

TANKÓ GYULA

# A validálási folyamat érvelésalapú megközelítéseinek áttekintése

## An overview of argument-based assessment validation

The aim of the article is to present an overview of the evolution of validity and of the argument-based approaches to language assessment validation. It is argued that there have been three seminal paradigm shifts in language assessment (the validity types taxonomy elaborated by Cronbach – Meehl 1955; the unitary view of validity proposed by Messick, 1989; and the assessment use argument developed by Bachman – Palmer 2010). The article briefly outlines and contrasts the argument-based validation procedures that make use of Toulmin's (1958) argument model (Mislevy's evidence-centered assessment design, Mislevy – Steinberg – Almond 2003; Kane's interpretation/use argument, Kane 2013; and Bachman & Palmer's assessment use argument published in 2010). It is contended that the last one proposes the most user friendly and straightforward justification procedure, which has real potential to empower not only assessment specialists but also further stakeholders.

*Keywords:* evolution of validity, informal reasoning, argument-based validation procedures, contrastive analysis of justification procedures, empowering assessment specialists

## 1. Bevezetés

Minden tesztfejlesztőnek vagy egy tesztet valamilyen célra felhasználó tesztalkalmazónak, aki lehet vizsgáztató, nyelvtanár, az adatgyűjtéshez egy mérőeszközt használó kutató vagy akár egy nyelvvizsgaközpont, arra kell törekednie, hogy a nyelvtudást nemcsak a lehető leghatékonyabb, hanem egyben a legmegbízhatóbb módon mérje. A tesztfejlesztő vagy tesztalkalmazó sokszor ilyen kérdéseket fogalmaz meg:

- Mit jelent a teszten elért pontszám?
- Miért értelmezhetem úgy a pontszámot, hogy a nyelvtanuló egy a Közös európai referenciakeret által meghatározott szintnek megfelelő nyelvtudással rendelkezik; képes egy képzést idegen nyelven sikeresen elvégezni; el tudja látni egy nyelvígé-nyes munkakör feladatait; egy bizonyos mértékig tanulta meg a nyelvkönyv egyik fejezetének tananyagát; egyes nyelvtani szerkezeteket és szókincset megbízhatóan, míg másokat bizonytalanul vagy egyáltalán nem képes használni?

- Miért dönthetek úgy a tesztpontszám alapján, hogy a nyelvtanuló egy bizonyos nyelvvizsga-bizonyítványt megkaphat; felvételt nyer egy képzésre vagy egy munkakörbe; készen áll, hogy elkezdje a következő fejezet tananyagát, vagy újra át kell néznie még valamit a nyelvkönyv éppen befejezett fejezetében; esetleg tudását egy középfokú nyelvtanfolyamon tudja a leghatékonyabban továbbfejleszteni?

Ezekkel a kérdésekkel a szakember tudatosan vagy sem, az általa végzett mérés vagy használt teszt validitását boncolgatja. A validitás márpedig vitathatatlanul minden mérési folyamat sarokköve.

Felmerül tehát egy sor kérdés: Mi a validitás? Mi a validálás? Hogyan validálható egy teszt a gyakorlatban? Hogyan ellenőrizhető egy teszt validitása? Ragaszkodnunk kell a validitáshoz mint nehezen átlátható elvont fogalomhoz, vagy észszerűbb az egyes jellemzői mentén gondolkodni a tesztfejlesztés és -felhasználás során? Validálni kell egy tesztet, vagy elég igazolni egy meghatározott mérési célra való alkalmazhatóságát? Mivel a validitás nehezen meghatározható, illetve a mérés és értékelés gyakorlatára nehezen lefordítható fogalom, ezért a validálás mint folyamat csak kevesek által érthető eljárásnak, a validitás meglétének vizsgálata pedig megvalósíthatatlan feladatnak tűnhet. Viszont a nyelvi mérési és értékelési elmélet legfrissebb fejleményei egyértelműen afelé mutatnak, hogy ennek nem kell feltétlenül így lennie. Ha a validitás fogalmát nem a hagyományos módon közelítjük meg, akkor az általa leírt tartalom közérthetővé válik, és a validálási folyamat érvelés alapú megközelítésével a validálás menete világos eljárássá szelídül, melyet a vizsgálfejlesztő és felhasználó is meg tud érteni, illetve hatékonyan tud kivitelezni.

A validitás jelentésével és validálási folyamattal kapcsolatos fenti kérdésekre keresve a választ, ennek a tanulmánynak a céljai a következők: (i) a validitás fogalmi evolúciójának és az érvelés alapú validálás eredetének, illetve strukturális felépítésének rövid áttekintése, hogy megértsük, miből fakad a validitás fogalmának komplexitása, illetve hogy honnan származik és mi képezi az érvelésen alapuló validálási eljárást; (ii) a Toulmin-féle érvelési modellre alapozott validálási eljárásmodok bemutatása és rövid kontrasztív elemzése annak érdekében, hogy az olvasó nagy vonalakban megértse azokat, és így össze tudja tudja őket hasonlítani.

A tanulmány e célok mentén először a validitás és validálás fogalmakat tárgyalja, azután leírja az érvelés alapú validálási eljárások kiindulópontját képező érvelési modellt, majd röviden bemutatja és összehasonlítja az érvelési modellen alapuló validálási eljárásokat.

## 2. A validitás fogalmának fejlődése

A validitás fogalmának fejlődése önmagában egy tanulmány lehetne, de most csupán a tanulmány szempontjából fontos mozzanatokat emelem ki. A szakemberek már az ötvenes éveket megelőzően a validitás több típusát különböztették meg (a kritérium-alapú validitás prediktív és megegyező altípusait, valamint a tartalmi validitást), majd Cronbach és Meehl (1955) egy új validitástípust, a szerkezeti validitás fogalmát vezették be, mely a teszteredmény és a teszt által mért mentális konstruktum(ok) közötti kapcsolatot írta le. Ezzel elindítottak egy paradigmaváltást a mérés és értékelés tudományágban.

De hamarosan megkérdőjeleződött a validitás három egyaránt fontos és jól elkülönülő típusa alapján történő meghatározása (Cronbach 1980, 1988, Ebel 1961). A kutatók amellet kezdtek állást foglalni, hogy a validitástípusok egy összetett méréselméleti fogalom szorosan összefüggő attribútumai. Ezt a gondolatmenetet formalizálta Messick (1989), aki az egységes konstruktumvaliditás-fogalom megalkotásával a validitás fogalmának új jelleget adott úgy, hogy a kritériumalapú és a tartalmi validitást a szerkezeti validitás komponenseinek nyilvánította, mivel mindkettő a megszerzett pontszám jelentéstartalmához járul hozzá. Így egy fő validitástípus keletkezett: a szerkezeti validitás, melynek fontos attribútumává vált a teszt megbízhatósága (Chapelle 1999).

Messick (1989) előírta annak a vizsgálatát is, hogy az eredmény értelmezése döntéshozatali szempontból nézve mennyire releváns, milyen következményei vannak, illetve azok mennyire hasznosak. Így az egységes konstruktumvaliditás-fogalom szerkezetéből adódóan a pontszámértelmezés mellett a pontszámalkalmazás és annak következményeinek vizsgálata is a modern validálási folyamat kötelező elemévé vált. A validitás fogalmának Messick-féle újragondolása egy újabb paradigmaváltást eredményezett a mérés és értékelés tudományágban.

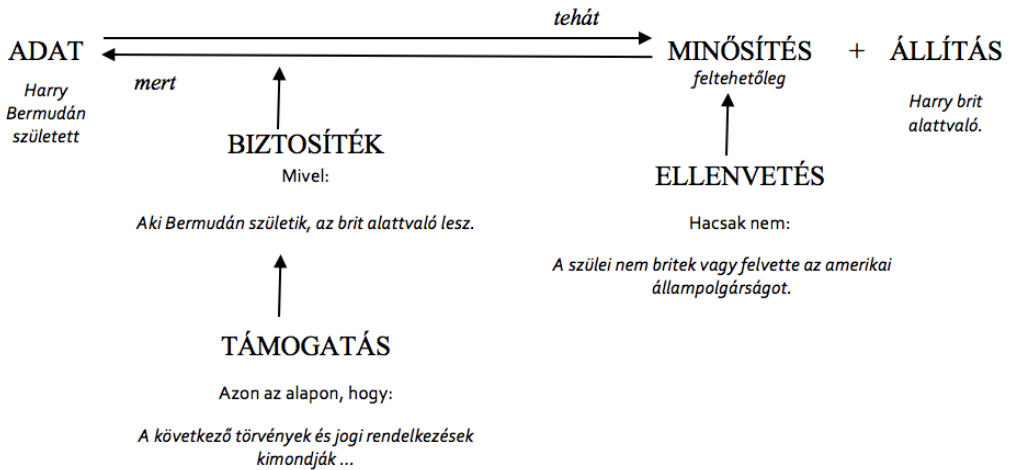
## 3. Az érvelésalapú validálás eredete és szerkezeti alapjai

A felismerés, miszerint a validitást a validitás igazolására felépített érvrendszer határozza meg (Cronbach–Meehl 1955, House 1980, 2014) a validáláselmélet további fontos fejleménye volt. Így Cronbach (1988) a validálás gyakorlatát egy érvelésen alapuló, és a validálási körülmények változása esetében ismételten elvégezendő formatív értékelési folyamatként határozta meg, melynek célja a teszteredmény magyarázattal történő *értelmezése*. Az értelmezés lehet leíró értékelés, előrejelzés vagy akár egy döntési javaslat (Cronbach 1971), és az értelmezés validitási fokának különböző szintjei lehetnek (Cronbach–Meehl 1955), mivel egy érvrendszer meggyőző ereje a begyűjtött bizonyítékok függvényében változik. Ugyan néhányan még sokáig ragaszkodtak hozzá (Alderson–Clapham – Wall 1995, Henning 1987, Hughes 1989, Kontra–Kormos 2007), illetve teszi azt a OH-NyAK (2019) a mai napig (lásd Fogalomtár), a validitás fogalmának újraértelmezése végérvényesen megcáfolta azt a korábbi meghatározást, miszerint

az a valid vizsga, ami azt méri, amit mérni szándékozik, és így a *valid teszt* fogalom elveszítette időszerűségét.

