

Szegedi Tudományegyetem  
Nyelvtudományi Doktori Iskola  
Angol Alkalmazott Nyelvészeti Doktori Program

**Az Euroexam International vizsgaközpont English for  
Academic Purposes nyelvvizsga íráskészség feladatainak  
kontextus és értékelési validitás vizsgálata**

**Doktori értekezés magyar nyelvű tézisei**

**Fűkőh Borbála**

**Témavezető:  
Dr. Barát Erzsébet**

**Szeged  
2020**

A nagy tétű (high-stakes) nyelvvizsgák fontos szerepet játszanak Magyarország oktatási rendszerében. A nyelvvizsgákat az Oktatási Hivatal Nyelvvizsgáztatási Akkreditációs Osztálya felügyeli, amely a magyarországi vizsgarendszerek szakmai követelményeinek egységesítéséért felelős (Oktatási Hivatal, 2020).

Idegen nyelvű képzésben való továbbtanuláshoz a magas szintű nyelvtudás mellett nélkülözhetetlen azoknak a készségeknek a megléte, amelyek képessé teszik a hallgatókat arra, hogy a tanulmányaikat megfelelő szinten végezzék. Az angol nyelvű képzések felvételi eljárásaihoz a nyelvtudás mellett az oktatásban való részvételhez szükséges nyelvi készségek bizonyítása is szükséges. Erre a célra az egész világon legelterjedtebbek az IELTS, a Pearson Academic és a TOEFL iBT – English for Academic Purposes (EAP) vizsgái, amelyek az angol nyelvet oktatási közegben használók számára készültek. Ezek a vizsgák mindegyike szintező vizsga, minden egyes országban, ahol elérhetők, azonos formában kerülnek meghirdetésre, nyilvánosak a szintillesztésükkel kapcsolatos tanulmányok, így biztosítva az átláthatóságot és a megbízhatóságot.

A fent említett angol nyelvvizsgák ugyan elérhetőek és letehetőek Magyarországon, azonban magyar akkreditációval nem rendelkeznek, tehát nem minden esetben vonzóak azon vizsgázók számára, akik a felsőoktatási felvételijük során esetleg magyarországi idegen nyelvű képzéseket is megcélznak. Az Euroexam International vizsgaközpont célja tehát egy olyan tudományos szaknyelvi vizsga létrehozása, amely valós alternatívát nyújt a fent felsorolt EAP vizsgák mellett azon a magyar vizsgázók számára, akik angol nyelvű felsőoktatásban szeretnék folytatni a tanulmányaikat itthon vagy külföldön.

A disszertáció az Euroexam International Vizsgaközpont English for Academic Purposes nyelvvizsga fejlesztési folyamatáról számol be, különös tekintettel a két íráskészséget mérő feladat (formális tranzakciós email és esszéírás validitásának vizsgálatára. A kutatás az írásfeladatok konstruktumának érvényességére fókuszált (Fűkőh, 2019a, 2019b), mely kvalitatív és kvantitatív módszereken alapult. A validitás vizsgálat szakaszainak célja annak bizonyítása, hogy az Euroexam International EAP vizsgájának írásbeli feladatai tükrözik azokat a készségeket, amelyekre a hallgatóknak angol nyelvű képzést nyújtó felsőoktatási programokban szükségük van.

A dolgozat két fő részre osztható. Az első rész a vonatkozó szakirodalom áttekintése négy fejezetben. A kutatás háttérének bemutatása után a 2. fejezet összefoglalja az írás jellegével és az írás folyamatával kapcsolatos releváns elméleti munkákat, különös

tekintettel az oktatásban használatos írás jellegére és műfajaira. A 3. fejezet az írás értékelésének témájával foglalkozik. Az íráskészség értékelésének általános tárgyalása mellett az idegen nyelvű íráskészség és annak értékelése is megjelenik. Az értékelés témakörében külön foglalkozom az értékelői szigorral valamint a Közös Európai Referenciakeretben foglaltakkal (KER, Council of Europe, 2001; 2018). A fejezet utolsó része az értékelési skálák és a lista alapú értékelés használatát, a két értékelési eszköz előnyeit és hátrányait tárja fel. A 4. fejezet áttekinti a tesztfejlesztés és a validitási vizsgálatok elméleteit, melyek közül hangsúlyosan tárgyalom Bachman és Palmer (1996; 2010) jószágmutatóit. A fejezet második része a nyelvi mérésben megjelenő validitási fogalmak bemutatásával foglalkozik. Ennek keretében összefoglalom az érvényesség hagyományos és új fogalmait, valamint részletesen tárgyalom Weir (2005) szocio-kognitív elméleti keretét és a lokalizáció fogalmát (O’Sullivan & Dunlea, 2015), amely különös jelentőséggel bír a nemzetközi egyetemi közegben tett nyelvvizsgákban.

Az 5. fejezet a kutatáson alapuló vizsgafejlesztés és a validitási vizsgálatok módszereit tárgyalja. Ez a fejezet fogalmazza meg a kutatási kérdéseket is. Az értekezés második részében a saját empirikus kutatásomat három fejezetben mutatom be. Ezek a fejezetek követik az Euroexam English for Academic Purposes vizsga fejlesztési folyamatának szakaszait. A 6. fejezet felvázolja a fejlesztés kezdeti szakaszát, azaz a fejlesztési folyamat tervezését, a szakterület elemzését, a konstruktum vizsgálatára irányuló kutatást, valamint a külső szakértői testület javaslatainak összefoglalását. A 7. fejezet a validitási vizsgálat két szakaszát mutatja be: a vizsgafeladatok tesztelésére irányuló kismintás próbát és a nagymintás pretesztet. A vizsgaspecifikáció véglegesítése előtt fontos összeállítani a vizsgakonstruktum és a feladatok részletes leírását, valamint a feladatok próbaverzióját. Ezek segítségével empirikus adatok alapján szükséges vizsgázói és értékelői visszajelzéseket gyűjteni. Fontos, hogy a próba és a preteszt alanyai a célpopulációnak megfelelőek legyenek, valamint a preteszt populációnak elegendő adatközlőt kell tartalmaznia ahhoz, hogy a felvett adatokat klasszikus tesztelméleti módszerekkel elemezni lehessen.

