

A KRIPPENDORFF-ALFA (KALPHA) ALKALMAZÁSA A GYAKORLATBAN: KETTŐNÉL TÖBB KÓDOLÓ KÖZÖTTI EGYETÉRTÉS VIZSGÁLATA DICHOTÓM VÁLTOZÓK ESETÉBEN¹



KESZEI Barbara

ELTE Pszichológiai Intézet, Pszichológiai Doktori Iskola
BGE KVIK Pedagógiai Intézet
keszei.barbara@ppk.elte.hu

BÖDDI Zsófia

ELTE TÓK Neveléstudományi Tanszék
boddi.zsofia@tok.elte.hu

GESZTEN Dalma

BME GTK Ergonómia és Pszichológia Tanszék
gesztend@erg.bme.hu

HÁMORNIK Balázs Péter

BME GTK Ergonómia és Pszichológia Tanszék
hamornik@erg.bme.hu

DÚLL Andrea

ELTE Pszichológiai Intézet
BME GTK Szociológia és Kommunikáció Tanszék
dull.andrea@ppk.elte.hu

ÖSSZEFOGLALÓ

Háttér, célkitűzések: A kvalitatív vizsgálatok az 1990-es évek óta egyre elfogadottabb és megbecsült részét képezik a pszichológiai kutatásoknak. A folyamatot tovább erősíthetjük, ha a kvalitatív és kvantitatív módszereket egymást kiegészítve, segítve használjuk a kuta-

¹ A tanulmány írása az ELTE Felsőoktatási Intézményi Kiválósági Program (1783-3/2018/FEKUTSRAT) keretében valósult meg az Emberi Erőforrások Minisztériuma támogatásával.

tásban. Ennek egyik módja lehet, ha a kódoláson alapuló kutatások reliabilitását elősegítjük, ellenőrizhetővé tesszük, például a kódolók közötti megbízhatóság kiszámításával és közlésével. A kategóriarendszereket kidolgozó és azokat tesztelő kutatók egyik kevésbé publikált témája a reliabilitás problémája több kódoló és dichotóm kategóriák esetében. Az ebben az esetben alkalmazandó statisztikai eljárás nem tartozik a kutatók alapvető statisztikai eszköztárába. Jelen vizsgálatban – példaként – célunk egy olyan megbízható kategóriarendszer kidolgozása volt, amellyel óvodás gyermekek interakciói válnak megbízhatóan leírhatóvá.

Módszer: Óvodai csoportszobákban természetes megfigyelést végeztünk, a gyerekek viselkedéséről videófelvételeket készítettünk. A kidolgozott kódrendszer célja a gyermekek tevékenységének, illetve azok társas szintjének és az interakciók jellemzőinek vizsgálata volt. Az elsődleges kódrendszert 8 független kódoló használta, amit egy 53 perces minta-videó-anyagon teszteltek.

Eredmények: A 8 kódoló közötti egyetértés kiszámítására a Krippendorff-alfát (KALPHA) alkalmaztuk. Ez jelenleg az egyetlen SPSS-be is könnyen importálható statisztikai eljárás, ami lehetővé teszi, hogy kettőnél több kódoló esetében dichotóm, binomiális változóknál is tudjunk megbízhatóságot számolni. A végleges kategóriarendszer kialakításához a KALPHA-értékek vizsgálata jelentős mértékben hozzájárult: ezek alapján lehetett megbízhatóan eldönteni, hogy mely kategóriákat érdemes összevonni vagy akár kihagyni a kategóriarendszerekből. Az így kialakított kódrendszert két független kódoló sikeresen használta, mint azt a kiszámított Cohen-kappa-értékek igazolták.

Következtetések: A Krippendorff-alfa (KALPHA) a magyar kutatási és statisztikai gyakorlatba könnyen beilleszthető. A használata a kódolók közötti egyetértés vizsgálatának hiánypótló eleme.