Cronbach (1988) megközelítésével megegyező módon Messick (1995) is az érvelést tartja a validálás alapfolyamatának. Szerinte a validálás egy érvelésen alapuló tudományos vizsgálódás, melynek során különböző jellegű bizonyítékokat sorakoztatunk fel annak megvizsgálása érdekében, hogy a begyűjtött empirikus bizonyítékok és elméleti okfejtések milyen mértékben támasztják alá a tesztponyszám alapján megfogalmazott értelmezéseket és azok felhasználását, illetve az értelmezések és azok alkalmazásának potenciális következményeit.

Az érvelési folyamat részeként Messick (1995) a pontszám alapján megfogalmazott értelmezésekkel összefüggő következtetéseket és a pontszám alkalmazásával, más szóval a pontszám alapján meghozott döntésekkel, illetve az azokból származó cselekvésekkel összefüggő következtetéseket különböztet meg. Szerinte a validálási folyamat során alá kell támasztani minden egyes következtetést, illetve bizonyítani kell, hogy a rendelkezésre álló bizonyítékok a lehetséges alternatív következtetéseket kevésbé támasztják alá. Ezért az érvelési folyamat részének tekinti az igazolást és a cáfolatot is.



**1. ábra. Toulmin (1958) argumentációs modellje kiegészítve az okozat → ok (mert) relációval**

A validálás érvelésen alapuló folyamatként való meghatározása szükségessé tette az érvelés menetének pontos és gyakorlatban alkalmazható leírását. Erre a célra a validálással foglalkozó szakemberek a legalkalmasabbnak Toulmin (1958, 2003) argumentációs modelljét találták. Ez az argumentációs modell az informális logikából ered, és a valós életben előforduló érvszerkezeteket mint döntésekhez vezető racionális folyamatokat írja le, lehetővé téve azok elemzését vagy akár létrehozását. A modell (1. ábra) egy következtetést

eredményező érv séma (mindegyik nyíl egy következtetési műveletet jelöl), melynek hat komponense van. A sémát balról jobbra haladva szokás olvasni (ok  $\rightarrow$  okozat reláció), de ellenkező irányba is működik, például egy állítás bizonyítási háttérének létrehozása vagy feltárása során.

Az alapértelmezett kiindulási pontot a tények vagy *adatok* képezik. Ez a tényállás, amelyet nem vonunk kétségbe, vagy amennyiben igen, akkor az könnyen verifikálható. Egy teszt esetében a nyelvvizsgáló megfigyelt viselkedése minősül adatnak (például a feladatlapon megjelölt válaszok, vagy egy vizsgán megírt panaszlevél). Az adat alapján egy *állítást* fogalmazunk meg, ami nem más, mint egy következtetés. A logikai lépést az adattól az állításig egy *biztosíték* garantálja, ami nem más, mint egy, a mindennapi életben vagy diszciplínaszpecifikusan elfogadott igazoló állítás vagy általánosítás, ami lehet egy törvény, elv vagy szabály (ezt bővebben Toulmin–Rieke–Janik 1984 tárgyalják). Ez a három komponens megegyezik a szillogizmus alkotóelemeivel: az adat a kis premissza, a biztosíték a nagy premissza, és az állítás a következtetés.

A Toulmin-modell további három komponenst határoz meg. Amennyiben szükséges, a biztosítékot alá lehet támasztani egy *támogatással*. Az érvrendszer fontos része továbbá az *ellenvetés*, mely kivételeket, kizáró okokat vagy az állítás elvetésének feltételeit, vagyis azon körülményeket írja le, melyek az állítás erejét korlátozzák vagy az állítást megcáfolják, ezáltal egyben az érvrendszer egészét gyengítve vagy cáfolva. Közvetlenül ehhez kapcsolódó komponens a *minősítés*, mely általában egy megszorító vagy nyomatékosító határozószó vagy határozói szerkezet, mely vagy az adat, biztosíték és támogatás összeadott hatása alapján, vagy az ellenvetés függvényében jelöli az állítás erősségi fokát (Toulmin–Rieke–Janik 1984).

Nem véletlenül esett a validálási folyamatot vizsgáló kutatók választása Toulmin argumentációs modelljére. A modell egy logikusan felépített sémát kínál, ami az érvelési folyamat minden komponensét tartalmazza, és az igazolást a cáfolattal ötvözi. Egyaránt alkalmas a modell a mindennapi életben, illetve a különböző tudományágakban előforduló érvelési folyamatok leírására, elemzésére vagy létrehozására. Teret enged az állításokkal kapcsolatos bizonytalanságnak, ami a validálási folyamat szempontjából nagyon fontos, hiszen bármely teszt a felsorakoztatott validitást alátámasztó bizonyítékok függvényében egy bizonyos fokú validitással rendelkezik (Cronbach–Meehl 1955, House 2014, Messick 1989).

A validitásfogalom fejlődése, és az érvelésalapú validálás eredetének, illetve érvszerkezetének áttekintése azt mutatja, modern megközelítésben a validitás egy összetett és egységes fogalom, melynek a mindent magában foglaló központi komponense a szerkezeti validitás. A validálás pedig a teszteredmények értelmezésének, alkalmazásának és következményeinek indokolására szolgáló logikán és empirikus bizonyítékokon alapuló koherens érvelési folyamat, melynek operacionalizált leírása Toulmin (2003) argumentációs modellje segítségével valósítható meg. A validálási folyamatban

nemcsak a tesztfeljesztő, hanem a tesztet használó minden egyes félnek részt kell vennie (Chapelle 1999). Végezetül, a validálás potenciálisan egy lezáratlan folyamat, ugyanis amennyiben a validálási körülmények változnak, a megváltozott körülményeket figyelembe véve a validálást újra el kell végezni (Cronbach 1988).

Következésképpen, a validitás egy többszörösen összetett fogalom, és a validálás egy bonyolult folyamat, melynek gyakorlati megvalósítása sok kérdést vet fel, és sokat vitatott. A validálási folyamat operacionalizálására több kutató dolgozott ki validálási eljárás módokat. Ezek közül a legismertebbek és legbefolyásosabbak a Mislevy (2003), Kane (2013), illetve Bachman és Palmer (2010) által javasolt eljárás módok<sup>1</sup>.

Ezekben az eljárás módokban több közös vonás felismerhető. Mindegyik célja a pontszámok értelmezésének és alkalmazásának validálása. A validálás érdekében mindegyik egy érvrendszer felépítését írja elő, mely egy érvelési láncból áll, és amelyben minden érv a Toulmin érvelési modell komponenseivel írható le. A retorikából ismert struktúra szerint (van Eemeren–Grootendorst–Snoeck 2002) az érvelési láncban egy érv kimeneti pontja egy másik érv bemeneti pontját képezi. Tehát az érvelési lánc szerkezete és az érvek struktúrája biztosítja a validáláshoz szükséges érvelési keretrendszert, de azt az egyes eljárás módok máshogy és eltérő sikerességgel töltik fel tartalommal a validálási folyamat kivitelezésére javasolt eljárás módok között. A validálási eljárás módok áttekintése azt mutatja, hogy azok egy evolúciós láncot képeznek, ugyanis egymásra épülnek, és egymáshoz képest kifinomultabbá válnak. Annak megítélése érdekében, hogy a validálási eljárás módok közül melyik ültethető át a gyakorlatba a legegyszerűbben, a következő részben bemutatásra kerül a három legbefolyásosabb modell legújabb változata és azok rövid, a gyakorlati alkalmazhatóságuk szempontjából elvégzett elemzése.