A kvalitatív és kvantitatív adatelemzés eredményei két szempontból fontosak. Egyrészt a verbális protokollok elemzése és a nagymintás preteszt segít megállapítani az írási feladatok érvényességét. Másrészt a vizsgálatok rávilágítottak az értékelési validitás egyes kérdéseire, amelyeket a 8. fejezetben problematizáltam. Ebben a fejezetben áttekintem az előző szakaszokban az értékeléssel és értékelői magatartással kapcsolatos

kérdéseket, és bemutatom egy műfaj- és szintspecifikus lista alapú értékelési eszköz fejlesztését. Végül a kutatás eredményeit a 9. fejezetben foglalom össze, amely visszatekint a szekvenciális felépítésű validitási vizsgálatra, valamint a lista alapú értékelés módszer alkalmasságát és előnyeit tárgyalja.

### **Az értekezés hipotézisei és kutatási kérdései**

Jelen kutatás célja az Euroexam English for Academic Purposes vizsga írásfeladatainak validitás vizsgálata, azaz annak feltárása, hogyan tükrözi az íráskonstruktum a felsőoktatásban szükséges képességeket és készségeket, valamint azt tekintem át, hogyan lehetséges annak elérése, hogy az eredmények megbízható pontszámokon alapuljanak, amelyek elfogulatlan értékelés alapján születtek. A dolgozat bemutatja az írásfeladatok kutatáson alapuló tesztfejlesztési folyamatát, különös figyelmet fordítva az 1. feladat (formális tranzakciós e-mail) kontextus validitására és a 2. feladat (érvelő esszé) értékelési validitására. A kutatási kérdések a következők:

**1. kutatási kérdés:** Érvényes feladattípusnak tekinthető a tranzakciós email egy English for Academic Purposes nyelvvizsgában?

A fő kérdés tehát, hogy az érvelő szövegtípusokon kívül a tranzakciós írás érvényes feladatként jelenhet-e meg az EAP vizsgákban. A kutatási hipotézis azt állítja, hogy a tranzakciós írás is része a hallgatók nyelvi és kommunikációs repertoárjának. A szakterület tudományos és szakmai nyelvhasználata mellett az egyetemisták életének része a tanulmányaik megszervezése, az egyetemi adminisztrációval és a regisztrációval kapcsolatos kérdések kezelése. A tanulmányi követelményeknek való megfelelés mellett a hallgatóknak meg kell felelniük a tanulmányaikkal kapcsolatos formális kommunikáció szabályainak is. Ezen feltételezés alapján a tranzakciós írás szintén az egyetemi élethez köthető nyelvhasználat része.

A disszertáció 6. fejezete empirikus kutatással és szakértői megítéléssel vizsgálja ezt a kérdést. A szakterület elemzésének részeként egy kismintás empirikus vizsgálatot végeztem a következő másodlagos kutatási kérdésekre fókuszálva:

- a) Melyek a leggyakoribb írásbeli műfajok az egyetemi hallgatók és az egyetemi oktatók, dolgozók közötti kommunikációban?
- b) A hivatalos angol nyelvű írásbeli kommunikáció része-e az egyetemisták célnyelvi használatának?
- c) Mennyire fontos a célnyelvi kommunikációban a formalitás?

A kérdések kvalitatív módszerekkel igyekeztek feltárni a feladat kontextus validitását. Az első kutatási kérdésre kismintás vizsgálat eredményei alapján születettek válaszok, a másodlagos kutatási kérdésekkel pedig a szakértői értékeléshez használt kérdőív foglalkozott. Ezek együttesen szolgálták az előzetes feladattervezés alapjául.

Az értékelési validitást illetően a kutatás mind kvantitatív, mind kvalitatív vizsgálaton alapul. A 8. fejezetben szereplő verbális protokollok és statisztikai elemzések eredményei mutatják be a lista alapú értékelés előnyeit a skála alapú értékeléssel szemben. A kutatási hipotézisemben azt állítom, hogy egy feladat- és szintspecifikus lista alapú értékelési eszköz javítja a 2. feladat értékelésének objektivitását és megbízhatóságát. A hipotézist a következő kutatási kérdéseken keresztül teszteltem:

**2. kutatási kérdés:** Az értékelési skálához képest a lista alapú értékelés javítani tud-e

- az érvelő esszék értékelésének minőségén és
- az értékelők megbízhatóságán?

A klasszikus tesztelmélet alkalmazásával végzett elemzés során a másodlagos kutatási kérdések a következők:

- a) A lista alapú értékelés megbízhatósága (Cronbach alfa) elég magas-e a magyar akkreditációs követelmények teljesítéséhez?
- b) Hogyan működik a lista alapú értékelési eszköz az item nehézség és minőség szempontjából?
- c) A lista alapú értékelés képes-e elkülöníteni a populációban az alacsonyan és magasan teljesítőket?
- d) A lista alapú értékelés befolyásolja-e az esszé feladat sikerességi rátáját?

