Ágnes (2006b): *Az idegen nyelvi érettségi reformja*. Kézirat. Országos Közoktatási Intézet, Budapest.
- Einhorn Ágnes és Major Éva (2006): Az idegen nyelvek – vizsgafejlesztés nemzetközi kontextusban. In: Horváth Zsuzsanna és Lukács Judit (szerk.): *Új érettségi Magyarországon. Honnan, hová, hogyan? Egy folyamat állomásai*. Országos Közoktatási Intézet, Budapest. 127–137.
- Ennis, R. H. (1999): Test Reliability: A Practical Exemplification of Ordinary Language Philosophy. *Philosophy of Education Society*. University of Illinois, Urbana-Champaign. (http://www.ed.uiuc.edu/eps/PES-Yearbook/1999/ennis_body.asp, 2006. augusztus 19.)
- Fazekas Márta (2006): Standardok a nyelvi érettségiben: régi-új érettségek és nyelvvizsgák. In: Kósa Barbara és Simon Mária (szerk.): *Új vizsga – új tudás? Az új érettségi hatása az iskolakezdéstől a záróvizsgáig*. Országos Közoktatási Intézet, Budapest.
- Figueras, N. (2005): Testing, testing, everywhere, and not a while to think. *ELT Journal*, **59**. 1. sz. 47–54.
- Figueras, N. (2006): *Linking exams to the CEF levels: The Manual and the Catalan experience*. (Az előadás elhangzott az INTO EUROPE. European Standards in Language Assessment konferencián, 2006. február 9-én, Budapesten.)
- Glaboniat, M. és Müller, M. (2006): Note „Sehr gut!“ – Aber in Bezug worauf? Referenzrahmen und Profile Deutsch in ihren Auswirkungen auf Prüfungen und Tests. *Fremdsprache Deutsch. Zeitschrift für die Praxis des Deutschunterrichts*, **34**. sz. 14–21.
- Glaboniat, M. és mtsai (2002): *Profile deutsch. Gemeinsamer europäischer Referenzrahmen*. Langenscheidt, Berlin stb.
- Glaboniat, M. és mtsai (2005): *Profile Deutsch A1-C2 (Version 2.0)* Langenscheidt, Berlin stb.
- Gronlund, N. E. (1988): *How to construct achievement tests*. Englewood Cliffs, NF: Prentice Hall.
- Grotjahn, R. (2000a): Testtheorie: Grundzüge und Anwendungen in der Praxis. In: Wolff A. és Tanzer H. (szerk.): *Sprache – Kultur – Politik. Beiträge der 27. Jahrestagung DaF 1999*. Materialien Deutsch als Fremdsprache, FaDaF, Regensburg. 304–341.
- Grotjahn, R. (2000b): Determinanten der Schwierigkeit von Leseverstehensaufgaben: Theoretische Grundlagen und Konsequenzen für die Entwicklung des TestDaF. In: Bolton, S. (szerk.): *TestDaF: Grundlagen für die Entwicklung eines neuen Sprachtests Beiträge aus einem Expertenseminar*. Goethe-Institut, München; Gilde Verlag, Köln. 7–55.
- Grotjahn, R. (2002): *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*. AKS Kiadó, Bochum.
- Grotjahn, R. (2006): *Testen und Prüfen. Ein Überblick*. (Az előadás elhangzott a 34. Jahrestagung des Fachverbandes Deutsch als Fremdsprache, Themenschwerpunkt 2: Testen und Prüfen in DaF/DaZ, konferencián, 2006. június 8-án, Hannoverben, írott változata megjelenik: Wolff, A.; Hunstiger, A.; Koreik, U. (szerk.): *Chance Deutsch: Schule - Studium – Arbeitswelt. Beiträge der 34. Jahrestagung DaF 2006*. Materialien Deutsch als Fremdsprache, FaDaF, Regensburg.)
- Horváth György (1991): *Az értelem mérése*. Tankönyvkiadó, Budapest.
- Horváth György (1997): *A modern tesztmodellek alkalmazása*. Akadémiai Kiadó, Budapest.
- Horváth Zsuzsanna és Lukács Judit (2006): A megvalósult vizsga. Eredmények és iskolai hatások. *Új Pedagógiai Szemle*, **9**. sz. 26–47.
- Hymes, D. H. (1972): On communicative competence. In Pride, J. B. és Holmes, J. (szerk.): *Sociolinguistics: Selected readings*. Penguin, Harmondsworth. 269–293.
- Katona Lucia (1995): A nyelvtudás fogalmának új értelmezése. *Iskolakultúra*, **5**. 1–2. sz. 67–75.
- Katona Lucia és Dörnyei Zoltán (1991): Az idegen nyelvtudás mérésének új módja a C-teszt. *Iskolakultúra*, **1**. 4. sz. 42–51.
- Közös Európai Referenciakeret: nyelvtanulás, nyelvtanítás, értékelés. KER (2002) OM – PTMIK, Budapest – Pilisborosjenő.*