## 4. Érvelésalapú validálási eljárás módok

### 4.1. A Mislevy-féle bizonyítékalapú tesztfeljesztési keretrendszer

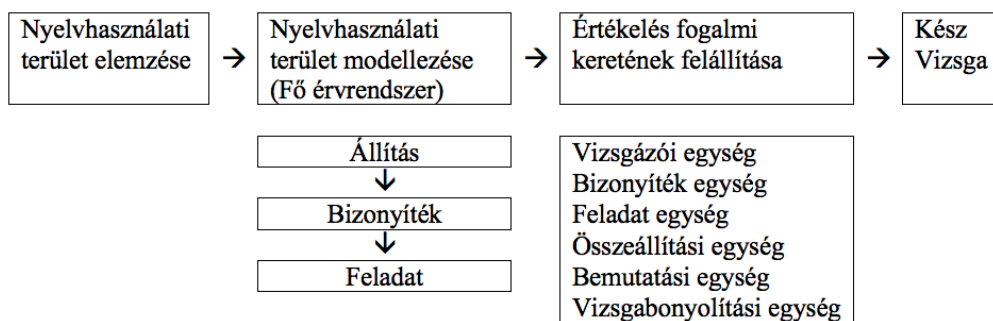
A bizonyítékalapú tesztfeljesztési keretrendszert a kidolgozói (Mislevy 2003, Mislevy–Steinberg–Almond 2002, 2003) tesztek létrehozására és mérőeszközként használt tesztek elemzésére alkalmas módszertani útmutatónak szánták a pedagógiai értékelésben érdekelttek számára<sup>2</sup>. Míg a keretrendszer elsősorban a tesztfeljesztési folyamatot mutatja be, a mérőeszközök vizsgálatának kulcsfontosságú részét képező érvelésen alapuló validálási folyamatokat is körvonalazza. Ugyanis Mislevy (2003) szerint a

<sup>1</sup> Más érvelésalapú validálási eljárás módok is léteznek, de azok nem a Toulmin-féle argumentációs modellel épülnek (lásd pl. Weir 2005, melynek összefoglalását Tankó 2015, validálási tanulmánya tartalmazza).

<sup>2</sup> Fontos megjegyezni, hogy ebben a tanulmányban a Mislevy-féle keretrendszert a nyelvvizsgáztatásban érdekelt felek szempontjából írom le, és ezt az egyes alkotóelemek specifikusabb megnevezése is tükrözheti (pl. az eredeti „terület” összetevő megnevezése ebben a tanulmányban a „nyelvhasználati terület”, mivel a téma a nyelvi mérés, nem a pedagógiai célú mérések összessége).

pedagógiai értékelés nem más, mint egy okfejtés, melynek során pár speciális helyzetben nyújtott teljesítmény alapján arra következtetünk, a vizsgázó milyen ismeretekkel rendelkezik, vagy általában mire képes.

A keretrendszer három egymásra épülő összetevőből áll (2. ábra). A *Nyelvhasználati terület elemzése* fázisban, a célcsoport várható nyelvhasználati területeinek szükségletelemzése történik. Ebben a fázisban gyűjtik össze a tesztfelkészítők a nyelvhasználati területtel kapcsolatos alapvető információkat, melyek forrása lehet például tananyag, kutatási eredmények vagy elméletek.



## 2. ábra. A bizonyítékalapú vizsgafelkészítési keretrendszer felépítése

A második összetevő, a központi szerepet betöltő *Nyelvhasználati terület modellezése*, három tesztfelkészítési produktumból áll, melyek együtt a keretrendszer fő érvrendszerét képező érvelési láncot alkotják. Az első ezek közül az *Állítás*, melynek célja a nyelvvizsgázó nyelvtudásáról a vizsgateljesítmény alapján megfogalmazott kijelentéseknek pontos és érthető leírása (pl. a nyelvvizsgázó több forrásból származó információt és érvelést tud a tudományos próza követelményeinek megfelelő formában szintetizálni.). A második produktum a *Bizonyíték*, melynek célja annak meghatározása, hogy a vizsgázóról megfogalmazott állítások alátámasztásához pontosan milyen jellegű információra van szükség (pl. a fentebbi szemléltető állítás esetében a nyelvvizsgázó által megírt tudományos szövegre). Az utolsó produktumot a *Feladat* képezi, melynek célja azon nyelvhasználati helyzetek meghatározása, melyek révén az előző lépésben leírt vizsgázói tudás meglétét igazoló bizonyítékok megszerzése megvalósulhat (pl. a nyelvvizsgázó meghallgat egy előadásrészletet, majd elolvas egy szöveget, mely az előadás témájához kapcsolódik, és a két forrásból kijegyzetelt tartalom alapján megír egy esszét).

A keretrendszer harmadik összetevője az *Értékelés fogalmi kerete*, ami nem más, mint egy technikai vizsgaleírás, mely tartalmazza például a teszt szerkezetére, tartalmára, szintjére vagy az utasításokra vonatkozó előírásokat. Ezen belül hat meghatározott funkciójú egység (angolul *model/design object*) található. A *Vizsgázoí egység* azt írja

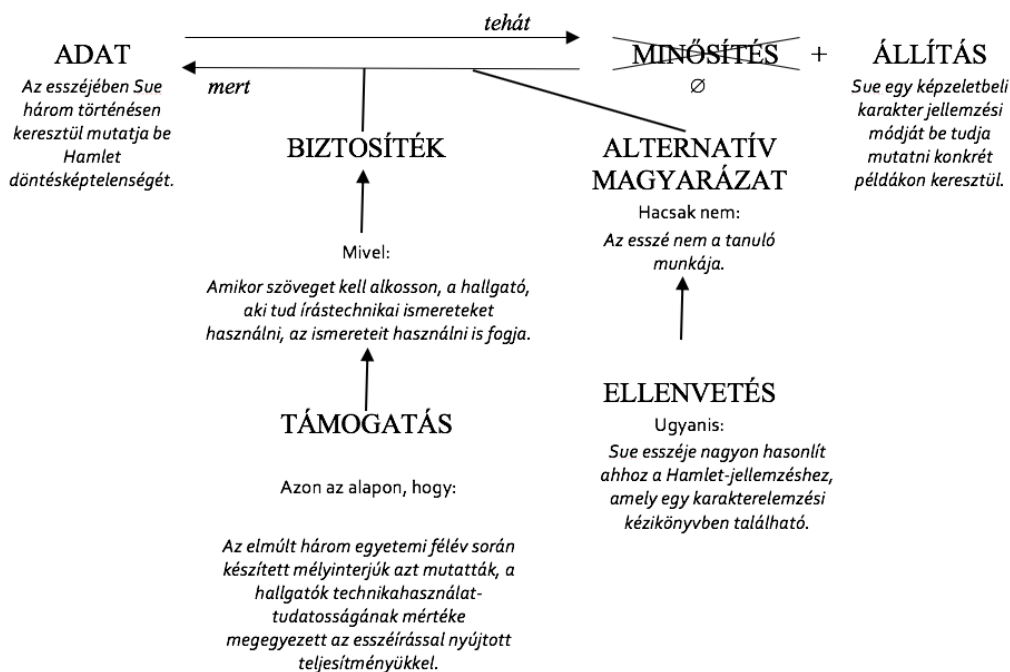


le, a vizsga mit mér. A *Bizonyíték egység* a mentális konstruktumok méréséhez szükséges bizonyítékokat, illetve a pontozási eljárást részletezi. A *Feladat egység* szerepe a bizonyítékok megszerzését lehetővé tevő nyelvhasználati helyzetek meghatározása. Az *Összeállítási egység* azokat a kritériumokat tartalmazza, melyek segítségével a vizsgában szereplő feladatokat kiválasztják (pl. egy készséget hány darab és milyen típusú itemmel kell mérni). A *Bemutatói egység* az itemek és feladatok elrendezését, a vizsgázóval történő interakció jellegét, és a vizsgázói produktumok rögzítési módját szabályozza. És végül a *Vizsgalebonyolítási egység* tartalmazza a vizsga formátumára (pl. számítógépes adaptív vagy hagyományos papíralapú vizsga), és a vizsgabiztonságra vonatkozó információkat, vagy például az egyes feladatokra és vizsgarészekre megszabott időkorlátokat.

A mérőeszköz validitásának vizsgálata szempontjából Mislevy (2003) az *értékelési érvelést* tartja kulcsfontosságúnak, ami a Nyelvhasználatiterület-modellezés komponensátívelő érvelési lánc. Ez a fő érvrendszer a mérőeszközzel elicított teljesítményt, vagyis amit a nyelvvizsgáló teszhelyzetben mond, ír vagy tesz, a nyelvvizsgálóról megfogalmazott állításokkal köti össze. Az állítás a teljesítmény alapján megfogalmazott következtetés (pl. ha a nyelvvizsgáló által alkotott szöveg tartalmazza a forrásanyagokban szereplő tartalom szintézisét, a szöveg logikusan felépített, kohezív és nyelvileg jól megformált, illetve lexikogrammatikai jellemzői és stílusa tekintetében megfelel az „x” nyelvű tudományos próza sajátosságainak, akkor az arra enged következtetni, hogy a nyelvvizsgáló több forrásból származó információt és érvelést tud a tudományos próza követelményeinek megfelelő formában szintetizálni). Az így létrejött állítások alapján a nyelvvizsgálóról döntések születnek (pl. egy bizonyos osztályzatot kap, átmegy vagy megbukik egy vizsgán, felveszik vagy sem egyetemre). Mislevy (2003) szerint egy teszt validitását a teszt fő érvrendszerének validitása határozza meg.