A nyelvvizsgákon elért eredmény több szempontból fontos a vizsga résztvevői és a többi érdekelt fél (stakeholders) számára, akik a következők lehetnek: egyetemek, vizsgaközpontok, vizsgafejlesztők és értékelők. Az Euroexam értékelési folyamatának egy korábban feltárt fő problémája, hogy vizsgázóknak és az értékelőknek eltérő felfogásuk van a sikeres teljesítményről (Lukácsi, 2017). Ezen túlmenően a szakirodalomból ismert, hogy az értékelés nagyban függ az értékelők személyes megítélésétől és a holdudvar hatástól (halo effect) (Knoch, 2009). Köztudott, hogy még „a jól képzett, tapasztalt értékelők is szisztematikusan különböznek egymástól a rutinszerűen alkalmazott pontozási kritériumok értelmezésében” (Eckes, 2009, 5. o.). Az Euroexam International korábbi kutatásai (Lukácsi, 2017; 2018, 2020) bebizonyították, hogy egy szint- és műfajspecifikus

lista alapú értékelési eszköz növeli a B2 szintű tranzakciós írási feladat pontozásának objektivitását és megbízhatóságát.

Az Euroexam értékelőivel végzett, a 7. fejezetben bemutatott verbális protokollok célja az volt, hogy feltárjam, hogyan viszonyulnak az értékelők a C1 szintű esszé feladathoz az értékelési folyamat során. A verbális protokollok arra is rávilágítanak, hogy az általuk a jól megformált esszéhez társított jellemzők hogyan és miért különböznek egymástól. Az értékelés érvényességével és az értékelési eszköz összefüggésével kapcsolatban fogalmaztam meg a következő kutatási kérdést, amelyet kvalitatív módszerekkel vizsgáltam:

**3. kutatási kérdés:** A lista alapú értékelés növelheti-e az értékelők műfaji tudatosságát?

7. fejezetben bemutatott verbális protokoll elemzés alapján a 8. fejezet a lista alapú értékelési eszköz fejlesztési folyamatát ismerteti. A lista szint- és feladatspecifikus, valamint ellenőrző kérdéseket tartalmaz annak érdekében, hogy a konstruktum állandóság biztosítva legyen, és közös alapot biztosítson az esszé műfajának értékeléséhez.

A vizsgafejlesztés során az objektív és elfogulatlan értékelés bizonyítása az egyik legfontosabb követelmény. Ez az egyik legfontosabb szempont, amit a nemzetközi felsőoktatási intézmények figyelembe vesznek egy vizsgaeredmény beszámítása során, hiszen az egyetemre történő belépéshez szükséges nyelvi követelmények a felvételi folyamat „kapuőrei” (Nagy, 2000). Annak biztosítása érdekében, hogy a felsőoktatási intézménybe pályázók valóban rendelkeznek a nyelvvizsga-bizonyítványukban szereplő tudással, azaz azokkal a készségekkel, melyek sikeressé teszik őket a felsőoktatási tanulmányaik során, olyan értékelési eszközt kell használni, amelynek segítségével nem csupán megbízhatóan lehet értékelni, hanem az eszköz használata esetlegesen pozitívan visszahat a vizsgázók képességeire is. A nyelvvizsgák ugyanis pozitív hatással lehetnek a tanítási gyakorlatra, a készségfejlesztésre, és „hatással lehetnek az egyes vizsgázók karrierjére vagy életlehetőségeire” (Taylor, 2005, 154. o.). Mivel az angol nyelvtudást „árucikknek” lehet tekinteni (Cameron, 2000), ezért kiemelt fontosságú, hogy a vizsgázók és az egyetemre jelentkezők tisztában legyenek az oktatási szakterületen használt különböző diskurzusokkal (Weninger & Khan, 2013).

Az íráskészség mérésére használt feladatok értékelési eljárása fontos szerepet foglal el a nyelvi méréssel kapcsolatos kutatások között. A szubjektív értékelésű írásfeladatok értékelési validitásával, valamint az értékelési skálákkal kapcsolatos vizsgálatok nagy

számban vannak jelen a szakirodalomban. Megállapítható, hogy az írásproduktumok értékeléséhez a legtöbb értékeléssel kapcsolatos kézikönyv az értékelési skálát ajánlja (Alderson et al., 1995; Bachman & Palmer, 1996; McNamara, 1996; Shaw & Weir, 2007; Weigle, 2002, Weir, 2005). Ugyanakkor számos publikáció (Eckes, 2009; Knoch, 2009; Knoch, 2011, Lukácsi, 2017; 2018; 2020; McNamara, 2000) és maga a KER 9. fejezete is felvetette az értékelési skálák alkalmasságának pontos definiálást, és hangsúlyozta a következetes értékelés szükségességét. Egy, a skálánál objektívabb értékelési eszköz használata várhatóan csökkenti az értékelők közti különbségeket és növeli műfajtudatosságukat. Ennek következtében további előnyei is várhatóak a lista alapú értékelésnek. Megjósolhatunk egy pozitív visszahatást (washback hatás), amelynek következtében a vizsgázók műfaji tudatossága is fejlődik, valamint íráskészségük javul, amely növeli a percepciójuk és a valós teljesítményük közötti összefüggés megítélésnek helyességét.