Kulcsszavak: reliabilitás, megfigyelők közötti egyetértés, KALPHA, dichotóm változók, statisztikaipróba-kódolás

BEVEZETÉS

A kutatómódszertan kétféle, gyakran talán mesterségesen is egymással szembe állított – kvalitatív és kvantitatív – megközelítése egyaránt törekszik a megbízható és érvényes vizsgálatok lebonyolítására. Ugyanakkor kijelenthetjük, hogy abban már nagy eltérések vannak, hogy a megbízhatóság és az érvényesség alatt melyik kutató, irányzat vagy paradigma pontosan mit ért (pl. Golafshani, 2003). Tanulmányunk elsősorban azoknak kíván segítséget nyújtani, akik a kutatásokban az adatok (pl. megfigyelés,

interjú során felvett hang- és képanyagok) elemzését kódrendszerekkel, számszerűsíthető formában végzik.

MEGFIGYELŐK KÖZÖTTI EGYETÉRTÉS VIZSGÁLATA

Többféle módon lehet értelmezni a megfigyeléses kutatásokban a reliabilitást. Tanulmányozhatjuk az ugyanabban az időpontban több megfigyelő közötti egyetértés mértékét, vizsgálhatjuk a különböző időpontokban végzett megfigyelések közötti eltérés mérté-

két (stabilitást), illetve a kettő kombinációját (különböző időpontok, különböző megfigyelők) (Falus, 2004).

A reliabilitásvizsgálatok egyik fontos kérdése tehát a megfigyelők közötti egyetértés vizsgálata. Erre nagy hangsúlyt helyez mind a társadalomtudományi kutatómódszertani szakirodalom (pl. Sántha, 2012; De Swert, 2012; Fletcher és mtsai, 2011; Hallgren, 2012; Landers, 2015; Antoine és mtsai, 2014), mind az orvosi kutatások szakirodalmában (pl. Kottner és mtsai, 2011; Koo és Li, 2016).

Gyakori hibák a megfigyelők közötti egyetértést vizsgáló kutatásokban

A következőkben azokat a kérdéseket vesszük sorra, amelyek a kódolást használó, esetleg több kódolóval dolgozó kutatások esetében merülnek fel, ahol a megfigyelők közötti egyetértés vizsgálata nélkülözhetetlen. Először is érdemes szétválasztani a belső és a külső megbízhatóságot erősítő, illetve veszélyeztető tényezőket (Sántha, 2012). A külső megbízhatóságot befolyásoló tényezők közé sorolhatjuk a dokumentálást, a kutatói pozíció vagy álláspont megfogalmazását, a megfelelő mintavételi eljárást, az egységes instrukciók használatát, egymástól valóban függetlenül kódoló megfigyelőket (pl. Sántha, 2012; Hayes és Krippendorff, 2007).

A belső megbízhatóság kapcsán is számos probléma merülhet fel a kutatásokban, amely gyakran a megfigyelők közötti egyetértésre, illetve a megfigyelési kategóriákra vonatkozik. A megismételhetőség a megbízhatóság egyik legfontosabb kritériuma. A megismételhetőséget úgy becsülhetjük meg, hogy megmérjük a független megfigyelők közötti egyetértés mértékét

(Krippendorff, 2004). Minél több kódoló értelmezi, kódolja hasonlóan a megfigyelt anyagot, annál nagyobb a valószínűsége, hogy bármilyen kódoló hasonlóan látná azokat. Fontos azonban megjegyeznünk, hogy a megfigyelők közötti egyetértés statisztikai módszerekkel való vizsgálata ugyan utal a reliabilitás mértékére, de nem azonos azzal (Krippendorff, 2004).

Előfordul az is, hogy a kutatók nem végzik el vagy nem publikálják a véglegesen használt változókon és ingeranyagon használt megfigyelők közötti egyetértés vizsgálatát (Hallgren, 2012). Ezt Lombard és munkatársai szerint jól példázza, hogy pl. az 1994–98 között publikált, az elemzésükbe bevont 200 kommunikációs megközelítésű tartalomelemzési kutatásban mindössze 69%-ban közöltek valamilyen reliabilitás mutatót a szerzők (Lombard és mtsai, 2002).

Gyakori probléma az is, hogy a kutatók nem értelmezik a megfigyelők közötti egyetértés mértékét a saját kutatási kérdéseikre vonatkoztatva. Például, ha fontos megmutatni, hogy a kódolók önállóan hasonló következtetéseket tudnak levonni az általuk megfigyelt jelenségekről, akkor hasznos lehet az megfigyelők közötti egyetértés kvalitatív értelmezésének összehasonlítása a hasonló eszközökből származó, korábbi kutatások eredményeivel. Továbbá annak megállapítása, hogy a vizsgált jelenségnél vagy módszernél a megfigyelők közötti egyetértés milyen értékei tekinthetők magasnak, elfogadhatónak vagy rendkívül alacsonynak (Hallgren, 2012).