Mislevy (2003) az érvelési lánc validitásának vizsgálatára egy érvszerkezetet leíró modellt javasol, melynek szerkezeti felépítéséhez Toulmin (2003) argumentációs modelljét vette alapul, illetve egészítette ki egy további elemmel a jogi bizonyítási eljárás során felmerülő bizonyítékok elemzésére kidolgozott ábrázolási módszerből (lásd Anderson–Schum–Twining 2005, Wigmore 1913). Mislevy adaptációja két szempontból kifogásolható: elhagyta a Toulmin által javasolt Minősítés elemet, amivel a validitás fokának jelölése szűnt meg. Továbbá az érvszerkezetet leíró modelljének kiegészített komponenseinek megnevezése hibás. A modell leírásában Mislevy (2003) az Ellenvetést alárendeli az Alternatív magyarázatnak (3. ábra), ami nem logikus, ugyanis Toulmin (2003, Toulmin, Rieke és Janik, 1984) értelmezése alapján az Ellenvetés az állítás erejét korlátozó *körülményeket* hivatott leírni. A modellt bemutató példa vizsgálatakor egyértelművé válik, hogy Mislevy Ellenvetés és Alternatív magyarázat elemei hibásan elnevezett, de funkcionális elemek. Az Alternatív magyarázat a következtetést megcáfoló érvrendszer állítása, az Ellenvetés pedig annak az alátámasztása.





### 3. ábra. Az argumentációs modellt bemutató példa kiegészítve a hiányzó komponenssel (Mislevy–Steinberg–Almond 2003: 12)

Az adaptált argumentációs modell állításának erejét szabályzó része azért is kifogásolható, mert Mislevy és társszerzői (2003) az Alternatív magyarázatot alternatív hipotézisnek hívják. Ha tehát az ellenérv egyik típusa egy alternatív hipotézis, akkor az érvrendszer állítása egy nullhipotézis kell legyen. Ez viszont azért értelmetlen, mert a szerzők szerint is az érvrendszer két változó, pontosabban a pontszám és annak az értelmezése közötti összefüggés meglétét hivatott igazolni. Az Ellenvetés Toulmin (2003) szerint sem egy hipotézis, csak az érvrendszeren belüli következtetés erejét gyengítő vagy a következtetést megcáfoló körülmény.

Azt a felvetést, hogy a validálási folyamat részeként nemcsak az állítások alátámasztására érdemes törekedni, hanem azok megcáfolási lehetőségeit is vizsgálni kell, korábban már Cronbach és Meehl (1955) is megfogalmazták. Azért kritizálták a validálási kutatásokat, mert azok kizárólag a megfogalmazott állítások alátámasztására törekedtek, megfelelően az állítás alternatív magyarázatainak megvizsgálásáról és, amennyiben lehetséges, azok falszifikációjáról. Messick (1989) hasonlóképpen azt hangsúlyozta, hogy a megszerzett pontszámon alapuló következtetések validálásának kétféle módon kell megtörténnie, mégpedig annak egyidejű bemutatásával, hogy amíg

egy állítást egymástól független bizonyítékok támasztanak alá, addig a lehetséges alternatív következtetések kevésbé meggyőzően bizonyíthatók. A bizonyítás, illetve cáfolás a validitási folyamat szerves és egyenrangú elemeiként kezelendők, és az alternatív következtetések az Ellenvetés egyik lehetséges típusát képezik, ahogy azt a következő részben tárgyalt érvelésalapú validálási eljárás mód esetében látni is fogjuk.

A bizonyítékalapú vizsgafejlesztési keretrendszer további kétséges pontjai közé sorolható a validitás szemlélete, mely vitathatóan leszűkített. A keretrendszer ugyanis a pedagógiai értékelés azon megközelítésén alapszik, miszerint az értékelés mindössze a vizsgahelyzetben, illetve a valós vizsgahelyzeten kívül történő nyelvhasználat összefüggéseit vizsgálja, ami az előrejelző validitás problémakörébe tartozik, és csupán a validitás egyik szelete. Az általánosíthatóság Messick (1995) egységes konstruktumvaliditás-fogalmában is csak egyik a hat szerkezeti validitást meghatározó aspektus közül.

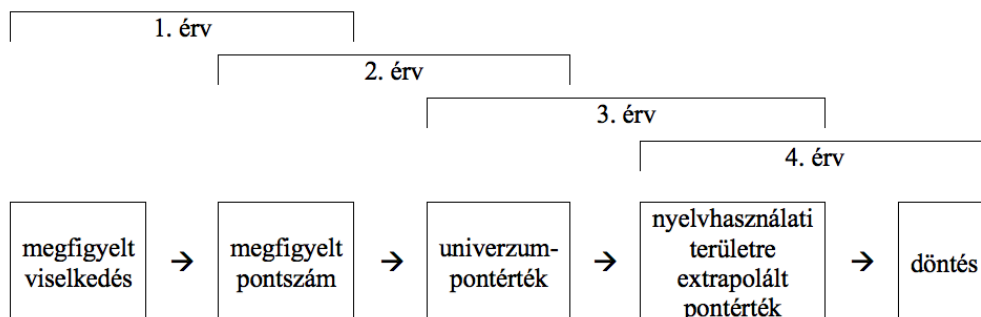
Továbbá, ahogy arra McNamara és Roever (2006) rámutattak, a keretrendszer nem foglalkozik a vizsgapontszámok felhasználásával a vizsgázóról megfogalmazott állításokon és az azokra alapozott döntéseken túlmenően. Márpedig a döntések következményeinek figyelembevételét az egységes konstruktumvaliditás fogalmának meghatározásakor már Messick (1989) is előírta.

Végezetül azt is érdemes kiemelni, hogy ugyan a keretrendszerben felépített érvelések elemzésére javasolt argumentációs modell egy logikus és jól kidolgozott elméleti modell, de az érvstruktúra vizsgálata itemszintű elemzést ír elő (Mislevy 2003), ami a gyakorlatban gazdaságosan nem kivitelezhető. Mislevy az értékelési érvelésre gondosan kidolgozott példákat hoz több elméleti megközelítés alapján (trait-elmélet, behaviorizmus, kognitív pszichológia és szociokulturális elmélet), de azok bonyolultsági foka az argumentációs modell valós életbeli felhasználhatóságát negatívan befolyásolja.

## **4.2. A Kane-féle tesztpontszám értelmezési és alkalmazási érvrendszer**

Kane (2010, 2012, 2013) a pontszámértelmezés és alkalmazás validálási folyamatát egy olyan keretrendszerként írja le, mely segítségével a pontszám alapján megfogalmazott állítások, illetve az állítások alapján meghozott, pl. diagnosztikai vagy válogatási döntések elfogadhatóságát lehet vizsgálni. A keretrendszer alapját a Toulmin-féle argumentációs modellre (2003) épülő érvrendszer képezi. A többször módosított keretrendszerét Kane (1999, 2006) végül a pontszámok értelmezési és alkalmazási érvrendszerének nevezte el (Kane 2013). A keretrendszer négy érvből áll, ugyanis négy állítást tartalmaz, és mindegyik állításhoz önálló alátámasztás tartozik. A négy érv mindegyike kötelező elem, és együtt egy érvláncot képeznek, melyen mind a tesztfejlesztés, mind a validálási folyamat során szigorú sorrendben az elsőtől az utolsóig kell végighaladni. Ha egy állításnak a bizonyítása nem sikerül, a tesztfejlesztés vagy

validálási folyamat az állítást tartalmazó érvnél megszakad. Tehát egy valid teszt mindig egyik állítását alá kell támasztani, és a teszt validitásának erősségi fokát a leggyengébb érv határozza meg.



