

# Az induktív gondolkodás fejlődésének összehasonlító vizsgálata: online felmérések Magyarországon és Finnországban

Csapó Benő,\* Risto Hotulainen,\*\* Pásztor Attila\*\*\* és Molnár Gyöngyvér\*\*\*\*

A tanulmány az induktív gondolkodás finn és magyar tanulók körében elvégzett összehasonlító vizsgálatának eredményeit mutatja be. Arra keresi a választ, hogy a Magyarországon kidolgozott mérőeszközök azonos módon működnek-e Finnországban, egy másik pedagógiai kultúrájú oktatási rendszerben, és a különböző életkorú diákok körében, hogyan fejlődik a diákok gondolkodása, továbbá milyen különbségek vannak a fiúk és a lányok között a két iskola-rendszerben. A felmérés résztvevői Vantaa város és az SZTE Oktatáseméleti Kutatócsoport partneriskoláinak nyegedik és hetedik évfolyamos diákjai közül kerültek ki. Bár egyik minta sem tekinthető reprezentatívnak, így a teljesítményeket nem lehet közvetlenül összehasonlítani, a felvett adatok alkalmasak a kutatási kérdések megválaszolására. A mérés eszköze egy több korábbi vizsgálatban már alkalmazott induktív gondolkodás teszt volt (figurális és numerikus analógiák és sorozatok), amelynek kiközvetítésére az eDia online tesztplatformon keresztül került sor. A finn diákok tableten, a magyarok asztali számítógépen (egér, billentyű) oldották meg a feladatokat. Az eredmények azt mutatják, hogy a tesztek reliabilitása a finnre való átültetés során nem változott lényegesen. A skála-invariancia elemzések szerint a teszt másik környezetben való felhasználása során megfelelt a konfigurális és a részleges metrikus invariancia feltételeinek, azonban a metrikus invarianciának már nem. Váratlan eredmény, hogy a magyar diákok mindkét életkorban jobban teljesítettek, mint a finnek, különösen a numerikus feladatokban. Mivel e felmérés mintái nem voltak reprezentatívak, a megfigyelt különbség megerősítése és okainak felderítése további vizsgálatokat igényel.

**Kulcsszavak:** induktív gondolkodás, technológia alapú mérés, eDia, összehasonlító vizsgálat

## Bevezetés

A tanulmány egy finn–magyar kutatási együttműködés első eredményeit mutatja be. Az adatfelvétel két, a Szegedi Tudományegyetemen folyó hosszabb távú kutatási program metszetében helyezkedik el. Az egyik program az általános értelmi képességek mérésével és fejlődésének vizsgálatával foglalkozik, és egészen az 1970-es évekig nyúlik vissza, a másik pedig a technológia-alapú tesztelésben rejlő, tanulást és tanítást segítő lehetőségek feltérképezésére fókuszál, amelynek megvalósítása közel egy évtizedes múltra tekint vissza. E két területen alakult ki kutatási együttműködés a Szegedi Tudományegyetem kutatócsoportjai és Helsinki Egyetem pedagógiai mérési központja (*Centre for Educational Assessment*) között.

Az induktív gondolkodás, illetve annak egyes komponensei központi szerepet kaptak a finn kutatók által elindított „tanulás tanulása” (*Learning to Learn*, L2) programban (Hautamäki, Arinen, Eronen, Hautamäki, Kupianen, Lindblom, Niemivirta, Pakaslahti, Rantanen & Scheinin, 2002; Vainikainen, 2014). A szegedi kutatók pedig vizsgálatok sokaságában használták az induktív gondolkodás tesztek különböző változatait, közvetlenül

\* egyetemi tanár, Szegedi Tudományegyetem Neveléstudományi Intézet, e-mail: csapo@edpsy.u-szeged.hu

\*\* egyetemi tanár, Centre for Educational Assessment, University of Helsinki, e-mail: risto.hotulainen@helsinki.fi

\*\*\* tudományos munkatárs, MTA-SZTE Képességfejlesztés Kutatócsoport, e-mail: attila.pasztor@edu.u-szeged.hu

\*\*\*\* egyetemi tanár, Szegedi Tudományegyetem Neveléstudományi Intézet, e-mail: gymolnar@edpsy.u-szeged.hu

magának az induktív gondolkodás fejlődésének vizsgálatára, vagy háttérváltozóként az általános értelmi képességek jellemzésére (l. pl. Csapó, 1994; Pásztor, 2016).

A tanulmány alapjául szolgáló adatfelvételt közvetlenül a finn kutatók szakértői felkérése indította el. Egy Helsinki melletti nagyváros, Vantaa, 2015-ben a digitális tartalmakhoz való hatékonyabb hozzáférés érdekében 16 000 tabletet vásárolt a város diákjainak és tanárainak. A Helsinki Egyetem kutatói kaptak arra megbízást, hogy az új informatikai eszközök oktatási alkalmazásának hatását megvizsgálják, akik, ismerve a Szegedi Tudományegyetem Oktatásméleti Kutatócsoportjában folyó munkát, a szegedi fejlesztéseket és kutatási tapasztalatokat is bevonták a projektbe. A hatásvizsgálatok gyors és hatékony kivitelezésének az bizonyult, hogy az eDia-rendszerben (Molnár, 2015; Csapó & Molnár, 2017, 2019; Molnár & Csapó, 2017, 2019; Molnár, Makay & Ancsin, 2018) található feladatok egy részét átültettük finn nyelvre, valamint a finn kutatókkal együttműködve további specifikus online mérőeszközök készültek.

A felhasznált mérőeszközök egyike az induktív gondolkodás teszt volt, amelynek eredményeiből a tanulmány keretein belül a negyedik és hetedik évfolyamon végzett méréseket mutatjuk be. Miután a vizsgált terület közvetlenül nem kapcsolódik a tantervi tananyaghoz, alkalmas annak az elméleti és gyakorlati szempontból egyaránt fontos kérdésnek a megvizsgálására, hogy mennyiben marad a mérési skála invariáns, ha a teszteket egy nyelvről egy másikra lefordítjuk, és a méréseket különböző életkorokban kissé különböző informatikai eszközökkel vesszük fel. Az eredmények tanulságosak abból a szempontból is, hogy a finn diákok osztálytermi környezetben, a saját személyes használatra szolgáló tabletyükön oldották meg a teszteket, míg a magyar tanulók az iskolák e célra – általában asztali számítógépekkel – felszerelt számítógépes termeiben. Így a Vantaa tablet vizsgálat tapasztalatai a későbbi magyar fejlesztések számára is hasznosak lehetnek.

## *Elméleti háttér*

### *Az induktív gondolkodás fejlődésének vizsgálata*

Az induktív gondolkodás az általános értelmi képességek egyik legtöbbet vizsgált komponense. Népszerűségének egyik oka az, hogy a legtöbb intelligenciateszt induktív feladatokat tartalmaz (pl. a Raven teszt: Raven, 2000), és az induktív gondolkodás számos más általános képességgel is kapcsolatban áll (l. Molnár, Greiff & Csapó, 2013). Például a PISA (*Programme for International Student Assessment*) felmérésben is használt dinamikus problémamegoldás feladatok megoldásának első részében szükség van a szabályindukcióra (OECD, 2014; Csapó & Molnár, 2017). Az induktív gondolkodás fontos szerepet játszik a tanulásban, az új tudás létrehozásában és a megszerzett tudás alkalmazásában, a tudástranszferben is (Csapó, 1998; Vainikainen, Hautamäki, Hottulainen & Kupiainen, 2015).

Az induktív gondolkodás korábbi értelmezése az analógiákból (hasonlóságok felismerése két mintázat, pl. szópárok vagy számpárok között) és a hasonló egyedi mintázatok közös sajátosságainak általánosításán (szabályindukció) alapszik (Fisher, 2015; Feeney, 2017). Mind a mai napig sokféle teszt, illetve vizsgálat épül erre a megközelítésre, amiben benne van az is, hogy az indukció mindig tartalmaz némi bizonytalanságot.

Egy másik megközelítés a gondolkodás műveleti szerkezetéből indul ki, egy ilyen modellt dolgozott ki például Karl Josef Klauer. E modell szerint az induktív gondolkodás lényege hasonlóságok, különbözőségek, valamint hasonlóságok és különbözőségek felismerése tulajdonságok vagy relációk között (l. pl. Klauer & Phye, 2008). E rendszert alapul véve Klauer fejlesztő gyakorlatokat is kidolgozott óvodások, az iskola első éveiben járók és tanulásban akadályozott serdülő diákok számára (Klauer, 1989, 1991, 1993). Később a modell alapján

számos kísérletet végeztek, többek között iskolai tananyagok tartalmait is felhasználva (ezeket is áttekinti: Kluwer & Phye, 2008).