### **Módszertan**

A fejlesztés célja egy helyi igényeknek megfelelő tudományos szaknyelvi vizsga létrehozása, valamint a konstruktum érvényességének megállapítása és bizonyítása. A kevert módszerű kutatásra alapuló, a konstruktum validitásának megállapításához kapcsolódó vizsgálat módszertana az előkészítési szakaszban Bachman és Palmer (1996) jóságmutatóit használja; a fejlesztési szakaszban, a specifikáció elkészítéséhez, a teszt itemek létrehozásához és próbájához Weir (2005) szocio-kognitív keretét valamint Read (2015) gyakorlatát követi.

Amint azt a szakirodalmi áttekintés során bemutattam, Bachman & Palmer (1996; 2010) bevezette azt a modellt, melyben a vizsgáknak hat fontos tulajdonságát határozták meg, és hangsúlyozták ezek figyelembevételét a vizsgafejlesztés során. Ahhoz, hogy a vizsga érvényes legyen, egyszerre kell teljesítenie hat feltételt, amelyeket jóságmutatónak neveztek. A vizsga a hat feltétel szerint akkor mér jól, ha biztosítjuk (a) az értékelés megbízhatóságát, (b) a konstruktum érvényességét, (c) a célnyelv autentikus használatát, (d) a vizsgázó több készségének együttes használatát, (e) a pozitív teszthatást, valamint (f) a vizsgafeladatok egyszerű gyakorlati lebonyolíthatóságát. Bár Bachman és Palmer korai munkája a mai napig meghatározó a vizsgafejlesztési szakirodalomban, később Bachman (2005) rámutatott, hogy ezek a kategóriák egymástól függetlenek, tehát nem feltétlenül lehetséges az összes jóságmutatót magas szintre emelni. Meggyőződni csupán arról tudunk, „hogy a fejlesztés során hozott döntéseink védhetők, és hiteles bizonyítékokkal

támasztottuk őket alá a validálási folyamat során” (Bachman, 2005, 5. o.). El kell fogadni, hogy bizonyos kompromisszumokat szükséges kötni, és a „tökéletes vizsga” helyett be kell érünk az „elég jó vizsga” fogalmával. A hat jóságmutató közül azok, amelyek az EAP vizsgafejlesztés szempontjából a legfontosabbak (a) a megbízhatóság, (b) a konstruktum érvényessége és (c) a célnyelv autentikus használata (Chan, 2013). A legfontosabb tehát olyan feladatok tervezése, amelyek a különböző vizsgaidőszakokban összehasonlítható eredményeket nyújtanak, valóban azt mérik, amit mérni akarunk, és a célnyelv valós használatára építenek.

Bachman és Palmer (1996) fent kiemelt három jóságmutatója Weir (2005) szociokognitív elméleti keretében is megjelenik. A konstruktum érvényességét illetően Bachman és Palmer azt állítja, hogy a konstruktum érvényességét az adott kontextusban kell vizsgálni. Ezt az elképzelést fejlesztette tovább Weir (2005), aki a konstruktum érvényességét összefoglaló értelemben használja, és az érvényesség új aspektusait vezeti be a validitás fogalmába. Elméleti keretében a konstruktumot a kontextus határozza meg, és az autentikus nyelvhasználat a kontextus érvényességének szerves részeként jelenik meg. Egy írásfeladat esetén tehát a kontextus érvényességének vizsgálata nem más, mint egy feladat nyelvi és tartalmi igényeinek feltérképezése, valamint a valós írási feladatok vizsgálata a célnyelvi használatban. Weir szintén új fogalomként vezeti be az értékelési validitást, amelybe a megbízhatóság fogalmát építette be (Weir, 2005; Shaw & Weir 2007).

Az értekezés a két feladat érvényességének kérdését két különböző módon kezeli. A kutatás kvalitatív és kvantitatív részei egymást kiegészítőnek tekinthetők, a kevert módszerű kutatás szekvenciális, iteratív struktúrát mutat, amelyben az egyes szakaszok eredményei és következtetései a következő szakaszok tervezésében épülnek be (Creswell, 2009, 14. o.). A validálás négy szakaszát és az általam kifejezetten az Euroexam EAP vizsga írásbeli feladatainak fejlesztésére tervezett folyamatokat az 1. táblázat mutatja be.



## 1. Táblázat

### *Euroexam EAP vizsga – Validálási szakaszok*

<b>1. szakasz</b>	<b>2. szakasz</b>	<b>3. szakasz</b>	<b>4. szakasz</b>
<b>1 &amp; 2 feladat</b>	<b>1 &amp; 2 feladat</b>	<b>1 &amp; 2 feladat</b>	<b>2. feladat</b>
<b>Előkészítő szakasz</b>	<b>A specifikáció és a teszt itemek létrehozása és kipróbálása</b>	<b>A feladatsor pretesztelése</b>	<b>Az értékelési validitás fejlesztésének érdekében lista alapú értékelési eszköz fejlesztése</b>
Tervezés	A szakterület elemzés eredményeinek modellálása és kismintás próba	A teszteredmény- validitás megfigyelése	Az értékelési lista és ellenőrző kérdések fejlesztése
A szakterület elemzése	Vizsgálói percepciók gyűjtése	Vizsgálói kérdőív	Verbális protokoll elemzés
A konstruktum vizsgálata			
Szakértői elemzés	Vizsgálói és értékelői adatok – verbális protokoll elemzés	Vizsgaeredmények - statisztikai elemzés	Vizsgaeredmények és értékelői visszajelzés – statisztikai elemzés

A fejlesztési folyamat szakaszainak lépéseit az Euroexam C1 szintű általános nyelvi vizsga íráskonstruktumának figyelembevételével tettem meg. Azaz az első három, standard validálási szakaszban mindkét feladattípust vizsgáltam, a vizsgálat fókuszja azonban eltérő a két feladat esetében. A két feladat együtt jelenik meg a kezdeti fejlesztési szakaszban, majd a kismintás próba és a nagymintás preteszt szakaszában. A 4. szakaszt annak érdekében adtam a validitási vizsgálathoz, hogy feltárjam, miként lehet javítani az esszé feladat értékelési validitását.