A megfigyelők közötti egyetértés mértéke, illetve annak értelmezése különösen fontos, ha az ezek alapján létrehozott változókat később hipotézisvizsgálatokra is használni kívánjuk. A megfigyelők közötti alacsony egyetértés azt jelzi, hogy

a megfigyelt értékek sok mérési hibát tartalmaznak, így ez növelheti a II. típusú hibák valószínűségének megjelenését. Ebben az esetben a „zaj” olyan erős lehet, hogy a kutató nem észleli a ténylegesen létező kapcsolatot, és így hibásan ítéli meg a hipotézisét (Hallgren, 2012).

A megfigyelők közötti alacsony egyetértést minden esetben érdemes alaposan megvizsgálni, a lehetséges okokat feltárni, illetve legalább alternatív magyarázatokat kínálni. Ezen okok között felmerülhet, hogy nagyon szűk tartományban kódoltak a megfigyelők, nem jól sikerült a betanításuk, a megfigyelt jelenség nehezen megfigyelhető, illetve nehezen számszerűsíthető vagy kategorizálható a megfigyelés tárgya. A kutatások kritikus pontja, hogy a kutatók milyen gondolatmenetet használtak bizonyos változók megtartása vagy kihagyása esetében. A kutatási beszámolókból gyakran kimarad a fenti, kutatómódszertani szempontból fontos kérdések bemutatása és megvitatása (Hallgren, 2012).

Tanulmányunkban leginkább abban a problémakörben kívánunk segítséget adni, amely a megfigyelők közötti egyetértés statisztikai igazolásához kötődik. Ennek oka, hogy a kutatások gyakran nem a kutatási elrendezésnek megfelelő statisztikai eljárást használják, vagy nem jól értelmezik azokat a reliabilitás, illetve a kódolók közötti egyetértés mérésekor (Hallgren, 2012).

A megfigyelők közötti egyetértés vizsgálatára gyakran használt statisztikai mutatók

Annak eldöntéséhez, hogy milyen statisztikai mutatót használjunk a megfigyelők közötti egyetértés mértékének vizsgálatára, praktikusán két szempontot kell figyelembe

vennünk: (1) hány megfigyelő közötti egyetértést akarunk vizsgálni (kettő vagy annál több megfigyelő); illetve (2), hogy milyen skálán kódoltak a megfigyelők (nominális, ordinális, intervallum- vagy arányskála). Ezen kívül a minta mérete is befolyásolja, hogy melyik statisztikai eljárás tekinthető megbízhatónak. Meglehetősen nagy ellentmondásokkal találkozhatunk a szakirodalomban arra vonatkozóan, hogy melyik reliabilitást mérő eljárás a leginkább megbízható (pl. Fletcher és mtsai, 2011; Krippendorff, 2004; Sántha 2012), melyik statisztikai próba áll a legközelebb a „valóság”hoz”, melyik próba mér túlságosan szigorúan vagy melyik éppen túl megengedő (Lombard és mtsai, 2002).

A megfigyelők közötti egyetértés vizsgálatára többek között a következő statisztikai próbák jelentek meg a szakirodalomban: Bennett-S és átalakított verziói, pl. a Guilford-féle-G, Benini- β , Cohen-kappa, Cronbach-alfa, Fleiss-K, Goodman-Kruskal- λ , Osgood-Holsti-féle-CR, Pearson-r, Scott- π , százalékos egyetértés (pl. Felcher és mtsai, 2011; Lombard és mtsai, 2002; Krippendorff, 2004; Hayes és Krippendorff, 2007). Az egyes eljárások részletes alkalmazhatósági feltételeit itt nem célunk bővebben kifejteni.