Az induktív gondolkodás tesztek közös sajátossága, hogy valamilyen megfigyelés alapján fel kell ismerni a mintázatokban megnyilvánuló kapcsolatokat, szabályosságokat (szabályindukció), majd a felismert szabály alapján döntést hozni és megoldani a feladatot. A tesztek tartalma többnyire verbális, képi, geometriai, numerikus, de ritkábban egyéb tartalom is előfordul.

Korábbi vizsgálatainkban felvázoltuk az induktív gondolkodás fejlődési folyamatait papíralapú tesztekkel (Csapó, 1994, 1997, 2001), majd online tesztekkel (Molnár & Csapó, 2011; Pásztor, 2016). Elemeztük a problémamegoldó gondolkodással való kapcsolatát (Molnár et al., 2013; Wu & Molnár, 2018a, 2018b) és kísérleti körülmények között foglalkoztunk fejlesztésével is (Molnár, 2011; Pásztor, 2014).

### *A technológiaalapú tesztelés és a médiahatás vizsgálata*

A papíralapú tesztelésről a számítógépes mérésekre való átállás során az egyik központi kérdés az, hogy mennyiben befolyásolja a diákok eredményeit a feladatok különböző médiumon történő kiközvetítése (Wang, Jiao, Young, Brooks & Olson, 2008; Csapó, Molnár & R. Tóth, 2009; Csapó, Molnár & Nagy, 2014). A kérdés egyrészről vizsgálható abban a vonatkozásban, hogy a papír- és a számítógép-alapú tesztek mennyiben tekinthetők ekvivalensnek. További kérdésként merülhet fel, hogy ugyanazon online tesztek eltérően viselkednek-e különböző digitális eszközök (például tablet és asztali számítógép) alkalmazása során.

A papír- és a számítógép-alapú tesztek összehasonlítására irányuló vizsgálatok az ezredforduló után kerültek előtérbe (Hülber & Molnár, 2013). A vizsgálatok tanulsága szerint a médiahatást számos változó befolyásolhatja, függhet a vizsgálandó konstruktumtól, a feladatok típusától, a technikai feltételektől (például monitorméret, felbontás), a minta jellemzőitől (például nem, kor, családi háttér, számítógépes jártasság), és természetesen ezek a hatások egymástól nem feltétlenül függetlenek (Hülber & Molnár, 2013).

Az e tanulmányban bemutatásra kerülő területen médiahatás-vizsgálatokra már korábban is sor került (Csapó, Molnár & R. Tóth, 2009). Az elemzések általános konklúziója az, hogy a teljes teszt szintjén mind papír-, mind számítógép-alapon megbízhatóan alkalmazhatóak a feladatok, a különbségek nem számottevők. A médiahatást a DIFER teszt egyes résztesztjei, valamint egy figuratív itemeket tartalmazó induktív mérőeszköz felhasználásával is vizsgáltuk első évfolyamos tanulók bevonásával (Csapó, Molnár & Nagy, 2014). Az induktív feladatok esetében a teszt megbízhatósága mindkét médiumon megfelelő volt, a teljesítmények között nem volt szignifikáns különbség. Részletesebb invarianciaelemzések arra is rámutattak, hogy a két médium között a teszt működésében sem volt különbség (Csapó, Molnár & Nagy, 2014). A figuratív itemek használata során megjelenő médiahatás elemzéséhez további támpontot adhatnak a népszerű Raven-teszt felhasználásával végzett idevágó vizsgálatok. Az eredmények egyöntetűen azt támasztják alá, hogy a Raven-teszten nem jelentkezik médiahatás a papír-alapú és a számítógépes adatfelvétel között (Williams & McCord, 2006; Arce-Ferrer & Guzmán, 2009).

Az eddigi kutatási eredmények szerint az itt bemutatásra kerülő vizsgálatban is alkalmazott induktív gondolkodás feladatok a papír-alapú és a számítógépes adatfelvétel összehasonlításában nincsenek jelentősen kitéve a média befolyásoló hatásának. Amennyiben a feladatok nem tartalmaznak hosszabb szövegeket, illetve a feladatelemek kellőképpen nagyméretűek, akkor feltételezhetően a tabletek kisebb kijelzője sem képvisel jelentős torzító hatást az asztali számítógépeken végzett mérésekhez képest.

## Összehasonlító vizsgálatok és a mérési skálák invarianciája

A mérési invariancia tesztelése alapvető fontosságú a különböző tulajdonságú minták és rész-minták (pl. fiú-lány, évfolyam szerinti bontás, különböző országokban, nyelvi környezetben felvett adatok stb.) teljesítményének összevetése során. Az eredmények összehasonlításakor ugyanis fel kell tételeznünk, hogy a két adatfelvétel során mért látens változó azonos felépítésű, struktúrájú és működésű (Byrne, 2008). Ha ezek a feltételek nem állnak fenn, vagy nem tudjuk, hogy fennállnak-e, nem lehetünk abban biztosak, hogy ugyanannak a konstruktnak a mérését valósítottuk meg, jelen esetben Magyarországon és Finnországon.

Dichotóm változók esetében a mérési invariancia tesztelése (Vandenberg & Lance, 2000; Muthén & Muthén, 2010) több lépésben történik többcsoportos megerősítő faktoranalízisek sorozata segítségével. A mérési modell (*basic model*) meghatározása után három beágyazott modell egyre szigorúbb feltételek melletti ellenőrzésére (I. Byrne & Stewart, 2006) került sor (Schroeders & Wilhelm, 2011; Greiff, Wüstenberg, Molnár, Fisher, Funke & Csapó, 2013).

Az invariancia első szintje a konfigurális invariancia, ami azt vizsgálja, hogy a mérési modell struktúrája invariáns-e a csoportok között (Byrne, 2008), azaz jelen esetben mindkét csoportban (magyar és a finn minta) ugyanúgy konceptualizálják-e a mért konstruktnot (Milfont & Fischer, 2010), az egyes vizsgált változók ugyanazt a látens változót mérik-e mindkét csoportban. Az invariancia tesztelésének második szintjén, a metrikus invariancia vizsgálata során, a mért változók által meghatározott látens faktorok faktorsúlyainak összevetése valósul meg. Tesztelése csak akkor releváns, ha a konfigurális invariancia fennáll, azaz modellje illeszkedik az adatokra. A szigorú metrikus invariancia, ami a csoportátlagok összehasonlíthatóságának feltétele, az invariancia harmadik szintje, tesztelése csak akkor releváns, ha fennáll a metrikus invariancia. A szigorú metrikus invariancia megléte biztosítja, hogy a látens változók skálájának nemcsak a mértéke egyezik meg (metrikus invariancia), hanem a skála viszonyítási pontja is azonos, azaz a látens változókat mérő két skála megegyezik egymással (Blunch, 2008). Ha a látens változó skálái megegyeznek, a látens változók átlagai összehasonlíthatóak egymással (Byrne, 2008). A szigorú metrikus invariancia megléte jelen esetben nem előfeltétel, azaz nem feltétele az eredmények összehasonlíthatóságának (Meredith, 1993; további alkalmazásokat illetően I. Davidov, Schmidt & Billiet, 2011).

Az egyre szigorúbb feltételeknek megfelelő modellekhez való illeszkedés egyik indikátora az RMSEA (*Root Mean Square Error of Approximation*) illeszkedésindex, illetve az illeszkedésindexhez tartozó konfidencia intervallum alsó és felső értéke. Ha mindhárom érték 0,05 alatti, akkor egyértelműen kijelenthető, hogy az adatok jól illeszkednek az adott modellhez. Az RMSEA indexre építő elemzések mellett elvégezhető az illeszkedések mértékének azonosságát vizsgáló, mintaméretre érzékeny khi-négyzet teszt is. A tanulmány keretein belül mindkét típusú elemzésre sor került, szem előtt tartva a finn és a magyar minták nagyságának különbözőségét.

## Kutatási kérdések

A vizsgálatban résztvevő diákok két különböző iskolarendszerben tanulnak, ami két különböző pedagógiai kultúrát is jelent. Ezért a vizsgálat első kérdése, hogy a mérőeszköz egyformán működik-e a két ország különböző életkorú diákjain, a magyarról finnrre való átültetés hatására változik-e a reliabilitása, illetve azonosak-e a mérésre használt skálák.

A minta két életkorra oszlik, ami lehetőséget ad a fejlődés keresztmetszeti adatok alapján történő becslésére.