A jelen kutatáson alapuló vizsgafejlesztési folyamat a konstruktum validitást használja arra, hogy az eredményeket keretbe foglalja (Kane, 2013). A validitás vizsgálata mind elméleti, mind empirikus bizonyítékokon alapul, ahol a különböző validitások (kontextus és értékelési validitás) a köztük lévő kölcsönhatásuk révén kapcsolhatók össze (Shaw & Weir, 2007).

### **A kutatás eredményei**

Az 1. szakaszban egy kismintás előzetes vizsgálatot végeztem azzal a céllal, hogy definiáljam a konstruktumot és empirikus bizonyítékokat állítsak elő a tranzakciós írás kontextus validitására az adott szakterületen belül. Az 1. kutatási kérdés megválaszolásához (Érvényes feladattípusnak tekinthető a tranzakciós email egy English for Academic Purposes nyelvvizsgában?) egyetemi hallgatókkal ( $N = 5$ ) és oktatókkal,

valamit nem oktató felsőoktatási dolgozókkal ( $N = 6$ ) készítettem félig strukturált interjúkat. Megvizsgáltam, hogy a hallgatók hogyan kommunikálnak az egyetem munkatársaival, és azt, hogy az egyetemen használt diskurzus részének lehet-e tekinteni a tranzakciós írást. Az vizsgálatban használt másodlagos kutatási kérdésekre adott válaszok alapján megerősíthető az eredeti hipotézis: a tranzakciós írás az része az adott szakterületnek. Mivel a kevert módszerű kutatásom szerkezete iteratív jellegű volt, a kutatás 1. szakaszának eredményei beépültek az új vizsga előzetes specifikációjának elkészítésébe. A konstruktum definiálása és az feladatok előzetes ellenőrzése a kutatásban jelen lévő triangulálásnak köszönhetően külső szakértői elemzés tárgya is volt. A szakterület elemzése ebben a vizsgálati szakaszban nem empirikus adatokon, hanem már rendelkezésre álló ismereteken és tapasztalatokon és szakértői megítélésen alapult. A triangulálás értelmében a külső szakértők nem kapták meg az empirikus eredményeket, a független ítéletük szerint a javasolt tesztfeladatok és a konstruktum érvényesnek mondhatók az adott területen. A kutatásomban részt vevő három nemzetközi szakértő észrevételei az 1. szakasz végén elősegítették a specifikációk kidolgozását. Az 1. és a 2. feladat esetében tranzakciós és érvelő írást használva lehetővé vált az íráskonstruktum megtartása az *Akkreditációs Kézikönyv*-ben (Oktatási Hivatal, 2019) szereplő profil akkreditáció követelményeinek megfelelően.

Az 1. szakasz után a 2. szakasz célja a vizsgaspecifikációk és a 3. szakaszban pretesztre bocsátott feladatok elkészítése volt. A szakterület elemzéséhez hasonlóan a 2. szakaszban végzett modellezés is empirikus adatokra épült. A kvalitatív adatfelvétel során vizsgázói teljesítményt, valamint a vizsgázói és az értékelői interjúkat használtam annak megállapítására, hogy a két javasolt feladattípus megfelel-e a konstruktumnak. Ezen a ponton fontos kiemelni, hogy a 2. szakaszban alkalmazott kvalitatív adatfelvétel a vizsgafeladatok kipróbálására is szolgált, melyhez az 1. szakasz végén elkészített feladatokat használtam. A szakterület modellezésének és a feladatok kipróbálásának összekapcsolása két okból lényeges. Egyrészt a műfajok elméleti konstrukciója nem alkalmas a felhasználói visszajelzések gyűjtésére, másrészt a mintafeladatok kipróbálásával egyszerre figyelhetjük meg a vizsgázók és a vizsgáztatók stratégiáit, valamint a feladat jellemzőit, azaz értékelési érvényességéhez tudunk adatokat gyűjteni.

A kismintás próba során potenciális vizsgázókat ( $N = 6$ ) kértem meg a vizsgafeladatok megoldására, amely után félig strukturált interjúkat készítettem, majd végeztem el ezek és a megoldott feladatok elemzését. A 2. szakasz fő célja a teszt

specifikációinak véglegesítése volt, amely a próba során a vizsgázóktól felvett adatok alapján történt. Az értékelési validitásra vonatkozó adatok összegyűjtése érdekében a 2. szakaszban az Euroexam International tapasztalt akkreditált értékelői is részt ( $N = 3$ ) vettek. A három értékelő ugyanazon három vizsgázó írásproduktumait értékelte hangosan gondolkodás közben. Az értékeléshez az akkreditált Euroexam C1 értékelési skálát használták. A verbális protokoll elemzés eredményeként megállapítható, hogy az Euroexam értékelői eltérően vélekedtek a pontértékekről, ezáltal különböző szigorúsággal pontoztak, valamint eltérő értelmezésük volt a konstruktumról. Megállapítható volt, hogy az értékelési skála homályosan megfogalmazott deskriptorai akadályozták az értékelési folyamat objektivitását és a pontszámok megbízhatóságát.

A kutatás ciklikus felépítésének megfelelően a 2. szakasz végén a vizsgázói észrevételek alapján újraterveztem az 1. feladatot, valamint elkészítettem az Euroexam EAP vizsga részletes feladatspecifikációját. A 3. szakaszban, a nagymintás preteszteléshez a feladatok ezen formátumát használtam.