A kommunikatív tesztelés elméleti alapjai

- Krekeler, C. (2005): *Grammatik und Fachbezug in Sprachtests für den Hochschulzugang*. Dissertation beim Fachbereich für Geisteswissenschaften der Universität Duisburg-Essen.
- Krumm, J.-H. (2006): Müssen jetzt alle dasselbe können? Vor- und Nachteile der Globalisierungsprozesse im Sprachunterricht. *Fremdsprache Deutsch. Zeitschrift für die Praxis des Deutschunterrichts*, 34. sz. 30–33.
- Lado, R. (1961): *Language testing. The construction and use of foreign language tests*. Longman, London.
- Lienert, G. A. és Raatz, U. (1994): *Testaufbau und Testanalyse*. Beltz, Psychologie Verlags Union, Weinheim.
- Malone, M. (2000): *Simulated Oral Proficiency Interviews: Recent Developments*. ERIC Document ED447729. ERIC Clearinghouse on Assessment and Evaluation, Washington DC.
- Martyniuk, W. (2006): *Relating language examinations to the Common European Framework of Reference for Languages (CEFR)*. (Az előadás elhangzott az INTO EUROPE. European Standards in Language Assessment konferencián, 2006. február 10-én, Budapesten.)
- McNamara, T. (2000): Testing, testing. What is a language test? In: McNamara, T.: *Language Testing*. Oxford University Press, Oxford.
- Messick, S. (1996): Validity and washback in language testing. *Language Testing*, 13. 3. sz. 241–256.
- Miyata-Boddy, N. és Langham, C. S. (2000): *Communicative language testing - an attainable goal?* The British Council, Tokyo. 75–82.
- Molnár Gyöngyvér (2006): A Rasch-modell alkalmazása a társadalomtudományi kutatásokban. *Iskolakultúra*, 16. 12. sz. 99–113.
- Molnár Gyöngyvér (2007): A Rasch modell kiterjesztése nem dichotóm adatok elemzése: a rangskálás és parciális kredit modell. *Pszichológiai Szemle*, Benyújtás alatt álló kézirat.
- Morrow, K. (1981): Communicative Language Testing – Revolution or Evolution. In: Alderson, J. C. és Huges A.: *111-Issues in Language Testing*. (ELT documents). The British Council. 9–25.
- Neuner, G. és Hunfeld, H. (1993): *Methoden des Fremdsprachlichen Deutschunterrichts. Eine Einführung*. Fernstudieneinheit 4. Langenscheidt, Berlin stb.
- Nikolov Marianne, Pércsich Richárd és Szabó Gábor (2000): A puding próbája: Alapszintű angol feladatsorok bemérésének tapasztalatai. *Modern Nyelvoktatás*, 6. 4. sz. 3–28.
- Niska, H. (1997): *Testing community interpreters: a theory, a model and a plea for research*. Stockholm University, Sweden. (<http://lisa.tolk.su.se/00TEST.HTM>, 2006. május 8.)
- North, B. és mtsai (2003): *Relating Examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (CEF). Manual: Preliminary Pilot Version*. Council of Europe, Strasbourg, Language Policy Division.
- Oller, J. W. (1979): *Language tests at school*. Longman, London.
- Perlmann-Balme, M. (2001): Formen und Funktionen von Leistungsmessung und -kontrolle. In: Helbig, G., Götz, L., Henrici, G. és Krumm, H.-J. (szerk.): *Deutsch als Fremdsprache: ein internationales Handbuch*. de Gruyter, Berlin és New York. 994–1006.
- Perlmann-Balme, M. (2006): „Das kann ich schon!“ Kompetenzen testen, prüfen, zertifizieren. *Fremdsprache Deutsch. Zeitschrift für die Praxis des Deutschunterrichts*, 34. sz. 5–13.
- Petneki, K. (2002): Entwicklung des Abiturkonzepts für den Deutschunterricht. *Deutschunterricht für Ungarn*, 17. 1–2. sz. 35–42.
- Petneki, K. (2006): Problematik der Niveaustufen im neuen Abitur für Deutsch als Fremdsprache in Ungarn (Az előadás elhangzott a 34. Jahrestagung des Fachverbandes Deutsch als Fremdsprache, Themenschwerpunkt 2: Testen und Prüfen in DaF/DaZ, konferencián, 2006. június 9-én, Hannoverben, írott változata megjelenik: Wolff, A.; Hunstiger, A.; Koreik, U. (szerk.): *Chance Deutsch: Schule - Studium - Arbeitswelt. Beiträge der 34. Jahrestagung DaF 2006*. Materialien Deutsch als Fremdsprache, FaDaF, Regensburg.)
- Prodromou, L. (1995): The backwash effect: from testing to teaching. *ELT Journal*, 49. 1. sz. 13–25.