Korábbi vizsgálatokból tudjuk, hogy a fiúk és a lányok teljesítménye között jelentős különbségek lehetnek. Adataink alapján mindegyik országban és életkorban elemezhetjük a fiú-lány különbségeket.

Megvizsgáljuk továbbá, hogy mi befolyásolja a teszteredményeket, a rendelkezésre álló háttérváltozók közül mivel lehet értelmezni a teljesítmények különbségeit.

## Módszerek

### *A vizsgálat kontextusa*

A finn mérések a már említett projekthez kapcsolódtak, melynek keretében Vantaa városa 16 000 tabletet vásárolt az oktatási intézmények számára ingyenes wifi hozzáféréssel óvodáktól a középiskolákig (*The tablets in Vantaa: New joy and interest in learning*). A Helsinkí Egyetem pedagógiai mérési központja (Centre for Educational Assessment) 2015 és 2018 között végezte el az eszközfejlesztés hatásának nyomon követését, melynek megvalósításába bevonta az SZTE Oktatásméleti Kutatócsoport és az MTA SZTE Képességfejlesztés Kutatócsoport munkatársait. A követéses vizsgálat általános célja az új eszközök hatásának elemzése a tanítás-tanulási folyamatokra, valamint a diákok motivációjára. Ennek érdekében egy óvodai (hazai kontextusban nagycsoport), egy 4., egy 7., és egy 10. évfolyamos minta longitudinális vizsgálata indult el 2015 őszén. Az ezt követő adatfelvételek 2016 és 2017 tavaszán valósultak meg. A minta kialakítása a város különböző körzeteinek megfelelő reprezentálásával történt.

A programban több mérőeszköz alkalmazására is sor került. A szegedi kutatók által fejlesztett inductív gondolkodás tesztek mellett a kognitív hatások további elemzése céljából egy vizuális memória teszt is helyet kapott a vizsgálatokban. A gondolkodást mérő teszteken túlmenően több tanulói kérdőív is készült az eDia rendszerben, melyek a tabletek tanórai alkalmazására (például mely tantárgyakban, milyen gyakran, vagy, hogy milyen feladatok során használták), illetve a motivációjuk mérésére irányultak. Az olvasási nehézségek hatásának kiküszöbölése érdekében az óvodai mintán és annak követése során a feladatok instrukciói és a kérdőívek szöveges részei meghallgathatók voltak. Egy adott mérési ciklus lezárását követően az adatok rendszerezésével a finn kutatók készítették el az intézményeknek szóló visszajelentéseket.

A magyar adatfelvétel tág életkori intervallumban, 2–12. évfolyamig terjedő, átfogó gondolkodásmérés keretein belül valósult meg. A közel fél évig tartó adatfelvétel során a diákok négy mérési ciklusban oldották meg a különböző nehézségű, de évfolyamok között összehorgonyozott teszteket. Az adatfelvétel 2–3. évfolyamon 2 tanórát, 4–5. évfolyamon 3 tanórát, 6–11. évfolyamon 4 tanórát vett igénybe és a következő területeket érintette:

- Számítógépes egérhasználat (2–4. évfolyam)
- Induktív gondolkodás (2–11. évfolyam)
- Vizuális memória (2–11. évfolyam)
- Tanulási stratégiák (2–11. évfolyam)
- Interaktív problémamegoldó képesség (4–11. évfolyam)
- Kombinatív gondolkodás (5–11. évfolyam)
- Kreativitás (5–11. évfolyam)
- Internetes információkeresés hatékonysága (6–11. évfolyam).

A diákok olvasási képességében lévő különbségek teszten nyújtott teljesítménybefolyásoló hatását kétféle módszerrel igyekeztük kiküszöbölni. Alsóbb évfolyamon (1–3. évfolyamon) minden egyes feladat instrukciója

fülhallgató segítségével meghallgatható volt, illetve minden egyes évfolyam minden egyes tesztjében minimalizáltuk az olvasnivaló szöveg mennyiségét. A kiközvetített tesztek alapvetően grafikus feladatokat alkalmaztak, amelyek megoldása minimális olvasást igényelt (kivétel a kizárólag magasabb évfolyamos diákok számára kiközvetített internetes információkeresés hatékonyságát mérő teszt).

### *A felmérések mintái*

A felmérésben a Helsinki Egyetem *Vantaa Tablet Study* kutatási programjának iskolái és a Szegedi Tudományegyetem Oktatáselméleti Kutatócsoport eDia programjának partneriskolái vettek részt. A következő elemzésekhez mindegyik országból a negyedik és a hetedik évfolyam eredményeit vonjuk be. A két minta alapvető adatait az 1. táblázat tartalmazza.

Évfolyam		Finn	Magyar
4.	N	633	2491
	Életkor átlag	10,26	10,50
	Életkor szórás	0,304	0,437
	Lányok aránya	51%	49%
7.	N	733	1750
	Életkor átlag	13,36	13,58
	Életkor szórás	0,388	0,447
	Lányok arány	49%	51%

1. táblázat: A magyar és a finn minta alapadatai az adatfelvétel pillanatában

Figyelemre méltó, hogy, bár a finn diákok hétévesen kezdik az iskolát, a magyar diákok mindkét évfolyamon valamivel idősebbek. Ugyanakkor a szórás adatok jelzik, hogy a magyar diákok beiskolázási intervalluma sokkal szélesebb.

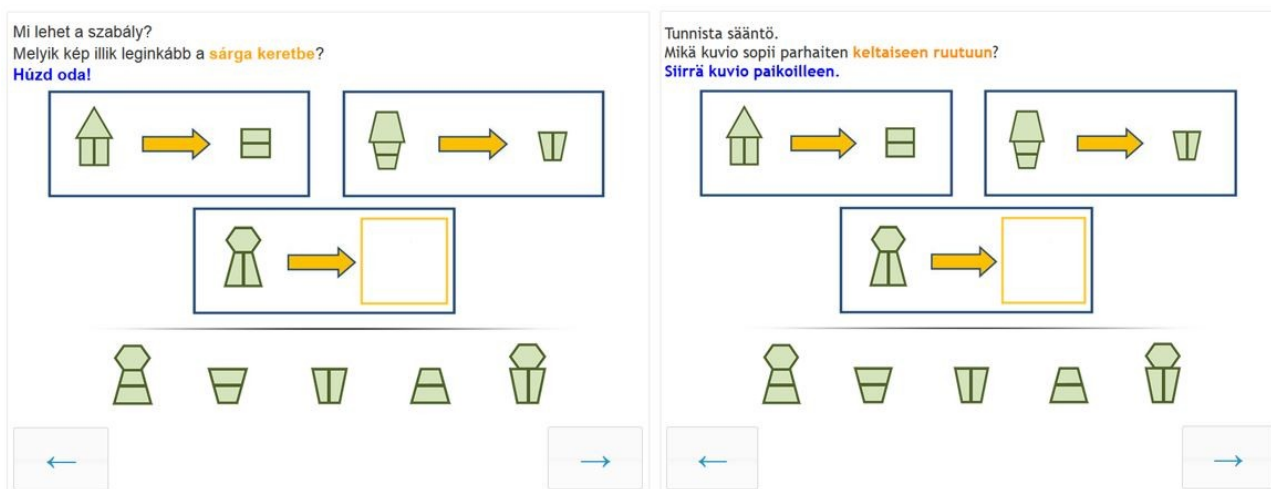
A mintavétel egyik országban sem felel meg a reprezentativitás elveinek, de kellően nagy és megfelelő szórással rendelkezik ahhoz, hogy a kutatási kérdések megválaszolásához szükséges elemzéseket elvégezzük. A PISA vizsgálatokból tudjuk, hogy a finn iskolák között nagyon kis különbségek vannak. Ennek alapján feltételezhetjük, hogy a Vantaa városból származó minta teljesítménye közel áll az országos átlaghoz, annál némileg jobb. Az eDia partneriskoláiról tudjuk, hogy közöttük a hátrányos helyzetű iskolák alulreprezentáltak, így feltételezhetjük, hogy az adott minta teljesítménye némileg jobb, mint az országos átlag.