#### 4. ábra. A pontszámok értelmezési és alkalmazási érvrendszere (Kane 2013)

Az első érv adat összetevője a *megfigyelt viselkedés* és az állítás összetevője a *megfigyelt pontszám* (4. ábra). Az első következtetés, amit a validálás részeként bizonyítani kell az, hogy a megfigyelt tesztpontszám a megfigyelt viselkedés, vagyis a megfigyelt vizsgázói nyelvi teljesítmény, értékeléssel megvalósított leképezése. Az érvelési láncban a második érv adat elemét a megfigyelt pontszám, az állítását pedig az *univerzum-pontérték* képezi. A második bizonyítandó következtetés arra vonatkozik, hogy a nyelvtanuló megfigyelt pontszáma megfelelőképpen reprezentálja az ő univerzum-pontértékét. Az univerzum-pontérték az itemek univerzumából véletlenül vett minta alapján kiszámított várható vizsgázói pontérték (Verhelst 2006). Ez a következtetés tehát akkor elfogadható, ha bizonyítást nyer, hogy a megfigyelt vizsgapontszám nem módosul (i) a vizsgafeladatok vagy értékelők változásával, tehát azok csereszabatosak, illetve (ii) a pontozási szabályokon vagy (iii) a vizsgaalkalmakon keresztül. A harmadik érv adata az univerzum-pontérték és állítása a *nyelvhasználati területre extrapolált pontérték*. A harmadik érvnek azt kell bizonyítania, hogy a nyelvtanuló univerzum-pontértéke megfelelően tükrözi az ő nyelvhasználati területre extrapolált pontértékét, vagyis az ő vizsgahelyzeten kívüli, de a vizsgált készségre vonatkozó nyelvhasználatát. A negyedik érv adatát a nyelvhasználati területre extrapolált pontérték, míg állítását a rajta alapuló *döntés* képezi. Az érv azt bizonyítja, hogy egy adott döntést (például hogy a nyelvtanuló készen áll folytatni tanulmányait egy egyetemen, vagy el tudja látni a munkaköri feladatait egy adott idegen nyelven) azért lehet meghozni, mert az extrapolált pontérték azt a bizonyos döntést indokolja. A négy érv által leírt következtetések ha-akkor szerkezetűek, és mindössze az állítások valószínűségét garantálják, nem minden kétséget kizáróan igazolják azokat.

Kane (2010) megközelítése alapján a pontszámértelmezési és -alkalmazási érvrendszer felállítása a tesztfejlesztéssel párhuzamosan történik. Ezt követi a meghatározott célra, egy bizonyos kontextusra, illetve egy adott populáció számára megfogalmazott pontszámértelmezési és -alkalmazási érvrendszer kritikai értékelése. Ez az érvrendszer-értékelés képezi magát a validálási folyamatot.

Ebben a megközelítésben a validálás egy ismétlődő folyamat, melynek során először a megfogalmazott értelmezési és alkalmazási érvrendszert kell megvizsgálni, majd szükség szerint újrafogalmazni. Ezt követően, az újrafogalmazott érvrendszert ismételtelen meg kell vizsgálni. A folyamat addig ismétlődik, amíg nem marad kétség afelől, hogy a pontszám alapján megfogalmazott pontszámértelmezési, illetve -alkalmazási állítások megfelelően alá vannak támasztva, nincs olyan bizonyíték, amely azokat megcáfolhatja, illetve nincs a pontszámok értelmezésének további elfogadható alternatívája (Kane 2012, 2013). Fontos kiemelni, hogy a pontszámok javasolt értelmezése és alkalmazása idővel módosulhat a változó szükségletek és tudományos magyarázatok fényében, ugyanis azok a validáláshoz szükséges bizonyítékok megváltozását eredményezhetik (Kane 2013).

Az értelmezési és alkalmazási érvrendszer vizsgálata ugyancsak több lépésből áll (Kane 2011, 2013). Először azt kell ellenőrizni, hogy az érvrendszer állításai érthetően vannak-e megfogalmazva. Ezt követi az egyes érvek kritikai értékelése. Az érveket annyira részletesen kell megfogalmazni, hogy a hihetőségük értékelése egy független tesztminősítő számára is lehetségessé váljon. Egyértelműen beazonosíthatóknak kell lenniük mind a következtetéseknek, mind pedig az alapfeltételezéseknek, különösképpen azoknak, melyek megkérdőjelezhetők. Továbbá az alapfeltételezések természetük-nél fogva maguktól értetődők, vagy pedig bizonyítékokkal megfelelően alátámasztott tartalmak lehetnek. Az alapfeltételezéseken alapuló következtetésekkel szemben támasztott követelmény az észszerűség. Nem utolsósorban a teljes érvelési lánc akkor elfogadható, ha logikus felépítésű, és kimerítő részletességgel írja le a pontszámok javasolt értelmezését vagy felhasználási módját.

Kane (2013) pontszám értelmezési és alkalmazási érvrendszerében a Toulmin-féle (2003) érvelési modell *Ellenvetés* elemének a megnevezése *Kivétel*. A megváltoztatott megnevezés viszont ugyanazt a funkciót takarja, mint az eredeti modellben. Kane tehát visszaállítja a Mисlevy (2003) által kevésbé sikeresen újraértelmezett *Ellenvetés* összetevőt az eredeti állapotába. Kane validálási módszere a pontszámok alkalmazásának hangsúlyos fókuszba állítása miatt is túlmutat a Mисlevy-féle megközelítésen. Kane (2011, 2013) rámutat ugyanis arra, hogy mivel az általa javasolt megközelítésben igyekezett Messick (1989) egységes konstruktumvaliditás-fogalmának átfogó jellegét megőrizni, fontosnak tartja a tesztalkalmazás hatásainak vizsgálatát. Ezáltal ebben az újragondolt validálási eljárás módban megjelennek a Messick által is fontosnak tartott szociális és politikai dimenziók.

Ellenben Kane (2013) validálási módszere sem könnyen alkalmazható a gyakorlatban. Nem írja le ugyanis azt, hogy pontosan mi mindent kell bizonyítani a validitás bizonyítási folyamat részeként. Tehát a validitás bizonyításához szükséges bizonyítékok pontos típusa továbbra sincs meghatározva, illetve az a kérdés is nyitott marad, hogy a bizonyítékok mindegyike szükséges-e a validitás bizonyításához.

Továbbá, ahogy az a pontszámok értelmezési és alkalmazási érvrendszerének bemutatása során tapasztalható (4. ábra), az érvelési lánc elemeinek értelmezése is erősen technikai jellegű, és széles körű pszichometriai ismereteket feltételez. Emiatt Kane (2011) állítása ellenére a javasolt validálási eljárás mód csak látszólag praktikus és egyszerű. Ez lehet a magyarázata annak, hogy a széles körű pozitív fogadtatása ellenére ezt a validálási módszert nem sokan használták, és különösen kevesen alkalmazták sok vizsgázóval dolgozó pedagógiai értékelések esetében (Shaw–Crisp 2015).

### **4.3. A Bachman és Palmer-féle tesztalkalmazási érvrendszer**

Az értékelési folyamat igazolása nevű tesztminősítési megközelítést Bachman és Palmer (2010) dolgozták ki, majd Bachman és Damböck (2018) tovább pontosították, illetve az osztálytermi értékelésre átdolgozták. Az „igazolás” megnevezés megfelel Bachman és kollégái által tudatosan kommunikálni kívánt közlésnek. Ahogy később látni fogjuk, Bachman és kollégái célirányosan kerülnek egyes méréselméleti fogalmakat. Az ő megközelítésükben az értékelési folyamat igazolása egy olyan eljárás, melynek segítségével bemutatható a tesztalkalmazás igazolhatóságának mértéke. Segítségével tehát tudományosan vizsgálható, hogy egy teszt alkalmazása mennyire helyénvaló.

Bachman és Palmer (2010) szerint bármely értékelési folyamat igazolásának feltétele egy tesztalkalmazási érvrendszer megfogalmazása, illetve az azt alátámasztó bizonyítékok összegyűjtése. A tesztalkalmazási érvrendszer egy olyan fogalmi keretrendszer, mely a tesztfejlesztés és -alkalmazás módszerét a méréselméleti követelményeknek megfelelően, de ugyanakkor a tesztfejlesztésben és -alkalmazásban érintett minden szakember számára érthető, követhető és a gyakorlatban kivitelezhető módon írja le. Az érintett szakemberek között első alkalommal jelenik meg a pontszámértelmezés történelmében explicit módon a nyelvtanár mint pedagógiai tesztfejlesztő, vagy mint minősítő döntéseket hozó tesztfelhasználó.