A következőkben azokat a statisztikai mutatókat ismertetjük, amelyeket az IBM SPSS 23 programcsomagban megjelennek. Nominális változók és két kódoló vizsgálatoknál a Cohen-kappa áll rendelkezésünkre. Ugyan ezt az eljárást is kritizálták (pl. Krippendorff, 2004), a szakirodalom zömmel elfogadja ennek használatát (pl. Lombard és mtsai, 2002; Sántha, 2012; Szokolszky, 2004). Ordinális, intervallum- vagy arányskálánál két főnél Cronbach-alfa, több megfigyelőnél a Kendall-W (Ketskeméty és mtsai, 2011) használható.

Az ICC (*intraclass correlation coefficient* – osztályon belüli korrelációs együttható) ugyan minden skálatípuson és több kódoló esetén is használható, azonban a szakirodalomban nincs egyetértés abban, hogy használata több megfigyelő esetében is megfelelően mér-e (Hallgren, 2012). Az ICC-n kívül a Krippendorff-alfa (KAPLHA) az a mutató, amelyik hasonlóan széles körben alkalmazható, ugyanakkor a használatával kapcsolatban eddig nem merültek fel szakmai kritikák. A KAPLHA a legtöbb megfigyelők közötti egyetértést mérő mutatóval szemben felmerülő elvárásnak megfelel.

- Két vagy több megfigyelő közötti egyetértés is vizsgálható az eljárással.
- A megfigyelők számát, a megfigyelők személyét és megfigyelések sorrendjét, illetve az esetleg kimaradó megfigyelőket, hiányzó adatokat is kezeli (Krippendorff, 2004; Hayes és Krippendorff, 2007).
- A reliabilitást vizsgáló mérőszám könnyen értelmezhető, nem függ az elemzésnél használt kategóriák számától vagy a skála méretétől. A konvencióknak megfelelően ez az érték 0 és 1 közé esik (Hayes és Krippendorff, 2007). A mutató abban az esetben lehet negatív, ha a megfigyelők a véletlennél rosszabb mértékben értenek egyet egymással (De Swert, 2012). A mérőszám kidolgozói a 0,66-os minimális kritériumértéket, illetve a 0,8-as „jó egyetértés” értéket is megadják. Tehát a mérőszám megmutatja, hogy a következő vizsgálatokban milyen mértékben támaszkodhatunk az adatokra (Krippendorff, 2004; Hayes és Krippendorff, 2007).
- A reliabilitás vizsgálat minden esetben illeszkedik az adatok mérési szintjéhez. Tehát a reliabilitást mérő mutató a teljes

adatállományunkat, illetve minden skálatípus adatait fel tudja dolgozni, így akár a különböző skálakon kapott reliabilitásértékek is összevethetők (Krippendorff, 2004; Hayes és Krippendorff, 2007).

- Többféle statisztikai programcsomagban is kiszámolható vagy beilleszthető (pl. Stata, SPSS).

A Krippendorff-alfa tudományterülettől függetlenül minden olyan kutatásban használható, ahol több megfigyelőtől származó adatokat kell összevetni. A kutatások a tömegkommunikációtól (Lombard és mtsai, 2002; Beam, 2003) kezdve, az interkulturálisan megjelenő tartalmak tartalomelemzésén (Moretti és mtsai, 2011), a természetes nyelvhasználati szokásokon (Antoine és mtsai, 2014), illetve az egészségpszichológián (Powell, 2016) és a marketingen (Hughes és Garrett, 1990) át az orvosi és igazságügyi tudományokig (Dedouit és mtsai, 2008) terjednek.

Összefoglalva tehát a Krippendorff-alfa széles körben elterjedt statisztikai mutató, tudja kezelni, ha kettőnél több kódoló szolgáltatja az adatokat, egyformán jól használható nominális, ordinális, intervallum- és arányskála esetében, valamint jól kezeli a hiányzó adatokat és a kis mintaméretet (Krippendorff, 2004; Hayes és Krippendorff, 2007). Mindezek mellett az eljárás az SPSS-be is importálható (De Swert, 2012).