### *A mérések eszközei*

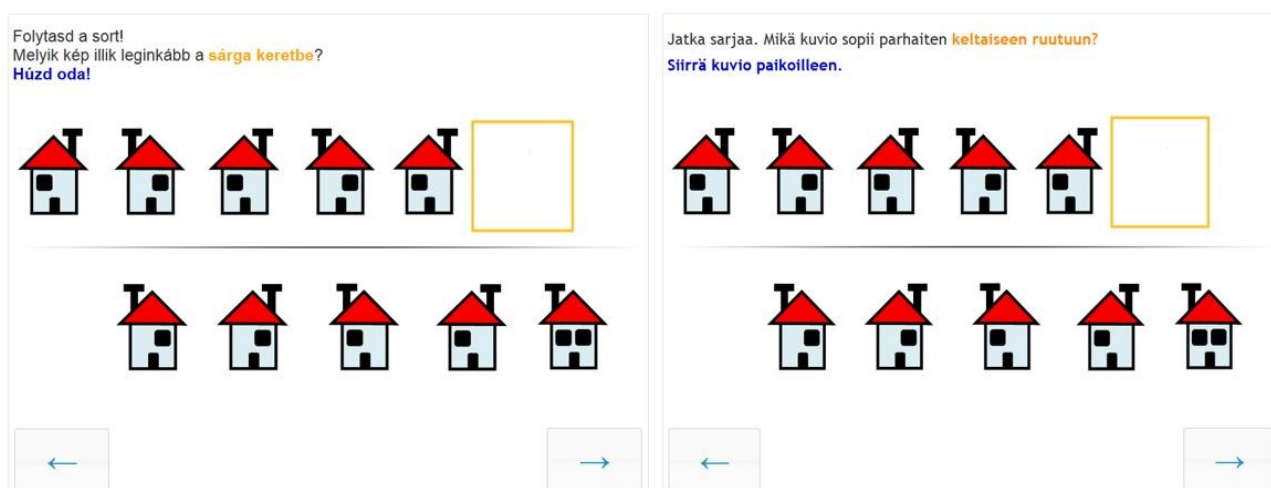
A felméréshez a korábban már több projektben (lásd például Csapó, 1998; Csapó et al., 2009; Molnár et al., 2013; Pásztor, 2016; Kambeyo & Wu, 2017; Pásztor, Molnár, Korom, B. Németh & Csapó, 2017; Wu & Molnár, 2018a, 2018b) használt induktív gondolkodás tesztet használtuk. A tesztet a konstruktum mérésre általánosan alkalmazott mérőeszközökkel összhangban dolgoztuk ki, amely négy résztesztet, numerikus analógiákat és sorozatokat, valamint figurális analógiákat és sorozatokat tartalmaz. A többféle tartalom és művelet növeli a mérőeszköz alkalmazása során nyert eredmények érvényességét. Az eddigi adatfelvételekből származó elemzések megmutatták, hogy a négy résztesztben szereplő különböző feladatformátumok az induktív gondolkodás

egymással szorosan összefüggő, de különböző aspektusait vizsgálják. A megerősítő faktorelemzés során az elméleti elvárásainknak leginkább megfelelő modellek illeszkedtek a legjobban az empirikus adatainkhoz, amelyben a négy részteszt külön látens faktorként jelent meg (Pásztor, 2016; Pásztor, Molnár, Korom, B. Németh & Csapó, 2017). Az egyes itemtípusokra egy-egy példát mutatunk be. Az 1. és 2. ábra numerikus, a 3. és 4. ábra figurális feladatokat szemléltet.

A jelen vizsgálatban szereplő négy tesztváltozat (FI4: finn, 4. évf.; FI7: finn, 7. évf.; HU4: magyar 4. évf.; HU7: magyar, 7. évf.) nem egyezik meg pontosan, de igen nagy átfedés van közöttük, így IRT skálázást alkalmazva igen nagy valószínűséggel közös skálán lehet az eredményeket kifejezni. A finn tesztek mindkét évfolyamon 54, a magyar tesztek 53 itemet tartalmaztak, 15 figurális sorozatot, 15 figurális analógiát (a magyar 14-et, 8 számanalógiát és 16 számsorozatot). A tesztek horgonyzásához mindegyik esetben 51 item állt rendelkezésre.



1. ábra. Példa figurális analógia feladatra



2. ábra. Példa figurális sorozat feladatra

Mi lehet a szabály?  
Melyik szám illik leginkább a sárga keretbe?  
Húzd oda!

Tunnista sääntö. Mikä numero sopii parhaiten keltaiseen ruutuun?  
Siirrä numero paikoilleen.

3. ábra. Példa számanalógia feladatra

Folytasd a sorozatot! Írd azokat a számokat a sárga keretekbe, melyek a legjobban odaillenek a számsor folytatásaként!

Jatka numerosarjaa. Siirrä keltaisiin ruutuihin ne numerot, joiden avulla numerosarja jatkuu oikein.

4. ábra. Példa számsorozat feladatra

### Eljárások

A Magyarországon megvalósuló méréseket az Oktatásméleti Kutatócsoport munkatársai koordinálták. Az egész országra kiterjedő partneriskolai hálózat intézményei előzetesen tájékoztatás kaptak a programban szereplő mérőeszközökről, a mérések lebonyolításához szükséges technikai feltételekről, valamint az eredmények eDia rendszerből való letöltésének módjáról. A programban való részvétel önkéntes volt. A regisztrációt követően az intézmények elküldték a mérésben részt vevő tanulók egyedi és anonim mérési azonosítóit, melyekkel a tanulók be tudtak lépni a feladatsorokba. A mérés objektivitásának biztosítása érdekében az adatfelvétel lebonyolításához mérési útmutató állt rendelkezésre, amely tartalmazta a tesztek elérhetőségét, a mérések kivitelezésének időintervallumát, és az adatfelvétel körülményeivel kapcsolatos instrukciókat. A mérések osztálytermi környezetben, az iskolák számítógépes termében valósultak meg az intézmények által biztosított mérési biztos felügyeletével, aki a technikai problémák kezelésén túl más segítséget nem nyújthatott a diákoknak. A diákok a tesztek befejezését követően azonnali visszajelzést kaptak a teljesítményükről. A pedagógusok az adatfelvétel lezárta után részletes, diákonként személyre szóló, szöveges és táblázatokkal, pókhálóábrákkal ellátott, objektív viszonyítási pontokat tartalmazó visszajelzést tölthettek le az eDia rendszerből. A diákonkénti



visszajelentés mellett az iskolaszintű, regionális, településtípus, valamint országos viszonyítási adatok mellett rendelkezésükre állt az osztály teszt és részteszt szintű átlagos teljesítménye is (Molnár & Csapó, 2019).

A finn méréseket a Helsinki Egyetem pedagógiai mérési központja koordinálta. A vizsgálatban részt vevő intézmények szintén tájékoztatást kaptak a vizsgálat céljairól, kereteiről. Az anonim tanulói mérési azonosítókat a finn kutatók előzetesen megküldték a szegedi kutatók számára, ezt követően feltöltésre kerültek az eDiarendszerbe. A mérések objektivitását a finn vizsgálatokban is mérési útmutató segítette, az adatfelvételt mérési biztosok felügyelték. A mérések a finn oktatási intézményekben, osztálytermi keretek között valósultak meg, az adatbeviteli eszközt ebben az esetben tabletek jelentették. A diákok számára az azonnali visszajelzés a finn mérések során is biztosított volt. Az adatfelvétel lezárta után az eredményeket a szegedi kutatók elküldték a Helsinki Egyetem pedagógiai mérési központja számára, a visszajelentéseket a mérési központ munkatársai készítették el és küldték meg az intézmények részére.

A finn és a magyar diákok különböző beviteli eszközökön oldották meg a feladatsorokat (tablet és asztali számítógép), ezért az eredmények értelmezése során felmerülhet a különböző médiumok befolyásoló hatása. A tableten és a képernyőn megjelenő ábrák azonosak voltak, a szövegdobozokban is megegyező módon jelentek meg a finn és a magyar szövegek. A rövid instrukciók szövegeinek méretét, valamint a képek méretét is kellőképpen nagyra lehetett állítani, hogy a feladatok jól láthatók legyenek a tableten történő megoldás során is. A lehetséges médiahatás mértékét tovább csökkentette, hogy a feladatok zárt végűek voltak, a válaszok megadáshoz nem volt szükség betűk vagy számok beírására.

## Eredmények

### *A mérőeszközök reliabilitása*

A méréshez felhasznált tesztek és az összehorgonyzáshoz használt résztesztek itemszámát és reliabilitásmutatóját a különböző mintákban a 2. táblázat tartalmazza. Az eredmények azt mutatják, hogy mind a négy teszt és az összeskálázásra használt 51 ítemes részteszt is kiemelkedően magas reliabilitásmutatóval rendelkezik.