A tesztalkalmazási érvrendszer négy állítást tartalmaz (5. ábra), melyek (i) a teszt-eredmények alkalmazásának következményeire, (ii) a következményeket eredményező döntésekre, (iii) a döntéseket alátámasztó értelmezésekre, és (iv) az értelmezések alapját képező eredményjegyzék (*assessment record*) formájában (pl. összpontszám, írásos vagy szóbeli minősítés, osztályzat) történő megjelenítésére vonatkozó kijelentéseket fogalmaznak meg. Az állítások tehát a megfigyelt nyelvvizsgálói teljesítményt a tesztalkalmazás következményeivel kötik össze, és négy egymásra épülő érvből álló érvrendszert képeznek. Amint ebben az érvrendszerben az egyik érv állítása igazolást

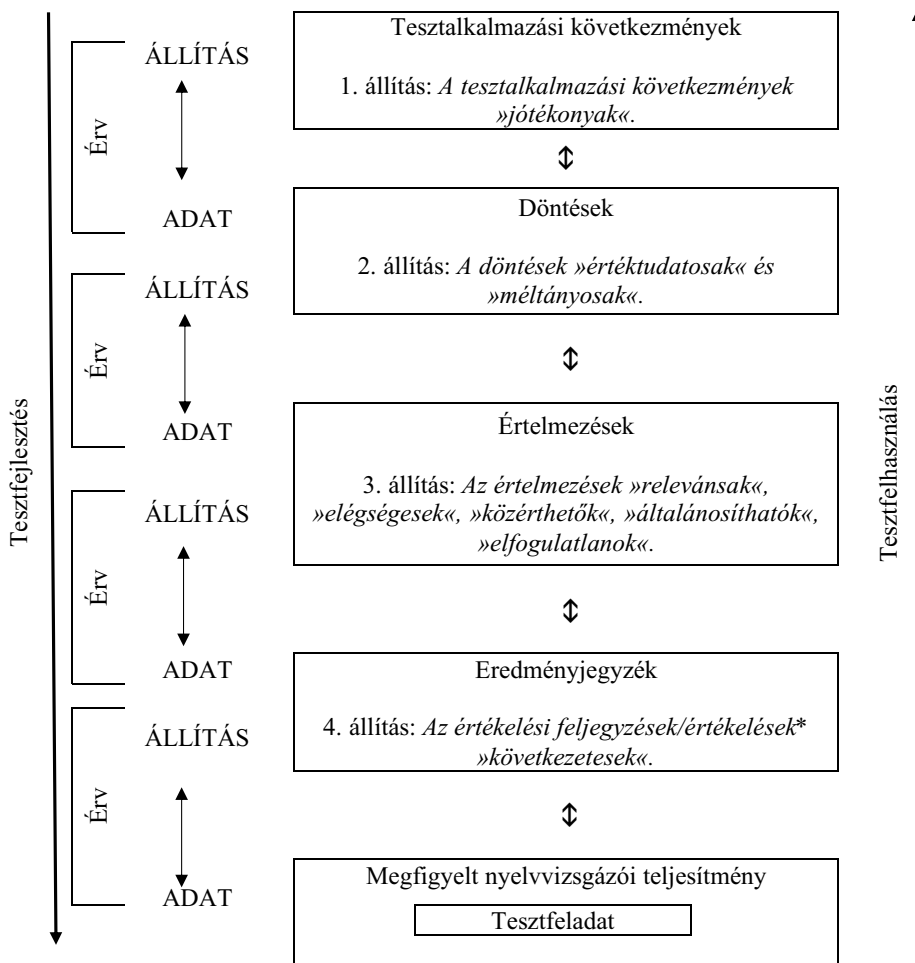
nyer, az a következő érv adatává válik. Az érvrendszer azzal arányos mértékben elégti ki az igazságosság kritériumát, amilyen mértékben az állítások közötti következtetések a tesztben érdekelt felek számára meggyőzően alátámaszthatók és igazolhatók. A tesztfejlesztő az érvrendszerben fentről lefelé, a teszthasználó pedig az ellenkező irányba haladva építi fel az okfejtését, melynek eredménye a tesztalkalmazás igazolása úgy a tesztfejlesztő által létrehozott, mint a vizsgáztató által használt teszt esetében.

Mindegyik állításnak két fő összetevője van: az *eredmény* és egy vagy több *eredményminősítő*. Az 5. ábrán az eredmények az állítások felett látható megnevezések, míg az eredményminősítők az állítások idézőjelekkel (»...«) kiemelt részei. Így a tesztalkalmazási következményekről megfogalmazott állításban a „tesztalkalmazási következmények” az eredmény és a „jótékonyosság” az eredményminősítő. A tesztalkalmazási érvrendszer igazolhatóságának egyik fontos feltétele, hogy abban az összes eredményminősítőnek, illetve a hozzájuk kapcsolódó bizonyítékoknak szerepelniük kell (személyes beszélgetés L. Bachmannal, 2019. március 23.). Ez a megközelítés, szemben Kane érvrendszerével, pontosan meghatározza, mit kell bizonyítani az igazolási folyamat során. A tesztalkalmazási érvrendszerben használt eredményminősítők meghatározásai az 1. mellékletben található.

Azon túl, hogy mit kell bizonyítani a tesztalkalmazási érvrendszer igazolása során, Bachman és Palmer a bizonyítás menetét is leírják, mégpedig az általuk javasolt érvrendszerre specifikusan, így ebből a szempontból is felülmúlják Mislevy és Kane megközelítéseit, akik az érvek sematikus leírására ugyancsak Toulmin (2003) érvelési modelljét használják. Továbbá Bachman és Palmer a modellt hatékonyan igazították saját céljaikhoz. Ellentétben Kane-nel, aki a Toulmin-modellt mindössze a validálási keretrendszere érvelésstruktúrájának általános leírásaként használja, illetve Mislevyvel, aki hibásan módosította a modell Ellenvetés komponensét, Bachman és Palmer (2010) adaptációja a céljaiknak megfelelő több kiegészítéssel sikeresen és meggyőzően pontosítja a Toulmin-modellt, bár egyik elemét nem a klasszikus módon használja.

Bachman és Palmer (2010) az érvelési modell *Állítás* összetevőjét úgy pontosítják, hogy a kijelentés topikja az eredmény, fókusz pedig az eredményminősítő. Az *Adat* számukra a megfigyelt vizsgáztató viselkedés vagy az érvrendszer részét képező és az érvelési láncban korábban szereplő érv állítása. A *Biztosíték* összetevő a Toulmin-modellben betöltött funkcióval rendelkezik, és ebben az érvrendszerben van a legpontosabban leírva: azazhogy a jellegét az Állítás tartalma határozza meg, vagyis mindegyik Állításhoz egyedi Biztosíték és ahhoz egyedi *Támogatás* tartozik. A szerzők az *Ellenvetés* összetevő két formáját különböztetik meg, és egy ellenérvrendszerrel egészítik ki. Az egyszerűbb változata a cáfolat, tehát például amíg az érv állítása az, hogy „A tesztalkalmazási következmények jótékonyak”, a cáfolat az ellenérvet fogalmazza meg, miszerint „A tesztalkalmazási következmények *nem* jótékonyak”. A tesztfejlesztő feladata ez esetben az ellenérv cáfolata egy cáfoló érvrendszer (lásd Tankó 2012: 83) felállításával. Az ellenvetés összetettebb változata





\* a megfigyelt vizsgázói teljesítmény rögzítése pl. osztályzat, szóveges értékelés vagy százalékos értékelés formájában

**5. ábra. Tesztalkalmazási érvrendszer (Bachman-Damböck 2018: 31)**

a tesztben érdekelt felek által megfogalmazott kételyek, melyek eloszlatása ugyancsak cáfoló érvrendszerek segítségével valósítható meg.

Ami viszont nem a hagyományos módon van jelen a Toulmin-modell Bachman és Palmer adaptációjában, de szemben Kane megközelítésével szervesen bele van építve az érvelési folyamatba, az a *Minősítés* összetevő. Ez nem összetévesztendő az érvelési rendszerben szereplő *eredményminősítő* fogalommal, de azzal összefüggésben van.

A *Minősítés* összetevő Toulmin értelmezése alapján az állítás erősségét jelöli, és emiatt kifejezetten fontos szerepe van, ugyanis a tudományos igazolás korszerű értelmezése szerint is (pl. Popper 1997) az állítás valószínűségi értékkel rendelkezik, amely a hamis és igaz végpontok között helyezkedik el. Ha egy ellenérvet sikerül megcáfolni, akkor elvileg egyáltalán nincs erre a komponensre szükség. Ellenben, ha az érdekelt felek által megfogalmazott kételyeket csak egy bizonyos mértékig lehet eloszlatni, akkor azt egy minősítéssel jelölni kell. Bachman és Palmer az állítások valószínűségi fokának klasszikus módon történő jelölése helyett a négy érv eredményminősítőibe építik bele a valószínűség fogalmát. Ahogy az *1. mellékletben* látható, minden meghatározás azzal a kifejezéssel kezdődik, hogy az eredményminősítő valaminek a *mértéke*.

Bachman és Palmer tesztalkalmazási érvrendszere nemcsak gyakorlati, de elméleti szempontból is felülmúlja a Mislevy-féle bizonyíték alapú tesztfejlesztési keretrendszer, illetve a Kane-féle tesztpontszám értelmezési és alkalmazási érvrendszer által leírt érvelés alapú validálási eljárás módokat. A tesztalkalmazási érvrendszer a tesztfejlesztési és -alkalmazási folyamatokat minden tesztben érdekelt számára átlátható és követhető módon írja le. Bachman és Damböck (2018) útmutatásai alapján akár a nyelvtanár is készíthet tanóráira egyedül vagy kollégáival közösen elméletileg jól megalapozott, az igazságosság kritériumának megfelelő és igazolható tesztek. Akár a nyelvtanár is megvizsgálhatja, mennyire igazolható egy mások által készített teszt, melyet a nyelvórán szeretne használni, vagy a tesztalkalmazási érvrendszer segítségével felül is vizsgálhatja a saját mérési és értékelési gyakorlatát.