Kutatási cél

Tanulmányunk fókuszában a kódolók közötti egyetértés pontos mérése áll. A következőkben példaként bemutatjuk egy megfigyelési kategóriarendszer kidolgozását, amelynek fejlesztésében kulcsfontosságú

része volt a megfigyelők közötti egyetértés mérése és értelmezése. A kategóriarendszert egy alapvetően kvalitatív megközelítésű kutatásban alkalmaztuk, amelynek célja integráló óvodai csoportokban megvalósuló szociális integráció, illetve ennek társas interakciókban történő megjelenésének vizsgálata volt. Ehhez szükséges volt egy viselkedés-megfigyelési kategóriarendszer kidolgozása, amellyel óvodás korú gyermekek interakciói válnak megbízhatóan leírhatóvá és kódolhatóvá (Böddi, 2017). A fő kutatásmódszertani kihívás az volt, hogy kettőnél több kódoló és dichotóm változók esetében a megfigyelők közötti egyetértést teszteljük. Tanulmányunkban bemutatjuk a Krippendorff-alfa (KALPHA) alkalmazási lehetőségeit, vagyis azt, hogy a kódolt, számszerűsített megfigyelési adatokon hogyan tudjuk mérni a megfigyelő közötti egyetértés fokát.

MÓDSZER

Kutatási elrendezés

A kutatásban az életkori sajátosságok miatt a természetes megfigyelés módszerét alkalmaztuk (ld. Böddi és mtsai, 2015). Az óvodai csoportszobákban délelőtt videókamerás felvételeket készítettünk. A kamerákat a csoportszoba három pontján helyeztük ki, ezzel lefedve a csoportszoba összes játékterét. Egy-egy felvétel kb. 90 percig tartott. A 90 percet zömmel szabad játékkal töltötték a gyermekek, de alkalmanként ebbe az intervallumba felnőtt által irányított tevékenység is megjelent. A kutatásban 3 budapesti óvoda 5 integráló csoportja vett részt. Egy óvodai csoportban általában három alkalommal készítettünk felvételeket. Minden vizsgált óvodai csoportba

minimum 1, maximum 3 sajátos nevelési igényű gyermek járt. Mindegyik gyermek szakértői véleménnyel rendelkezett. Az elsődleges kategóriarendszer tesztelésére használt felvételeket egy negyedik budapesti óvodában vettük fel.

A kutatást az Eötvös Loránd Tudományegyetem Etikai Bizottsága engedélyezte 2015-ben, a felvételek készítéséhez az óvodák vezetése, az óvodapedagógusok és a gyermekek szülei is hozzájárultak. Az adatgyűjtés problematikáit és azok feloldását korábban részletesen bemutattuk (ld. Keszei és mtsai, 2016).

Elsődleges kategóriarendszer

A megfigyelési kategóriarendszer kidolgozásához két forrásra támaszkodtunk. Egyrészt a szakirodalomban fellelhető integráló és inkluzív óvodai csoportokban végzett hasonló témájú kutatások kategóriáit (pl. File, 1994; Harper és McCluskey, 2002, 2003), másrészt a videófelvételeinken megfigyelhető további, gyakran megjelenő viselkedéselemeket vettük alapul.

Az elsődleges kategóriarendszer kialakításakor célunk volt valamennyi lehetséges csoportszobai tevékenységet és azok társas szintjét is a kialakított kódrendszerbe bevonni. Mivel a kutatás központi eleme volt a sajátos nevelési igényű gyermekek megfigyelése, ezért a kategóriarendszer kialakításában is az egyén megfigyelésére helyeztük a hangsúlyt. A tevékenységek társas szintjének besorolásához Parten (1932, idézi Harper és McCluskey, 2002, 2003) kategóriarendszerét tekintettük alapnak, amely a következő kategóriákat használja: magányos játék, párhuzamos játék, szemlélődő (*onlooker*) viselkedés és interakció (részletesen pl. Böddi és mtsai, 2015). Ez a kategó-

riarendszer a nemzetközi szakirodalomban a kora gyermekkori játék társas szintjének tanulmányozásában igen elterjedt (pl. Upton, 2011) hasonlóan hazánkban a Mérei-féle (1989) aktometriához. Azonban a Parten-féle kategóriarendszer egyszerűbben kezelhetőnek és könnyebben továbbfejleszhetőnek bizonyult a számunkra, szemben az aktometria tizennégy szintjével (Mérei, 1989).

Az aktometriában egyszerre figyeljük a gyereket és a csoportot (Mérei, 1989), így az egyén interakcióira fókuszáló kutatásunkban ez a módszer nem bizonyult szükségesnek. Mivel a kutatás fókuszában a felnőtt-gyerekek interakciók tanulmányozása is megjelent, ezért megalkottuk a *felőtt-gyermek interakció* kategóriáját, amelyet tovább bontottunk. Lényegesnek tartottuk a felnőtt jelenlétét, illetve a felnőttel folytatott kétszemélyes interakciókat a kategóriarendszerben megjeleníteni.