Évfolyam	Finn (54 ítem)	Magyar (53 ítem)	Finn horgony (51 ítem)	Magyar horgony (51 ítem)
4.	0,941	0,936	0,937	0,933
7.	0,956	0,937	0,953	0,935

2. táblázat: Az induktív gondolkodás teszt reliabilitásmutatója (Cronbach alfa)

A reliabilitásmutatók között nem voltak lényeges különbségek, ami azt jelzi, hogy a tesztek mindkét országban egyaránt megbízhatóan működnek, a tesztek lefordítása, illetve másik médium (asztali gép vs. tablet) használata nem okoz a reliabilitásban eltérést.

### *A skálázás invarianciája*

A korábban bemutatott érvek alapján felmerül annak szükségessége, hogy megvizsgáljuk, a tesztek egyforma skálákat szolgáltatnak-e a különböző részmintákon. Ezt kétféle releváns összehasonlításban tesszük meg. Elő-

szőr megnézzük, különböznek-e a skálák a két országban, vagyis elemezzük a skála országok közötti invarianciáját mind a negyedik, mind a hetedik évfolyamon, majd megnézzük a nemek szerinti skálainvarianciát külön a finn és a magyar minta alapján.

Első lépésben a nagy mintaméretre érzékeny khi-négyzet tesztet (Cheung & Rensvold, 2002) végeztük el. Az eredmények értelmezése során szem előtt tartva a finn és a magyar minták méretének különbözőségét. A khi-négyzet értéke alapján az induktív gondolkodás teszt mérési invarianciájának tesztelése során mind 4., mind 7. évfolyamon a metrikus invariancia modell illeszkedése nem, de a részleges metrikus invariancia modell illeszkedése már azonos volt a konfigurális modell illeszkedési szintjével ( $\Delta CFI < 0,01$  és nem szignifikáns  $\chi^2$ -teszt). A legszigorúbb feltételek mellett azonban már szignifikáns változás történt az illeszkedés fokában (3. és 4. táblázat). Ennek értelmében mindkét tesztelési környezetben a diákok ugyanúgy konceptualizálták a mért konstruktumot, a mérőeszköz ugyanannak a látens változónak a mérését valósította meg mindkét csoportban. A tesztelési (nyelvi és technológiai) környezet nem befolyásolta a diákok relatív helyzetét a mintában, azaz bármely két diák egymáshoz való viszonya független volt attól, hogy a két diák a tesztet a finn tesztelési környezetben vagy itthon oldotta-e meg (részleges metrikus invariancia).

Modell	$\chi^2$	df	$\Delta\chi$	$\Delta df$	p	CFI	TLI	RMSEA	RMSEA CI
(1) Konfigurális invariancia	2302,5	1170				0,988	0,987	0,020	0,019-0,021
(2) Metrikus invariancia	2331,0	1202	72,7	32	<0,01	0,988	0,987	0,020	0,019-0,021
(2.1) Részleges metrikus invariancia	2289,1	1199	48,4	29	>0,01	0,988	0,988	0,019	0,018-0,021
(3) Részleges szigorú metrikus invariancia	2214	1231				0,990	0,989	0,018	0,017-0,019

3. táblázat: A 4. évfolyamos finn és magyar tesztek skálainvarianciája

Modell	$\chi^2$	df	$\Delta\chi$	$\Delta df$	p	CFI	TLI	RMSEA	RMSEA CI
(1) Konfigurális invariancia	1936,6	1171				0,983	0,982	0,023	0,021-0,025
(2) Metrikus invariancia	1990,7	1202	66,3	31	<0,01	0,983	0,982	0,023	0,021-0,025
(2.1) Részleges metrikus invariancia	1963,4	1200	47,0	29	>0,01	0,984	0,983	0,023	0,021-0,024
(3) Részleges szigorú metrikus invariancia	2007,1	1232				0,983	0,983	0,023	0,021-0,024

4. táblázat: A 7. évfolyamos finn és magyar tesztek skálainvarianciája

A finn mintában a nemek szerinti skálainvariancia tesztelése során mindhárom egymásba ágyazott, egyre szigorúbb feltételeknek megfelelő modell jó illeszkedésű volt (CFI, TLI > 0,95; RMSEA < 0,06; l. 5. táblázat). A különböző feltételek meglétét tesztelő modellek illeszkedése alapvetően nem romlott szignifikánsan (szigorú metrikus invariancia esetén igen, de a részleges szigorú metrikus invariancia esetén nem;  $\Delta CFI < 0,01$  és a  $\chi^2$ -

különbség teszt nem szignifikáns), azaz a fiúk és lányok teljesítménye összehasonlítható, mind a fiú, mind a lányok esetén azonos konstruktum mérésére került sor. Mind a különböző nemű diákok relatív helyzete, mind teljesítményük, teszten elért eredményük, képességszintjük közvetlenül (manifeszt szinten is) összehasonlítható.

Modell	$\chi^2$	df	$\Delta\chi$	$\Delta df$	p	CFI	TLI	RMSEA	RMSEA CI
(1) Konfigurális invariancia	1675,4	1171				0,985	0,983	0,025	0,022-0,028
(2) Metrikus invariancia	1706,6	1203	44,9	32	>0,01	0,985	0,984	0,025	0,022-0,027
(3) Szigorú metrikus invariancia	1762,2	1239	81,6	36	<0,01	0,984	0,984	0,025	0,022-0,026
(3.1.) Részleges szigorú metrikus invariancia	1673,2	1233	45,7	30	>0,01	0,987	0,986	0,023	0,020-0,026

5. táblázat. Nemek szerinti skálainvariancia a finn minta alapján

A nemek szerinti skálainvariancia tesztelése eltérő eredményt hozott a hazai mintában (6. táblázat), miután a legszigorúbb feltételeknek megfelelő szigorú metrikus invariancia illeszkedése jelentősen különbözött a kevésbé szigorú feltételeket állító modellek illeszkedésétől. A konfigurális modell illeszkedése azonos volt a két mintán, azaz a mérési modell struktúrája invariáns a csoportok között, a fiúk és a lányok ugyanúgy konceptualizálják a mért konstruktumot, ugyanazt a fogalmi keretet használják a teszt itemeinek megválaszolása során, az egyes vizsgált változók ugyanazt a látens változót mérik mindkét csoportban, a fiúk és a lányok esetén azonos konstruktum mérésére került sor. Miután a részleges metrikus invariancia modell modellilleszkedése már azonos volt a konfigurális modell illeszkedési szintjével, megállapítható, hogy a fiúk és lányok relatív helyzete összehasonlítható.

Modell	$\chi^2$	df	$\Delta\chi$	$\Delta df$	p	CFI	TLI	RMSEA	RMSEA CI
(1) Konfigurális invariancia	2573,8	1171				0,984	0,982	0,024	0,023-0,025
(2) Metrikus invariancia	2700,6	1203	121,7	32	<0,01	0,983	0,982	0,024	0,023-0,025
(2.1) Részleges metrikus invariancia	2572,7	1195	38,7	24	>0,01	0,984	0,983	0,023	0,022-0,025

6. táblázat: Nemek szerinti skálainvariancia a magyar minta alapján

Második lépésben – a mintaméretből eredő különbségek miatt – az invarianciatesztelések alapját a mintaméretre kevésbé érzékeny becslési hiba négyzetes átlagának gyökét (az RMSEA mutatót) alapul vevő számításokat végeztük el és interpretáltuk. A vizsgált modellek illeszkedése abban az esetben jó, akkor beszélhetünk jó illeszkedésről, ha az RMSEA értéke kisebb/egyenlő 0,05 (Kline, 2011, Browne és Cudeck, 1993), valamint a konfidencia intervallum alsó és felső értéke is az említett tartományban van. Ezzel a módszerrel megállapítha-

tó, hogy az invariancia meglétének teszteléséhez szükséges egyre szigorúbb feltételeknek megfelelő modellek illeszkedése minden esetben jó (RMSEA illeszkedésindex értéke kisebb, mint 0,05), azaz minden egyes vizsgált esetben teljesül az összehasonlíthatóság feltétele.

### *A teszteken elért teljesítmények összehasonlítása*

A két országban a mintavétel különböző volt és egyik sem tekinthető olyan reprezentatív mintának, ami alapján az átlagos teljesítmények összehasonlításából messzemenő következtetéseket vonhatnánk le a finn-magyar különbségekre vonatkozóan. Mindenesetre, figyelembe véve a mintákkal kapcsolatos megjegyzéseket, érdekes megfigyeléseket tehetünk. Az országokon belüli részminták összehasonlításával kapcsolatban pedig nincsenek a mintavétel különbözőségeiből származó aggályok.