A 3. szakasz a fejlesztésben a nagy mintán alapuló vizsgát foglalta magában ( $N = 136$ ). Ennek a szakasznak az volt a célja, hogy kvantitatív módszerekkel tárja fel, hogy a tesztfeladatok megfelelően működnek-e annak érdekében, hogy az Euroexam EAP vizsga C1 standard szintjét a KER-hez képest meg lehessen állapítani, valamint bizonyítékot szolgáltatson a megfelelő értékelési validitásra. Az vizsgafeladatok pretesztelése mellett egy vizsgázói kérdőívet is használtam, amelynek segítségével a vizsgázók nyelvtanulási háttéréről és az önértékeléséről gyűjtöttem adatot, valamint a vizsgázók véleményét kértem a vizsga formájáról és tartalmáról.

A vizsgázói eredmények statisztikai elemzése bebizonyította, hogy az Euroexam EAP vizsga a C1 szintű írástudás érvényes mérésére alkalmas, azonban az adatok alapján további kérdések merültek fel a két írásfeladat értékelési validitásával kapcsolatban. Az adatelemzés során alkalmazott rekurzív folyamatoknak köszönhetően világossá vált, hogy a korábbi szakaszok eredményei mind egy bizonyos irányba vezetnek: rávilágítottak az értékelői különbségekre valamint az értékelés hibáira. A 2. és a 3. szakasz eredményei egymást erősítve arra a jelenségre hívták fel a figyelmet, hogy az értékelési skála használata jelentős mértékben torzítja az íráskészség értékelését.

A kutatás 4. szakasza az értékelési skála kiváltása érdekében egy lista alapú értékelési eszköz fejlesztésére összpontosított. A kutatás célja az volt, hogy növelni tudjuk

az esszé feladat értékelési validitását és az eredmények megbízhatóságát. A kutatás és fejlesztés Lukácsi (2017; 2018; 2020) korábbi vizsgálatainak módszerét követte. A szakasz célja egy olyan feladat- és szintspecifikus lista alapú értékelési eszköz kidolgozása volt az EAP vizsga esszé feladatához, amely használatával kompenzálni lehet az egyéni értékelési különbségeket és az értékelői hatást. A 4. szakasz két fő részből állt: dokumentumelemzésből (1. lépés) és empirikus kutatásból (2-8. lépés). Az empirikus kutatást további két részre bontható: a 2-6. lépések az ellenőrzőlista tételeinek megtervezésére és fejlesztésére összpontosítottak, a 7-8. lépések pedig a skála- és lista alapú értékelés pontszámainak összehasonlítására, a köztük lévő kapcsolat feltárására fókuszált.

A 4. szakasz eredményei alapján egyértelműen kijelenthető, hogy a javasolt lista alapú értékelési eszköz használata javítja az Euroexam EAP vizsga esszéfeladatának értékelési validitását. A **2. kutatási kérdés** (Az értékelési skálához képest a lista alapú értékelés javítani tud-e az érvelő esszék értékelésének minőségén és az értékelők megbízhatóságán?) és az ahhoz kapcsolódóan megfogalmazott másodlagos kutatási kérdésekre adott válaszok alapján a 4. szakasz végén a következő eredményekről lehet beszámolni: mivel a dichotóm itemek és az ellenőrző kérdések kisebb teret adnak a szubjektív értékelésre, a lista alapú értékelés növeli az értékelők és az értékelés objektivitását. Az értékelők közötti megbízhatóság magasabb szintjét különféle statisztikai elemzések bizonyították (teljes azonosság, ICC és Krippendorff alfa). A másodlagos kutatási kérdések konkrét statisztikai mutatók alapján levonható következtetéseket céloztak meg, melyek alapján elmondható, hogy az értékek megfelelnek a magyar akkreditációs követelményeknek (Oktatási Hivatal, 2019):

- a) A végleges 30 tételes ellenőrzőlista alapján kapott pontszámok megbízhatósága ( $\alpha = 0,90$ ) teljesíti az akkreditációs követelményt, mely szerint  $\alpha \geq 0,75$ .
- b) Az itemek nehézség és minőség tekintetében is megfelelnek az *Akkreditációs Kézikönyv* azon előírásainak, hogy a p-értékek több mint 80%-a és a diszkriminációs indexek (Ebel-féle D) 90% -a az elfogadható tartományba esik ( $0,70 \geq p\text{-érték} \geq 0,30$ ; Ebel-féle  $D \geq 0,30$ )
- c) Az Ebel-féle D magas értékei azt jelzik, hogy az lista alapú értékelési eszköz képes elkülöníteni a populáció magas és alacsony teljesítményű tagjait. Ez egyértelműen megfigyelhető a pontértékek eloszlásának diagramján is, amelyen

szélesebb pontszám tartományt figyelhetünk meg a központi tendencia jelenléte nélkül.

- d) Noha a pontértékek más eloszlást mutatnak, az esszéfeladat skála- és lista alapú eredményeinek összehasonlítása és páros t-próbával összevetett sikerességi rátája nem különbözik egymástól. Elmondhatjuk, hogy a lista alapú értékelés nem szigorúbb, mint a skálán alapuló értékelés, és nem befolyásolja a megfelelési százalékot.