Az elsődleges kategóriarendszer használatakor a kódolók a megfigyelt gyermek viselkedését figyelték meg, a gyermek tevékenységét, interakcióit kódolták. Az elsődleges kategóriarendszer a következő kategóriákat tartalmazta (ld. *1. táblázat*).

Tevékenységek: eszik-iszik, beszélget, nézeget (könyvet, tárgyat), alkot, játszik, lézeng, önkiszolgál, rakodik, konfliktus, egyéb.

Tevékenységek társas szintje: teljesen magányos tevékenység, magányos tevékenység (társas jelenléttel), párhuzamos (paralel) tevékenység, szemlélődés, interakció.

Az *interakción* belül a következő *alkategóriákat* különítettünk el: interakció gyermekcsoporttal, interakció felnőtt-gyermek-csoporttal, kétszemélyes interakció gyermekkel, kétszemélyes interakció felnőttel. A *kétszemélyes interakció felnőttel* kategórián belül, vagyis amikor a gyermek a felnőttel áll interakcióban és a felnőtt

figyelmét elsősorban ő köti le, további alkategóriákat határoztunk meg: beszélget, ad valamit, mutat valamire/valamit, odafordul felé, bólint, mosolyog rá, kinyújtja felé a kezét, odahajol, testi kontaktus. A beszélgetésen kívül az itt megjelenő alkategóriákat mind nemverbális jelzéseket kódolnak.

A kódolás folyamata

A próbakódolás

Az elsődleges kódrendszer próbakódolását először 6 független kódoló végezte (mindegyikük nő, átlag életkoruk: 23,33 év), akik az ELTE neveléstudomány szakjának mesterszakos hallgatói voltak, tehát a gyermekek viselkedése, illetve annak megfigyelése nem volt teljesen ismeretlen a számukra.

Az elsődleges kategóriarendszer teszteléséhez olyan videóanyagot használtunk, amelyet a kutatás következő szakaszában nem alkalmaztunk. A kódolók rövid betanításon vettek részt, amelynek során találkoztak a 3 kameraállásból felvett videóanyag kezelésével, kipróbálták, hogyan lehet egy-egy gyermeket nyomon követni a felvételeken. Megismerkedtek a kategóriarendszerrel, illetve a táblázattal. A gyermekek tevékenységét, illetve annak társas szintjét kellett kódolniuk. Amennyiben interakciót láttak a kódolók, akkor az interakció alkategóriái közül is választani kellett. A kódolók összesen 28 kategóriát kellett használni. A próbakódolás szakaszában a kódolók időközököt jegyeztek fel egy-egy kategória megjelenésekor, illetve amikor változás következett be a gyermek tevékenységében vagy interakciójában. Váltás esetében kellett az újonnan megjelent kategóriánál megjelölni. A kódolók ezt az eljárást nehezen követhetőnek találták, így ezt a következő szakaszban módosítottuk.

1. táblázat. Az elsődleges kategóriarendszer kategóriái (Böddi, 2017)

Tevékenység		
Eszik-iszik		
Beszélget		
Nézeget (könyvet, tárgyat)		
Alkot		
Játszik		
Lézensz		
Önkiszolgál		
Rakodik		
Konfliktus		
Egyéb		

Tevékenység társas szintje	Interakció jellege (ha van)	Kétszemélyes interakció felnőttel jellege (ha van)
Teljesen magányos tevékenység		
Magányos tevékenység (társas jelenléttel)		
Párhuzamos (paralel) tevékenység		
Szemlélődés		
Interakció, interaktív tevékenység	Kétszemélyes interakció gyermekkel	
	Kétszemélyes interakció felnőttel	Beszélget Ad valamit Mutat valamire/valamit Odafordul felé Bólint Mosolyog rá Kinyújtja felé a kezét Odahajol Testi kontaktus
	Interakció gyermekcsoporttal	
	Interakció felnőtt-gyermek csoporttal	

A teszt kódolás folyamata

A teszt kódolást 8 független megfigyelő (7 nő, 1 férfi, átlagéletkor: 21,88 év) végezte, akik az ELTE Tanító- és Óvóképző Kar hallgatói voltak. Ők tehát a próbakódolásban nem vettek részt.