A különböző résztesztek, illetve a teljes teszt átlagát, szórását, továbbá a két évfolyam finn-magyar összehasonlítására vonatkozó t-próbáját a 7. táblázat foglalja össze. A két ország adatainak összehasonlításából kiderül, hogy a magyar diákok a negyedikes figurális sorozatok részteszt kivételével minden esetben jobban teljesítettek, mint társaik. Ez az eredmény, figyelembe véve, hogy a finn diákok minden PISA teszten jobban teljesítettek, mint a magyarok, váratlan és nehezen értelmezhető. Különösen nehéz magyarázatot találni a jobb magyar teljesítményekre annak fényében, hogy a PISA 2012 dinamikus problémamegoldás teszten, aminek a megoldása igényli a szabályindukciót, a finn diákok sokkal jobban teljesítettek, mint a magyarok (OECD, 2014).

Ország		FI	FI	HU	HU	t	p	t	p
Évfolyam		4.	7.	4.	7.	4.	4.	7.	7.
Figurális itemek	átlag	61,17	68,04	64,45	74,99	3,201	0,001	7,24	0,001
	szórás	25,44	25,93	22,36	19,84				
Numerikus itemek	átlag	33,68	45,66	50,68	62,05	16,095	0,001	14,714	0,001
	szórás	22,28	27,86	24,08	24,17				
Figurális sorozatok	átlag	59,51	66,98	61,16	72,97	1,486	n.s.	6,124	0,001
	szórás	26,91	26,39	24,37	20,22				
Figurális analógiák	átlag	62,96	69,18	67,99	77,16	4,53	0,001	7,395	0,001
	szórás	27,71	28,86	24,23	22,48				
Numerikus sorozatok	átlag	35,27	45,18	48,41	58,12	16,551	0,001	15,313	0,001
	szórás	25,61	30,19	28,45	28,60				
Numerikus analógiák	átlag	30,49	46,62	43,33	59,26	10,903	0,001	10,237	0,001
	szórás	22,39	29,4	27,39	27,46				
Teljes teszt	átlag	48,23	57,51	57,97	68,90	10,142	0,001	11,904	0,001
	szórás	22,00	24,68	21,45	20,39				

7. táblázat: A horgony itemekre (51 item) vonatkozó teljesítménykülönbségek (%)

A különbségek mértékét könnyebben értelmezhetjük, ha mindegyiket a *Cohen-féle d* hatásméretben (a különbség elosztva az átlagos szórással) fejezzük ki. Ezeket az adatokat (a 7. táblázat adatai alapján) a 8. táblázat

foglalja össze. Itt látjuk most már egymással is összehasonlítható formában a finn–magyar különbségeket, amiből kiderül, hogy a magyar diákok igazi erőssége inkább a numerikus területen jelentkezik. Mind a negyedik, mind a hetedik évfolyamon a numerikus itemekben volt a nagyobb különbség, míg a figurális sorozatok különbsége a negyedik évfolyamon nem is volt szignifikáns.

d Részteszt	d (FI-HU)		d (4-7)	
	4.	7.	HU	FI
Figurális itemek	0,140	0,253	0,498	0,267
Numerikus itemek	0,759	0,634	0,470	0,475
Figurális sorozatok	0,073	0,224	0,527	0,280
Figurális analógiák	0,215	0,282	0,391	0,219
Numerikus sorozatok	0,738	0,652	0,351	0,354
Numerikus analógiák	0,468	0,483	0,580	0,617

8. táblázat: Az évfolyamok és az országok közötti különbségek horgony itemekre (51 item) vonatkozó teljesítménykülönbségek (%)

Ugyancsak a 8. táblázatban látható a két évfolyam közötti különbség szintén  $d$  értékben kifejezve. A figurális itemek tekintetében a magyar diákoknál nagyobb fejlődést jeleznek a tesztek, mint a finneké, míg numerikus területen kiegyenlítettebb a változás. Azt látjuk tehát, hogy a numerikus feladatok által mért készségek tekintetében a magyar diákok gyorsabban fejlődnek, mint a finnek, és mivel itt nem jelentkeznek a mintavétel különbözőségeiből fakadó aggályok, elfogadhatjuk, hogy itt valóban gyorsabb fejlődésről van szó.

A legtöbb kognitív területen a fiúk és a lányok teljesítménye között statisztikailag kimutatható különbség van. Ennek egyik legismertebb példája, hogy szövegértésben a PISA felmérések szerint a lányok szinte minden országban jobban teljesítenek a fiúknál. Például a 2015-ös PISA felmérésben a finn lányok 47 ponttal, a magyar lányok pedig 25 ponttal jobban teljesítettek, mint a fiúk. Matematikából Finnországban a fiúk 7 ponttal jobban, Magyarországon 8 ponttal gyengébben teljesítettek, mint a lányok, de itt egyik különbség sem szignifikáns (OECD, 2016).

Mindezek alapján indokolt a felmérésünk adatai alapján is megvizsgálni a fiúk és lányok teljesítményei közötti különbségeket. Az eredmények azt mutatják, hogy a teljes teszt szintjén negyedik évfolyamon egyik országban sincsenek szignifikáns különbségek, míg hetedik évfolyamon a lányok mindkét országban jobban teljesítenek (9. táblázat). A résztesztek szintjén már változatosabb képet látunk, negyedik évfolyamon a magyar fiúk kissé jobbak a lányoknál, a többi eredmény pedig azt mutatja, hogy a lányok előnye erősebben a figurális tartalmú feladatokban mutatkozik meg.

Évf.	Terület	Nem	HU			FI		
			Átlag	Szórás	t, p	Átlag	Szórás	t, p
4.	Teljes teszt	fiú	57,7	22,6	n.s.	48,2	23,2	n.s.
		lány	58,2	20,2		48,8	20,2	
	Numerikus analógiák	fiú	44,2	28,1	n.s.	32,5	24,7	2,040, p<0,05
		lány	42,4	26,5		28,9	19,6	
	Numerikus analógiák	fiú	55,5	26,4	2,219, p<0,05	35,9	26,8	n.s.
		lány	53,2	25,4		35,0	24,3	
	Figurális analógiák	fiú	66,6	26,0	-2,747, p<0,01	62,3	29,0	n.s.
		lány	69,3	22,1		64,3	25,6	
	Figurális sorozatok	fiú	59,7	25,4	-2,958, p<0,01	58,1	28,0	n.s.
		lány	62,6	23,1		61,5	25,1	
Teljes teszt	fiú	67,8	21,6	-2,107, p<0,05	55,2	26,2	-2,118, p<0,01	
	lány	69,9	19,0		60,1	22,4		
7.	Numerikus analógiák	fiú	59,2	29,1	n.s.	45,6	31,7	n.s.
		lány	59,3	25,8		48,0	26,6	
	Numerikus analógiák	fiú	62,7	26,4	n.s.	43,0	31,6	-2,154, p<0,05
		lány	64,0	24,9		47,8	28,2	
	Figurális analógiák	fiú	75,4	24,2	-3,155, p<0,01	65,8	30,6	-3,381, p<0,01
		lány	78,8	20,5		72,9	26,1	
	Figurális sorozatok.	fiú	71,5	21,2	-2,882, p<0,01	64,9	27,5	-2,283, p<0,05
		lány	74,3	19,13		69,3	24,62	

9. táblázat. Nemek szerinti különbségek az induktív gondolkodás teljesítményekben

Amint az a nemzetközi összehasonlító adatokból is ismeretes, a magyar iskolai osztályok létszáma a közepesnek felel meg, nagyjából megegyezik az OECD átlaggal (OECD, 2018. D2.1. táblázat). A vizsgálatunkba bevont minta adatai szerint a finn osztályok létszámai mindkét életkorban magasabbak a megfelelő magyar átlagoknál. A nagyjából két fő osztályonkénti különbség pedagógia szempontból nem jelentős, a kisebb osztálylétszám viszont a rendszer szintjén közel 10%-nyi többlet erőforrást igényel, ami felveti a rendelkezésre álló források hatékony felhasználásának kérdését (OECD, 2019).

Ország, évfolyam	átlag	szórás	relatív szórás (%)	r (p<0,001)
Finn 4. évf.	22,4	3,5	15,6	0,149
Magyar 4. évf.	20,7	5,3	25,6	0,275
Finn 7. évf.	20,6	3,6	17,4	0,217
Magyar 7. évf.	18,6	5,5	29,6	0,188

10. táblázat. Az osztálylétszám és az induktív gondolkodás kapcsolata

Hasonló kérdést vet fel, hogy negyedikben mindegyik országban magasabbak a létszámok, mint hetedikben. Ennek sokféle (pl. településszerkezeti, demográfiai) oka lehet, de érdemes megjegyezni, hogy pedagógiai szempontból fordítva lenne indokolt. Látjuk, hogy a magyar adatok szórása nagyobb, mint a finneké, ami jelzi, hogy nálunk osztályméret tekintetében is nagyobbak a különbségek. Feltehetően a kisebb települések kisebb iskolái miatt.