- Quetz, J. (2001): Der Gemeinsame Europäische Referenzrahmen. *Info DaF*, **28**. 6. sz. 553–563.
- Quetz, J. (2003): A1 – A2 – B1 – B2 – C1 – C2. Der Gemeinsame Europäische Referenzrahmen. *Deutsch als Fremdsprache. Zeitschrift zur Theorie und Praxis des Deutschunterrichts für Ausländer*, 1. sz. 42–48.
- Robinson, P. (1996): Task based testing, performance-referencing and ESL program development. In: Robinson P. (szerk): *Task Complexity and Second Language Syllabus Design: Data-based Studies and Speculations*. University of Queensland Working Papers in Language and Linguistics (Special Issue). CLTR, Brisbane. 95–116.
- Seiffert, C. (2003): Was ist und was will „Profile deutsch?“ *Deutsch als Fremdsprache. Zeitschrift zur Theorie und Praxis des Deutschunterrichts für Ausländer*, 4. sz. 238–240.
- Sigott, G. (1996): Quantifying Language Ability. *Effectiveness of language learning and teaching*, Graz, 42–64.
- Skehan, P. (1998): *A cognitive approach to language learning*. Oxford University Press, Oxford.
- Takala, S. (1998): *Language Testing: Recent Developments and Persistent Dilemmas*. ERIC Document ED460636. ERIC Clearinghouse on Assessment and Evaluation, Washington DC.
- Tschirner, E. (2001): Die Evaluation fremdsprachlicher mündlicher Handlungskompetenz: Ein Problemaufriss. *Fremdsprachen Lehren und Lernen*, 30. sz. 87–115.
- University of Cambridge ESOL Exams. (<http://www.cambridgeesol.org/exams/index.htm>, 2006. augusztus 20.)
- Young, R. F. (2000): *Interactional Competence: Challenges for Validity*. ERIC Documents ED444361. ERIC Clearinghouse on Assessment and Evaluation, Washington DC.
- Zeidler, B. (2006): *Niveaustufen vergleichbar machen*. (Az előadás elhangzott a 34. Jahrestagung des Fachverbandes Deutsch als Fremdsprache, Themenschwerpunkt 2: *Testen und Prüfen in DaF/DaZ*, konferencián, 2006. június 9-én, Hannoverben, írott változata megjelent: Wolff, A.; Hunstiger, A.; Koreik, U. (szerk.): *Chance Deutsch: Schule - Studium – Arbeitswelt. Beiträge der 34. Jahrestagung DaF 2006*. Materialien Deutsch als Fremdsprache, FaDaF, Regensburg.)
- Zertifikate und Prüfungen für Deutsch als Fremdsprache (2006): *Fremdsprache Deutsch. Zeitschrift für die Praxis des Deutschunterrichts*, 34. sz. 45.

ABSTRACT

TIBOR VÍGH: THEORETICAL PRINCIPLES OF COMMUNICATIVE LANGUAGE TESTING

This paper attempts to highlight central themes and issues in communicative language testing theory. It is structured into six parts. Part 1 interprets the terms of language measurement, evaluation and testing. Parts 2 and 3 deal with the question what we have to test, give an overview of the development of models of communicative competence and introduce the levels of the Common European Framework of Reference for Languages (CEFR). A further section of the paper deals with the methods of communicative language testing and discusses the problem of quantifying language ability. Part 4 describes the evaluation criteria of communicative language tests, the notion of test usefulness developed by Bachman and Palmer (1996) with its interlocking elements of reliability, construct validity, authenticity, interactivity, test impact and practicality. Part 5 presents relevant test types, and deals with norm- and criterion-referenced tests, performance and system-referenced tests and with problems of direct and indirect testing. This part focuses on the relationship between the models of communicative language ability, evaluation criteria and test types. Finally, the paper discusses the practice of communicative language testing and describes the procedure of relating language examinations to CEFR levels.

Magyar Pedagógia, **105**. Number 4. 381–407. (2005)

Levelezési cím / Address for correspondence: Vígh Tibor, Szegedi Tudományegyetem, Neveléstudományi Doktori Iskola, H–6722 Szeged, Petőfi S. sgt. 30–34.