Mivel a tesztalkalmazási érvrendszer nyelvtanárok számára is értelmezhető nyelvezetet használ, a benne szereplő fogalmak egyértelműen vannak meghatározva, és logikusan strukturált, segítségével bármelyik tesztfejlesztési és -alkalmazási vagy mérési és értékelési folyamat érthetővé és egyben ellenőrizhetővé válik. Popper azt állítja, „egy igazolás akkor *objektív*, ha elvileg bárki ellenőrizheti és megértheti” (1997: 54–55). Tehát az érvrendszer a tesztfejlesztésben és -alkalmazásban érdekeltnek lényegesen szélesebb köre számára teremt lehetőséget az objektív igazolásra. Ezáltal több szempontból megszűnik a méréseméleti szakembereknek kiváltságos státusza, ami pozitív visszacsatolást eredményezhet nemcsak a nyelvi mérés és értékelés gyakorlatára, hanem a nyelvoktatásra is. Például sokan azt vallják, rossz gyakorlat, ha a nyelvtanárok arra készítik fel a nyelvtanulókat, hogy jól teljesítsenek standardizált teszteken. Ez valóban lehet rossz gyakorlat, de csak akkor, ha például egy teszt nem életszerű nyelvhasználatot mér, nem életszerű feladatokat tartalmaz, vagy olyan ismereteket, képességeket, készségeket vagy stratégiákat mér, melyek nem teszik lehetővé az idegen nyelv valós élethelyzetekben történő célszerű és sikeres alkalmazását. Ellenben, ha a nyelvtanár a tesztalkalmazási érvrendszer segítségével kritikusan megvizsgál egy tesztet, akkor tájékozott döntést tud hozni, miszerint azt használni fogja vagy sem az oktatási gyakorlatában, arra ’trenirozza’ a tanulóit vagy sem. Ugyanis a jó tesztfeladat a

valós élethelyzetek kommunikációs feladatainak megoldásához szükséges stratégiákat ugyanúgy aktivizálja, ahogy azokat a nyelvtanuló egy valós élethelyzetben alkalmazná. Az érvrendszert használva olyan tesztet tud fejleszteni, melyek a sikeres nyelvelsajátításhoz szükséges nyelvtanulói és nyelvtanári viselkedési mintákat erősítik. Az idegen nyelvi érettségi feladatsorok értő vizsgálata pedig lehetőséget ad a nyelvtanárnak, hogy rámutasson a feladatsorok gyengéire, és építő kritikát fogalmazzon meg annak reményében, hogy az illetékesek változtatnak a tesztfejlesztési gyakorlatukon, így a tesztek javulnak, és előnyösen hatnak vissza a nyelvtanításra.

A tesztalkalmazási érvrendszer közérthetősége részben annak tudható be, hogy a szerzők tudatosan kerülnek a vitatottságuk miatt kompromittált és túlságosan technikai jellegű *validitás*, *megbízhatóság*, *autentikusság* vagy *teszthatás* fogalmakat (Bachman–Palmer 2010: 132). Az ezek helyett megjelenő eredményminősítők közérthetőbbek, és egyben a méréselméleti követelményeknek is megfelelnek. A figyelmes olvasó bizonyára ráismert a relevancia és elégségesség eredményminősítők mögött a tartalmi validitás jellemzőire, vagy a következetesség meghatározásakor a megbízhatóság fogalmának operacionalizált változatára. Ugyancsak felismerhetők például az általánosíthatóság, illetve jótékonyág eredményminősítők sajátosságaiban az általánosíthatóság aspektus vagy a következményes validitás formában is ismert hatásaspektus Messick (1989) egységes konstruktumvaliditás-fogalmának leírásából.

A tesztalkalmazási érvrendszer közérthetősége annak is köszönhető, hogy Bachman és Palmer (2010) kérdés-felelet formájában írják le a tesztfejlesztési és -használati igazolás menetét (164–166), illetve az egyes általánosított állításokhoz tartozó biztosítékokat, támogatásokat és ellenvetéseket gyakorlatban közvetlenül felhasználható módon részletezik a tesztfejlesztő vagy felhasználó számára (158–161). Továbbá Bachman és Palmer (2010), illetve Bachman és Damböck (2018) is a tesztalkalmazási érvrendszert számos konkrét és jól követhető példán keresztül mutatják be, melyek között egyaránt megtalálhatók kis és nagy vizsgáztói létszámú, az óvodától kezdve az egyetemig használható tesztek.

A tesztalkalmazási érvrendszert az is felhasználóbaráttá teszi, hogy a tesztfejlesztést egy olyan rendszerként írja le, melyben jól elkülöníthetőek a tesztfejlesztési fázisok a tesztfejlesztési folyamat két produktumától, melyek a tesztigazolás és a teszt-előállítás (Bachman–Palmer 2010: 140). A tesztigazolás részét képezi a fentebb leírt tesztalkalmazási érvrendszer. A Mislevy- és Kane-féle megközelítések közül egyik sem tudja a tesztfejlesztés és -alkalmazást egy ilyen jól áttekinthető rendszerként bemutatni. Bachman és Palmer megközelítésében az egyes fázisok, folyamatok, valamint az azok közötti kapcsolódási pontok is vizuálisan szemléltetve vannak.

A klasszikus fogalomtár kiváltása és az azok mögöttes tartalmának egységes rendszerbe foglalása egy újabb paradigmaváltás kezdetét jelzi a mérés és értékelés tudományágban. Ugyancsak a változás jele, egy olyan egyértelműségre és közérthetőségre való törekedés a nyelvi mérés és értékelés elméletében, ami ezt a tudományágot sikeresen gyakorlatközelbe tudja hozni.

## 5. Konklúzió

A tanulmányban tárgyaltak szerint egyértelművé vált, hogy a mérés és értékelés korszerű elméletei szerint a validitás a teszteredmények értelmezésének és felhasználhatóságának jellemzője. Ugyancsak egyértelművé vált, hogy a validálás egy érvelési folyamat, melynek szerepe a teszteredmények értelmezésének és alkalmazásának indoklása. Az ezek által jelzett paradigmaváltás már 1955-ben megkezdődött, amikor Cronbach és Meehl kijelentették, „Nem a tesztet validáljuk, hanem azt az elvet, melynek segítségével következtetéseket fogalmazunk meg” (Cronbach–Meehl: 297). Ehhez képest az újdonságot Bachman és Palmer (2010) értékelési folyamat igazolása nevű tesztminősítési megközelítése jelenti. A tanulmányban felvázolt három megközelítés közül nemhogy az övük a legígéretesebb, hanem az egyben egy újabb paradigmaváltást is jelent a mérés és értékelés elméletében.

Bachman és Palmer megközelítésében a tesztalkalmazás a kimeneti pont, következőképpen az értékelési folyamat igazolása a mérést egészében és folyamatában vizsgálja. Továbbá, az igazolhatóság feltételeként azt is vizsgálja, hogy egy teszt megfelel-e minden eredményminősítő által leírt feltételnek. Ezáltal pontosan teljesíti azokat a követelményeket, melyeket Messick (1989) progresszív mátrixaként ismert folyamatábrája támaszt. Az eddigi legmeggyőzőbb és legteljesebb módon képezi le ugyanis az egységes konstruktumvaliditás-fogalmának osztatlan és folyamatszerű jellegét. Emiatt kijelenthető, hogy Bachman és Palmer (2010) számára sikerült eddig a leghűebben operacionalizálni Messick (1989) validitáselméletét.

A Bachman és Palmer-féle megközelítés központi elemét képező tesztalkalmazási érvrendszer egy bonyolult elméletet képez le egyértelműen, és tesz minden tesztben érdekelt fél számára követhetővé és kivitelezhetővé egy okfejtési folyamat formájában. Az értékelési folyamat igazolása megközelítés tesztalkalmazási érvrendszere révén egy rég várt paradigmaváltás tanúi vagyunk, mely a méréselméletnek és a nyelvi mérés és értékelés elméletének felhasználóközelivé válását teszi lehetővé. Felmerül viszont a kérdés, hogy ha egyes kontextusokban az 1955-ben elkezdődött paradigmaváltás sem érezteti egyelőre a teljes hatását, mikorra várható érzékelése és gyakorlatba való átültetése az újnak? Ez a tanulmány reményeim szerint megtette azokat a lépéseket, melyek segítségével ez az újabb paradigmaváltás a magyar kontextusban is megvalósulhat.

Ami Bachman és Palmer (2010) nyelvi értékelési megközelítését potenciálisan a létező leghatásosabbá és ezáltal a legfigyelemreméltóbbá teszi az az, hogy tudást oszt meg, és ezáltal erősít. Ha a szülő, a nyelvtanár, az iskolaigazgató, a minisztériumi döntéshozó ennek a tudásnak a birtokában van, akkor annak pozitív visszacsatolási hatása várható nemcsak a mérési és értékelés gyakorlatra, hanem az oktatási tevékenységekre is. Ezt a megközelítést tehát a tanárképzési felsőoktatási programok részeként és tanártovábbképzéseken oktatni kell.