Ismeretes a szakirodalomból az is, hogy a nagyobb létszámú osztályokban általában jobbak a teljesítmények. E vizsgálat is azt mutatja, hogy az osztályméret mindegyik esetben pozitívan korrelál az induktív gondolkodás fejlettségi szintjével. A legerősebb korrelációt a magyar negyedik évfolyamon láttuk, a leggyengébbet a finn negyediken. Hetedikre megfordulnak a viszonyok, és a különbség is kisebb.

## Az eredmények értékelése

Elemzésünk egyik központi kérdése az volt, hogy a magyar kutatócsoport által kifejlesztett online induktív gondolkodás tesztek át lehet-e ültetni egy másik nyelvre, és másik informatikai eszközre úgy, hogy közben a mérési skála ne változzon lényegesen, azaz a mérőeszközöket a két országban ekvivalens módon lehessen használni. A mérőeszközök jóságának vizsgálatára egyrészt a klasszikus tesztelmélet megközelítését alkalmaztuk, és azt találtuk, hogy a teszt reliabilitásmutatói az átültetés és új kontextusban való alkalmazás során érdemben nem változtak.

Másrészt a megerősítő faktoranalízis eszközrendszerét alkalmazva megvizsgáltuk, hogy a mérési skálák a transzformáció során invariánsak maradnak-e. Az elemzések eredményei szerint a mérések kielégítették a konfigurális invariancia követelményeit, vagyis a tesztek mindkét országban ugyanazt a konstruktumot, ugyanazt a látens változót mérték, továbbá a diákok egymáshoz viszonyított rangsorának megállapítására alkalmasak voltak. A szigorú metrikus invariancia követelményeinek a két országban végzett tesztelés már nem felelt meg. További vizsgálatokat igényel annak megállapítása, hogy ennek oka a két nyelv és kultúra különbségeiben keresendő, vagy pedig abban, hogy az online tesztek megoldásához a két országban más technikai eszközöket használtunk. A fiúk és lányok esetében a mérőeszköz azonos módon működött, ami akkor is fennáll, ha az invariancia tekintetében a szigorúbb feltételekkel számolunk.

A különböző teljesítmények összehasonlítása során azt találtuk, hogy a magyar diákok jobb eredményt értek el, mint a finnek. Ez az eredmény váratlan, és a minták nagy mérete miatt mérési hibának kicsi a valószínűsége. Mivel a magyar diákok nem csak jobban teljesítenek, hanem a negyedik és a hetedik évfolyam között nagyobb a különbség, ami az intenzívebb fejlődésre utal, az eredmények figyelemre méltóak. Ugyanakkor a rendelkezésre álló adatok alapján a magyar diákok jobb eredményeire nem lehet érdemi magyarázatot találni. Egy ugyancsak a Vantaa-projektből származó másik elemzés szerint, amelyik az óvodásokat hasonlította össze, és

amelyben a magyar mintát Szeged város óvodáiból választottuk, egy induktív gondolkodás teszten még a finn gyerekek teljesítettek jobban (Hotulainen, Kupiainen, Pásztor, Molnár, & Csapó, 2018). Idősebb korban, a PISA mérések szerint ugyancsak a finn diákok eredményei jobbak. Mindezek alapján csak hipotézisként fogalmazhatjuk meg, hogy a magyar diákok jobb teljesítménye az online mérésekkel való gyakoribb találkozásnak köszönhető. Amint a vizsgálat kontextusának bemutatása során jeleztük, a magyar diákok nem egyszerűen csak több online mérésben vettek részt, hanem a mérések tartalma is a gondolkodásra irányult. Ezáltal mind a diákok, mind a pedagógusok fegyelmét felhívtuk a tanulás egy olyan dimenziójára, ami a hétköznapi iskolai munka során többnyire láthatatlan marad. A gondolkodási feladatokkal és azok méréseinek eredményeivel való gyakori szembesülés tudatosabbá tehetette a pedagógusok képességfejlesztő munkáját.

Néhány további összehasonlítás az irodalomból ismert adatokkal összhangban levő eredményt mutatott. A fiú-lány teljesítmények összehasonlításának eredménye illeszkedik abba a trendbe, miszerint a fejlett országokban a legtöbb területen a lányok teljesítménye gyorsabban javul, mint a fiúké, egyre több területen előzik meg a fiúkat. Méréseink eredményeiben szerepet játszhat, hogy a serdülőkor környékén a lányok biológiai, szociális és intellektuálisabb fejlődése gyorsabb, mint a fiúké, de az is lehet, hogy a lányok többet profitálnak az iskolai oktatásból.

Eredményeink ugyancsak megerősítették azt a korábbi tapasztalatot, miszerint pozitív korreláció van az osztálylétszám és az iskolai eredmények között. Ezt most egy olyan területen is megfigyelhettük, ahol a mérés tartalma nem kapcsolódik közvetlenül az iskolai tanuláshoz. Ennek azonban Magyarországon nem feltétlenül a nagyobb osztályokban folyó hatékonyabb pedagógiai munka lehet az oka, hanem inkább az, hogy a népszerűbb iskolákba a jobb érdekérvényesítő képességgel rendelkező szülők hatékonyabban be tudják juttatni a gyermekeiket. Ez utóbbi érvet gyengíti, hogy Finnországban is fennáll a korreláció, ahol az iskolai szelekció minimális mértékű.

Ahogy az előzőekben jeleztük, vizsgálatunk alapján számos érdekes kutatási kérdést lehet megfogalmazni. A Vantaa követéses program folyamatban van, a keletkező adatok további elemzésekre adnak lehetőséget. Az induktív gondolkodás fejlődésében tapasztalható finn–magyar teljesítményeket később, amikor már a finn diákok is huzamosabb ideje használják a tabletjeiket (lényegében ez már megvalósult, ma már nagyobb gyakorlatot szereztek az IKT eszközökkel való tanulásban, mint magyar kortársaik), az összehasonlításokat újra el lehet végezni. Emellett a diagnosztikus mérések gyakorlati hatásának felbecsülése szempontjából a finn–magyar összehasonlító elemzések számos további lehetőséget is nyújtanak, melyek kihasználására a későbbiekben kerül sor.

### *Köszönetnyilvánítás*

A tanulmány megírása alatt Pásztor Attila Nemzet Fiatal Tehetségeiért Ösztöndíjban részesült (NTP-NFTÖ-18-B-0100).



## Irodalom

---

1. Brown, T. (2006). CFA with equality constraints, multiple groups, and mean structures. *Confirmatory Factor Analysis for Applied Research*, New York: The Guilford Press.
2. Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York: Routledge.
3. Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. H. (2009). Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. *Quality & Quantity*, 43(4), 599–616.
4. Resing, W. C., Tunteler, E., & Elliott, J. G. (2015). The effect of dynamic testing with electronic prompts and scaffolds on children's inductive reasoning: A Microgenetic Study. *Journal of Cognitive Education and Psychology*, 14(2), 231–251.
5. Arce-Ferrer, A. J., & Guzmán, E. M. (2009). Studying the equivalence of computer-delivered and paper-based administrations of the Raven standard progressive matrices test. *Educational and psychological measurement*, 69(5), 855–867.
6. Blunch, N. L. (2008). *Introduction to structural equation modeling: Using SPSS and AMOS*. London: Sage.
7. Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. D. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
8. Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, 20(4), 872–882.
9. Byrne, B. M., & Stewart, S. M. (2006). Teacher's corner: The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*, 13(2), 287–321.
10. Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233–255.
11. Csapó, B. (1994). Az inductív gondolkodás fejlődése. *Magyar Pedagógia*, 94(1–2), 53–80.
12. Csapó, B. (1997). Development of inductive reasoning: Cross-sectional measurements in an educational context. *International Journal of Behavioural Development*, 20(4), 609–626.
13. Csapó, B. (1998). Az új tudás képződésének eszközei: az inductív gondolkodás. In Csapó, B. (Ed.). *Az iskolai tudás* (pp. 251–280). Budapest: Osiris Kiadó.
14. Csapó, B. (2001). Az inductív gondolkodás fejlődésének elemzése országos reprezentatív felmérés alapján. *Magyar Pedagógia*, 101(3), 373–391.
15. Csapó, B., & Molnár, G. (2017). Potential for assessing dynamic problem-solving at the beginning of higher education studies. *Frontiers in Psychology*, 8, 2022. doi: 10.3389/fpsyg.2017.02022
16. Csapó, B., & Molnár, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: The eDia System. *Frontiers in Psychology*, 10, 1522. doi: 10.3389/fpsyg.2019.01522
17. Csapó, B., Molnár, G., & Nagy, J. (2014). Computer-based assessment of school-readiness and reasoning skills. *Journal of Educational Psychology*, 106(2), 639–650. doi: 10.1037/a0035756
18. Csapó, B., Molnár, G., & R. Tóth, K. (2009). Comparing paper-and-pencil and online assessment of reasoning skills. A pilot study for introducing electronic testing in large-scale assessment in Hungary. In F. Scheuermann & J. Björnsson (Eds.), *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing* (pp. 120–125). Luxemburg: Office for Official Publications of the European Communities.
19. Davidov, E., Schmidt, B., & Billiet, J. (2011). *Cross-cultural analysis: Methods and applications*. New York: Psychology Press.