Ami a **3. kutatási kérdést** illeti (A lista alapú értékelés növelheti-e az értékelők műfaji tudatosságát?), a kvalitatív adatok elemzése alapján elmondható, hogy a lista alapú értékelési eszköz dichotóm elemei és az ellenőrző kérdések használata növeli az értékelők műfaji tudatosságát. A kutatás adatközlői a fejlesztés során a fokozott objektivitás érzéséről és a jól használható eszköz iránti pozitív hozzáállásról számoltak be.

A lista alapú értékelési eszköz fejlesztési eredményei alapján azt a következtetést lehet levonni, hogy minimalizálni lehet az eredmények értékelők általi torzítását, csökkenteni lehet az erős központi tendencia rossz gyakorlatát a pontértékek eloszlásában (Eckes et al., 2016), valamint az értékelőket az műfaji kritériumok közös megértése felé lehet irányítani. A lista alapú szint- és műfajspecifikus értékelési eszköz fejlesztési folyamata és annak eredményei alapján azt állíthatjuk, hogy az lista alapú értékelési eszköznek számos előnye van. Megállapítható, hogy a kutatási terv a jövőben adaptálható egy hasonló értékelési eszköz kifejlesztésére az Euroexam C1 szintű írásfeladatainak értékeléséhez is.

### **Összegzés**

A disszertáció célja az Euroexam International English for Academic Purposes (EAP) vizsga írásbeli feladatainak validitási vizsgálata volt, valamint egy lista alapú értékelési eszköz kidolgozása speciálisan az érvelő esszé műfajára. A disszertáció legfontosabb megállapításai az Euroexam International helyi fejlesztésű vizsgájának íráskészség feladatainak érvényességét érintik. A négy szakaszból álló kutatáson alapuló fejlesztési és validálási folyamat eredményei alapján megerősítést nyert, hogy az új EAP vizsga írásfeladatai alkalmasak az angol nyelvtudás mérésére az adott szakterületen. A kutatás további eredménye egy lista alapú értékelési eszköz kifejlesztése és validálása, amelynek használatával megnövelhető az írásfeladatok közül az érvelő esszé értékelési validitása.

## Hivatkozások

- Alderson, J. C., & Clapham, C. (1995). Assessing student performance in the ESL classroom. *TESOL Quarterly*, 29(1), 184–187.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2(1), 1–34.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Cameron, D. (2000). Styling the worker: Gender and the commodification of language in the globalized service economy. *Journal of Sociolinguistics*, 4(3), 323–347.
- Chan, S. H. C. (2013). *Establishing the validity of reading into writing test tasks for the UK academic context*. PhD thesis. University of Bedfordshire. Letöltve 2020. június 20.: <http://uobrep.openrepository.com/uobrep/handle/10547/312629>
- Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press. Letöltve 2020. június 20.: <https://rm.coe.int/1680459f97>
- Council of Europe (2018). *Common European framework of reference for languages: Learning, teaching, assessment. Companion volume with new descriptors*. Strasbourg: Language Policy Division. Letöltve 2020. június 20.: <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>
- Creswell, J. W. (2009). *Research Design. Qualitative, Quantitative and Mixed Methods Approaches*. Thousand Oaks, CA: Sage Publications.
- Eckes T. (2009). Many-facet Rasch measurement. In S. Takala (Ed.), *Reference supplement to the manual for relating language examinations to the common European framework of reference for languages: Learning, teaching, assessment (Section H)* (pp. 1–52). Strasbourg, France: Council of Europe/Language Policy Division.

- Fűkőh, B. (2019a május). *Research-based EAP test development: local needs and opportunities on an international context*. Paper presented at the 14<sup>th</sup> UPRT Conference, University of Pécs, Pécs, Hungary.
- Fűkőh, B. (2019b). Kutatáson alapuló tesztfelkészítés – írásfeladatok egy angol tudományos szaknyelvi vizsga számára. In: (Besznyák R. (Ed.) *Porta Lingua – 2019. Interdiszciplináris megközelítések a szaknyelvoktatásban és -kutatásban* (pp. 271–288). Budapest: Szaknyelvoktatók és -Kutatók Országos Egyesülete. <http://szokoe.hu/kiadvanyok/porta-lingua-2019>
- IELTS (2018). *IELTS test format*. Letöltve 2020. június 16.: <https://www.ielts.org/about-the-test/test-format>
- Kane, M. T. (2013). The argument-based approach to validation. *School Psychology Review*, 42(4), 448–457.
- Knoch, U. (2009). *Diagnostic assessment of writing: The development and validation of a rating scale*. Frankfurt: Peter Lang.
- Knoch, U. (2011). Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from? *Assessing Writing*, 16, 81–96. <https://doi.org/10.1016/j.asw.2011.02.003>
- Lukácsi, Z. (2017, May). *Developing a level-specific checklist for assessing writing*. Paper presented at the 14<sup>th</sup> EALTA Conference, Ciep, Sèvres, France.
- Lukácsi, Z. (2018). Írásművek lista alapú értékelése – avagy hogyan mérjük a nyelvvizsgán. *NyelvVilág*, XXI, 7–23.
- Lukácsi, Z. (2020). Developing a level-specific checklist for assessing EFL writing. *Language Testing*. <https://doi.org/10.1177/0265532220916703>
- McNamara, T. F. (1996). *Measuring second language performance*. New York, NY: Longman.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- Nagy, P. (2000). The three roles of assessment: Gatekeeping, accountability, and instructional diagnosis. *Canadian Journal of Education*, 25(4), 262–279.
- Oktatási Hivatal (2019). *Akkreditációs Kézikönyv 2019*. Letöltve 2020. június 16.: <https://nyak.oh.gov.hu/nyat/doc/ak2019/ak2019.htm>