## IRODALOM

- Alderson, J. C. – Clapham, C. – Wall, D. (1995): *Language test construction and evaluation*. Cambridge, UK: Cambridge University Press.
- Anderson, T. – Schum, D. – Twining, W. (2005): *Analysis of evidence (2nd ed.)*. Cambridge, UK: Cambridge University Press.
- Bachman, L. – Damböck, B. (2018): *Language assessment for classroom teachers*. Oxford, UK: Oxford University Press.
- Bachman, L. – Palmer, A. (2010): *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford, UK: Oxford University Press.
- Chapelle, C. A. (1999): Validation in language assessment. *Annual Review of Applied Linguistics*, 19, 254–272.
- Cronbach, L. J. (1971): ‘Test validation.’ In Thorndike, R. L. (ed.) *Educational Measurement*. Washington, DC: American Council on Education, 443–507.
- Cronbach, L. J. (1980): Validity on parole: How can we go straight? New directions for testing and measurement – Measuring achievement over a decade – Proceedings of the 1979 ETS Invitational Conference. San Francisco: Jossey-Bass, 99–108.
- Cronbach, L. J. (1988): Five perspectives on the validity argument. In: Wainer, H. – Braun, (eds.) *Test validity*. Hillsdale, NJ, US: Lawrence Erlbaum Associates, 3–17.
- Cronbach, L. J. – Meehl, P. E. (1955): Construct validity in psychological tests. *Psychological Bulletin*, 52/4, 281–302.
- Ebel, R. L. (1961): Must all tests be valid? *American Psychologist*, 16, 640–647.
- Henning, G. (1987): *A guide to language testing: Development, evaluation, research*. Cambridge, MA: Newbury House.
- House, E. R. (1980): *Evaluating with validity*. Beverly Hills, CA: Sage.
- House, E. R. (2014): Origins of the ideas in evaluating with validity. In: Griffith, – Montross-Moorhead, B. (eds.) *Revisiting truth, beauty, and justice: Evaluating with validity in the 21st century*. New Directions for Evaluation, 142, 9–15.
- Hughes, A. (1989): *Testing for language teachers* (1st ed). Cambridge, UK: Cambridge University Press.
- Kane, M. (1999): Validating measures of performance. *Educational measurement: Issues and practice*, 18/2, 5–17.
- Kane, M. (2006): Validation. In: Brennan, R. (ed.), *Educational measurement* (4th ed.). Westport, CT: American Council on Education and Praeger, 17–64.
- Kane, M. (2010): Validity and fairness. *Language Testing*, 27/2, 177–182.
- Kane, M. (2011): Validating score interpretations and uses: Messick lecture, language testing research colloquium, Cambridge, April 2010. *Language Testing*, 29/1, 3–17.
- Kane, M. (2012): Articulating a validity argument. In: Fulcher, G. – Davidson, F. (eds.) *The Routledge handbook of language testing*. New York, NY: Routledge, 34–47.
- Kane, M. T. (2013): Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50/1, 1–73.
- Kontra, H. E. – Kormos, J. (2007): *An introduction to language testing for teachers of English*. Budapest: Okker.
- McNamara, T. F. – Roever, C. (2006): *Language testing: The social dimension*. Malden, MA: Blackwell.

- Messick, S. (1989): 'Validity.' In: Linn, R. L. (ed.) *Educational Measurement*. New York, NY: Macmillan/American Council on Education, 13–103.
- Messick, S. (1995): Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50/9, 741–749.
- Mislevy, R. J. (2003): Substance and structure in assessment arguments. *Law, Probability, and Risk*, 2/4, 237–258.
- Mislevy, R. J. – Steinberg, L. S. – Almond, R. G. (2002): Design and analysis in task-based language assessment. *Language Testing*, 19/4, 477–496.
- Mislevy, R. J. – Steinberg, L. S. – Almond, R. G. (2003): On the structure of assessment arguments. *Measurement: Interdisciplinary Research and Perspectives*, 1/1, 3–62.
- OH-HYAK (2019): Fogalomtár. *Akkreditációs Kézikönyv (53–65)*.  
[https://nyak.oh.gov.hu/nyat/doc/ak2018/Akkreditacios\\_Kezikonyv\\_2018.pdf](https://nyak.oh.gov.hu/nyat/doc/ak2018/Akkreditacios_Kezikonyv_2018.pdf)
- Popper, K. R. (1997): *A tudományos kutatás logikája*. Budapest: Európa Könyvkiadó.
- Shaw, S. – Crisp, V. (2015): *Reflections on a framework for validation – Five years on*. Research Matters: A Cambridge Assessment Publication, 19, 31–37.
- Tankó, Gy. (2012): *Professional writing: The academic context* (2nd ed.). Budapest: Eötvös University Press.
- Tankó, Gy. (2015): Investigating the context validity of an English academic writing test. In: Holló, D. & Károly, K. (eds.) *Inspirations in foreign language teaching*. London, UK: Pearson Education, 128–150.
- Toulmin, S. E. (1958): *The uses of argument*. Cambridge, UK: Cambridge University Press.
- Toulmin, S. E. (2003): *The uses of argument* (Updated ed.). Cambridge, UK: Cambridge University Press.
- Toulmin, S. E. – Rieke, R. – Janik, A. (1984): *An introduction to reasoning* (2<sup>nd</sup> ed.). New York, NY: MacMillan.
- Van Eemeren, F. H. – Grootendorst, R. – Snoeck-Henkemans, A. F. (2002): *Argumentation analysis, evaluation, presentation*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Verhelst, N. (2006): Általánosíthatósági elmélet. In: Banerjee, J. – Kaftandjieva, F. – Takala, S. – Verhelst, N. *Szintillesztési módszertani segédlet. A nyelvvizsgák illesztése a Közös európai referenciakerethez című kézikönyv előzetes, kísérleti verziójához*. Budapest: Nyelvvizsgát Akkreditáló Testület – Nyelvvizsgáztatási Akkreditációs Központ, 128–150.
- Weir, C. (2005): *Language testing and validation*. Basingstoke, UK: Palgrave Macmillan.
- Wigmore, J. H. (1913): *The science of judicial proof*. Norwood, MA: J. S. Cushing.



## 1. MELLÉKLET

Állítástípus	Eredményminősítő
Tesztalkalmazási következmény	<i>Jótekonyság:</i> Annak a mértéke, hogy a teszt alapján meghozott döntések és a teszthasználát következményei mennyire kedvező hatásúak, és mennyire nem károsítják a tesztben érdekelt feleket.
Döntés	<i>Értéktudatosság:</i> Annak a mértéke, hogy a teszt alapján meghozott döntések és a teszthasználát következményei mennyire veszik figyelembe az aktuális nevelési és társadalmi értékeket, valamint a vonatkozó törvényeket, szabályokat és szabályozásokat. <i>Méltányosság:</i> Annak a mértéke, hogy a mért képesség azonos szintjén levő vizsgázókat milyen eséllyel sorolják be ugyanabba a csoportba.
Értelmezés	<i>Relevancia:</i> Annak a mértéke, hogy az értelmezések a tesztfelhasználó számára mennyire biztosítják a döntéshozáshoz szükséges információkat. <i>Elégségesség:</i> Annak a mértéke, hogy az értelmezések a tesztfelhasználó számára mennyire biztosítják a döntéshez szükséges mennyiségű információt. <i>Közérthetőség:</i> Annak a mértéke, hogy az értelmezések 1. mennyire tájékoztatják a tesztben érdekelt feleket a teszt által mért készség(ek)ről, 2. olyan terminológiával vannak-e megfogalmazva, ami a tesztben érdekelt felek számára érthető, és amit magukra tudnak vonatkoztatni. <i>Általánosíthatóság:</i> Annak a mértéke, hogy az értékelés alapján megfogalmazott értelmezések mennyire alkalmazhatók a vizsgázók nyelvhasználati területén belül, vagy mennyire terjeszthetők ki azokra a területekre. <i>Elfogulatlanság:</i> Annak a mértéke, hogy a feladatok formátuma és tartalma, illetve a vizsgaadminisztráció összes aspektusa mennyire mentes a vizsgázónak kedvezhető vagy ártatható részrehajlástól.
Eredményjegyzék	<i>Következetesség:</i> Annak a mértéke, hogy a diákok különböző mérések alkalmával (pl. különböző vizsgaalkalmak során, eltérő feladatokat tartalmazó tesztekkel vagy különböző javítókkal/értékelőkkel folytatott mérés esetén) azonos nyelvi képességen belül mért teljesítménye lényegében ugyanazokat az értékelési jegyzékeket eredményezi (a vizsgázó ugyanazt a pontszám, osztályzat, százalékos érték vagy szöveges értékelés formájában kifejezett eredményt éri el).

**Bachman és Palmer (2010: 253) tesztalkalmazási érvrendszerében használt eredményminősítői**