20. Feeney, A. (20178). Forty years of progress on category-based inductive reasoning. In L. J. Ball & V. A. Thompson (Eds.), *International handbook of thinking and reasoning* (pp. 167–185). New York: Routledge.
21. Fisher, A. V. (2015). Development of inductive generalization. *Child Development Perspectives*, 9(3), 172–177.
22. Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., & Csapó, B. (2013). Complex problem solving in educational contexts – Something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology*, 105(2), 364–379.
23. Hautamäki, J., Arinen, P., Eronen, S., Hautamäki, A., Kupianien, S., Lindblom, B., Niemivirta, M., Pakaslahti, L., Rantanen, P., & Scheinin, P. (2002). *Assessing Learning-to-Learn: A framework*. Helsinki: Centre for Educational Assessment, Helsinki University.
24. Hotulainen, R., Kupiainen, S., Pásztor, A., Molnár, G., & Csapó, B. (2018). Entering school with equal skills? A two-country comparison of early inductive reasoning. Paper presented at the Conference of European Association for Research on Learning and Instruction, Special Interest Group 1. Helsinki.
25. Hülber, L., & Molnár, G. (2013). Papír és számítógép alapú tesztelés nagymintás összehasonlító vizsgálata matematika területén, 1–6. évfolyamon. *Magyar Pedagógia*, 113(4), 243–263.
26. Kambeyo, L., & Wu, H. (2017). A gondolkodási képességek számítógép-alapú mérésének lehetőségei Namíbiában és Kínában. In G. Hunyady, B. Csapó, G. Pusztai & J. Szivák, (Eds.), *Az oktatás korproblémái* (pp. 179–189). Budapest: ELTE Eötvös Kiadó.
27. Klauer, K. J., & Phye, G. D. (2008). Inductive reasoning: a training approach. *Review of Educational Research*, 78(1), 85–123.
28. Klauer, K. J. (1989). *Denktraining für Kinder I*. Göttingen: Hogrefe, Verlag für Psychologie.
29. Klauer, K. J. (1991). *Denktraining für Kinder II*. Göttingen: Hogrefe, Verlag für Psychologie.
30. Klauer, K. J. (1993). *Denktraining für Jugendliche*. Göttingen: Hogrefe, Verlag für Psychologie.
31. Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York: Guilford Press.
32. Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543.
33. Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111–130.
34. Molnár, G. (2011). Playful fostering of 6- to 8-year-old students' inductive reasoning. *Thinking skills and Creativity*, 6(2), 91–99.
35. Molnár, G. (2015). A képességmérés dilemmái: a diagnosztikus mérések (eDia) szerepe és helye a magyar közoktatásban. *Géniusz Műhely Kiadványok*, 2, 16–29.
36. Molnár, G., & Csapó, B. (2011). Az 1–11 évfolyamot átfogó induktív gondolkodás kompetenciaskála készítése a valószínűségi tesztelmélet alkalmazásával. *Magyar Pedagógia*, 111(2), 127–140.
37. Molnár, G., & Csapó, B. (2017). Assessment-based, personalised learning in primary education. In J. C. Spender, T. Gavrilova, & G. Schiuma (Eds.), *Knowledge management in the 21st century: Resilience, creativity and co-creation. Proceedings IFKAD2017* (pp. 443–449). St. Petersburg: St Petersburg University.
38. Molnár, G., & Csapó, B. (2019). A diagnosztikus mérési rendszer technológiai keretei: Az eDia online platform. *Iskolakultúra*, 29(4–5), 16–32.
39. Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: relations and development. *Thinking Skills and Creativity*, 9, 35–45.
40. Molnár, G., Makay G., & Ancsin G. (2018). *Feladat- és teszt szerkesztés az eDia rendszerben*. Szeged: SZTE Oktatásméleti Kutatócsoport.
41. Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide*. Los Angeles, CA: Author.
42. OECD (2014). *PISA 2012 results: Creative problem solving. Students' skills in tackling real-life problems (Volume V)*. Paris: OECD.

43. OECD (2016). *PISA 2015 results: Excellence and equity in education (Volume I)*. Paris: OECD.
44. OECD (2018). *Education at a glance 2018*. Paris: OECD.
45. OECD (2019). How much would it cost to reduce class size by one student? *Education Indicators in Focus*, 66.
46. Pásztor, A. (2014). Lehetőségek és kihívások a digitális játék alapú tanulásban: egy induktív gondolkodást fejlesztő tréning hatásvizsgálata. *Magyar Pedagógia*, 114(4), 281–301
47. Pásztor, A. (2016). *Az induktív gondolkodás technológia alapú mérése és fejlesztése*. Ph.D. disszertáció. Szeged: SZTE BTK Neveléstudományi Doktori Iskola.
48. Pásztor, A., Molnár, Gy., Korom, E., B. Németh, M., & Csapó, B. (2017). Online assessment of inductive reasoning and its predictive power on inquiry skills in science. 17th biennial conference of the European Association for Research on Learning and Instruction (EARLI). p. 509. Tampere, Finland.
49. Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cognitive Psychology*, 41(1), 1–48.
50. Schroeders, U., & Wilhelm, O. (2011). Equivalence of reading and listening comprehension across test media. *Educational and Psychological Measurement*, 71(5), 849–869.
51. Vainikainen, M. P. (2014). *Finnish primary school pupil's performance in learning to learn assessments: A longitudinal perspective on educational equity*. Helsinki: University of Helsinki.
52. Vainikainen, M. P., Hautamäki, J., Hotulainen, R., & Kupiainen, S. (2015). General and specific thinking skills and schooling: Preparing the mind to new learning. *Thinking Skills and Creativity*, 18, 53–64.
53. Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70.
54. Wang, S., Jiao, H., Young, M. J., Brooks, T. E., & Olson, J. (2008). Comparability of computer-based and paper-and-pencil testing in K-12 assessment: A meta-analysis of testing mode effects. *Educational and Psychological Measurement*, 68(1), 5–24.
55. Williams, J. E., & McCord, D. M. (2006). Equivalence of standard and computerized versions of the Raven Progressive Matrices Test. *Computers in Human Behavior*, 22(5), 791–800.
56. Wu, H., & Molnár, G. (2018a). Interactive problem solving: Assessment and relations to combinatorial and inductive reasoning. *Journal of Psychological and Educational Research*, 26(1), 90–105.
57. Wu, H., & Molnár G. (2018b). Computer-based Assessment of Chinese Students' Component Skills of Problem Solving: A Pilot Study. *International Journal of Information and Education Technology*, 8(5), 381–356.

## *Comparative examination of the development of inductive reasoning: Online assessments in Hungary and Finland*

---

This study presents the result of a comparative assessment of inductive reasoning among Finnish and Hungarian students. It investigates if the instruments developed in Hungary work in the same way in different age groups in Finland, in another country with different pedagogical culture; how students' inductive reasoning develops and how boys' and girls' achievement differ in the two educational systems. The participants of the study were fourth and seventh grade students from Vantaa city and from the partner schools of the Research Group of Learning and Instruction, University of Szeged. Although the samples cannot be regarded as representative, thus the achievements cannot be directly compared, the data are suitable to answer the re-search questions. The instrument was an inductive reasoning test (figural and numeric analogies and series) previously used in several former studies administered by the eDia online testing platform. The Finnish students solved the tasks on tablets, while the Hungarian students used keyboards and mice. The results showed that the reliability of the tests did not change by transplanting them into Finnish. The scale variance analyses indicated that the test used in different environments met the requirements of configural and partial invariance, but not the metric invariance. It was an unexpected result that in both age groups the Hungarian students achieved better than the Finnish students, especially in numeric tasks. As the samples of this assessment were not representative, further studies are required to confirm the observed difference and to explore its causes.

**Keywords:** inductive reasoning, technology-based assessment, eDia, comparative study