- Oktatási Hivatal (2020). *Alapfogalmak*. Letöltve 2020. június 16.: <https://nyak.oh.gov.hu/doc/alapfogalmak-eng.asp>
- O’Sullivan, B., & Dunlea, J. (2015). *Aptis general technical manual*. London: British Council.
- Pearson Academic (2017). *Test format*. Letöltve 2020. június 16.: <https://pearsonpte.com/the-test/format/>
- Read, J. (2015). *Assessing English proficiency*. Palgrave: Macmillan.
- Shaw, S., & Weir, C. J. (2007). *Examining writing: Research and practice in assessing second language writing*, Studies in Language Testing (Vol. 26). Cambridge: UCLES/Cambridge University Press.
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2), 154–155.
- TOEFL iBT (2018). *Test content*. Letöltve 2020. június 20.: <https://www.ets.org/toefl/ibt/about/content/>
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Weir, C. J. (2005). *Language testing and validation: an evidence-based approach*. Basingstoke: Palgrave Macmillan.
- Weninger, C., & Khan, K. H.-Y. (2013). (Critical) language awareness in business communication. *English for Specific Purposes*, 32, 59–71.
- 137/2008. (V. 16.) Korm. rendelet az idegennyelvtudást igazoló államilag elismert nyelvvizsgáztatás rendjéről és nyelvvizsga bizonyítványokról. Letöltve 2020. június 16.: [http://njt.hu/cgi\\_bin/njt\\_doc.cgi?docid=119127](http://njt.hu/cgi_bin/njt_doc.cgi?docid=119127)

### **Releváns publikációk**

- Füköb, B. (2016a). Developing Writing Skills of Students of Business English. *NyelvVilág*, XX, 7–18.
- Füköb, B. (2016b). Nyelvismeret, nyelvtanulási motiváció és második idegen nyelv: a nyelvi és motivációs felmérés néhány tanulsága a BGE turizmus–vendéglátás szakos I. évfolyamos hallgatók egy csoportjában. In: T. Váradi (Ed.) *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből 2016: X.*



*Alkalmazott Nyelvészeti Doktoranduszkonferencia* (pp. 19–30). Budapest: MTA Nyelvtudományi Intézet.

- Fűköh, B. (2018). Student interviews in establishing the context validity of an EAP writing task. In: (R. Besznyák (Ed.) *Porta Lingua - 2018. Tudásmegosztás, értékközvetítés, digitalizáció– trendek a szaknyelvoktatásban és -kutatásban* (pp. 301–309). Budapest: Szaknyelvoktatók és -Kutatók Országos Egyesülete. <http://szokoe.hu/kiadvanyok/porta-lingua-2018>
- Fűköh, B. (2019). Kutatáson alapuló tesztfejlesztés – írásfeladatok egy angol tudományos szaknyelvi vizsga számára. In: R. Besznyák (Ed.) *Porta Lingua – 2019. Interdiszciplináris megközelítések a szaknyelvoktatásban és -kutatásban* (pp. 271–288). Budapest: Szaknyelvoktatók és -Kutatók Országos Egyesülete. <http://szokoe.hu/kiadvanyok/porta-lingua-2019>
- Fűköh, B. (2017). Language learning gains and motivation of L2 and L3 learners of ESP. In: M. Lehmann, R. Lugossy, M. Nikolov, & G. Szabó (Eds.) *UPRT 2017: Empirical Studies in English Applied Linguistics* (pp. 177–188). Pécs: Lingua Franca Csoport.
- Lukácsi, Z., & Fűköh, B. (2018). Vizsgafejlesztés a magyar felsőoktatásban: Egy egyetemi szaknyelvi vizsga kidolgozása és fogadtatása. *Új Pedagógiai Szemle*, 68(9-10), 42–63.
- Lukácsi, Z., & Fűköh, B. (2020). A tantervi hatékonyság vizsgálata az egyetemi szaknyelvoktatásban. *Modern Nyelvoktatás*, 26(1-2), 28–43.

### **Konferenciaelőadások**

- Fűköh, B. (2018. szeptember). Establishing the Context and Scoring Validity of an English for Academic Purposes test. In: J. Rambousek, I. Schöfrová, & J. Chamonikolasová (Eds.) *14th ESSE Conference Abstracts* (p. 138). Brno: Masaryk University.
- Fűköh, B. (2018. szeptember). Validity Arguments for English for Academic Purposes Test Tasks. In: J. Rambousek, I. Schöfrová, & J. Chamonikolasová (Eds.) *14th ESSE Conference Abstracts* (p. 133). Brno: Masaryk University.

- Füköh, B. (2019. május). *Research-based EAP test development: local needs and opportunities on an international context*. Paper presented at the 14<sup>th</sup> UPRT Empirical Studies in Applied Linguistics Conference, University of Pécs, Pécs, Hungary.
- Lukácsi, Z., & Füköh, B. (2018. november). *Assessing academic English among Central European students for UK University admissions*. Poster presented at the Language Testing Forum 2018 Conference, University of Bedfordshire, Luton, UK.
- Lukácsi, Z., & Füköh, B. (2018. június). *A truly special examination: A case study of a local exam development project*. Paper presented at the 13<sup>th</sup> UZRT Empirical Studies in Applied Linguistics Conference, University of Zagreb, Zagreb, Croatia.
- Lukácsi, Z., & Füköh, B. (2019. május). *Assessing academic English with a localised test for international admissions purposes*. Paper presented at the 16<sup>th</sup> EALTA Conference, SIG Assessment of Writing / Assessment for Academic Purposes, University of Dublin, Ireland.