A kódolók a próbakódolás során felmerült problémák megoldásával létrehozott új betanításon vettek részt, amelyben az idő kódok helyett más megoldást használtunk. A megfigyelendő videóanyagot 10 másodperces szakaszokra osztottuk. Az instrukció szerint a kódolók akkor tekinthettek egy történést jelenlevőnek, ha a 10 másodperces szakaszból legalább 5 másodpercig jellemző volt a gyermek tevékenységére. A kódolóknak összesen 360 darab 10 másodperces szakasz kapcsán kellett döntést hozniuk, így a teszt felvétel összesen 52,67 perc hosszúságú volt.

A teszt felvételeken egymás után több gyermek viselkedését is kódolták a megfigyelők. A kódolók vakok voltak mind az óvodák, mind az óvodai csoport tekintetében, valamint arról sem volt információjuk, hogy a megfigyelt gyermek sajátos nevelési igényű-e vagy sem.

Felmerülhet a kérdés, hogy miért volt szükség 8 kódolóra, hiszen 2 kódoló eredményei alapján is lehet megfigyelők közötti egyetértést számolni (Sántha, 2012). Ugyanakkor úgy gondoljuk, hogy különösen egy kategóriarendszer kidolgozásának kezdeti szakaszában mindenképpen szükség van több megfigyelőre – akár a kutatásvezető prekonceptióinak kiszűrésére vagy azért, hogy lássuk mely kategóriák könnyen használhatóak, és melyek lesznek azok, amik esetleg nehezen észrevehető különbségeket, jelenségeket céloznak meg és ezért alacsony megfigyelők közötti egyetértés értékeket kapnak.

A Krippendorff-alfa makró használata SPSS-ben

A kódolók közötti megbízhatóság vizsgálatához a 8 független kódoló között kellett reliabilitást számítani dichotóm változók esetében. Így a megfigyelők közötti megbízhatóság tesztelésére a Krippendorff-alfa (KALPHA) kiszámítására került sor (Hayes és Krippendorff, 2007; De Swert, 2012). A számításokat IBM SPSS Statistic 23 programmal végeztük, melynek az alapeszköztárába ugyan nem tartozik bele a Krippendorff-alfa (KALPHA), azonban egy, az SPSS-be egyszerűen illeszthető makró (Hayes, én.; De Swert, 2012) használatával az értékek könnyen kiszámíthatók. A makró (Hayes, én.) letöltése² után azt az SPSS-ben kell lefuttatni. Érdemes arra odafigyelni, hogy az SPSS újraindításakor a makró újra le kell futtatni, mert automatikusan nem kezdi el használni a program.

Az adattáblában az egyes megfigyelők által felvitt kódokat külön változóként kezeltük (pl. a játék_1_kódoló, játék_2_kódoló stb.). A hiányzó értékek nem okoznak problémát a KALPHA számításban, így az adattisztítással és a hiányzó adatok kezelésével nem kellett foglalkoznunk. Arra viszont oda kellett figyelni, hogy a KALPHA parancsba írjuk be, hogy milyen skálán kódoltak a megfigyelők (esetünkben ez az 1-es érték volt, mivel nominális skálát használtunk, ordinális változóknál 2, intervallumskálánál 3 és arányskálánál 4-es értéket kell használni) (De Swert, 2012).

Ahhoz, hogy a KALPHA-értékeket ki tudjuk számolni, az egyes kódok, megfigyelt értékek nem lehetnek konstansok (De Swert, 2012). Esetünkben ez nem merült

² <http://www.afhayes.com/public/kalpa.zip> (Letöltés ideje: 2019. 02. 02.)

fel problémaként, hiszen minden kategóriánk esetében voltak olyan időszakok, amikor az adott tevékenység vagy interakciótípus nem jelent meg (0), illetve amikor megjelent (1).

Az egyetlen probléma, amit a KALPHA nehezen kezel, az a ritka kategóriák megjelenése (De Swert, 2012). Ez a probléma az elsődleges kategóriarendszerünk átfogó jellege miatt sok kategóriánál megjelent. Viszont a ritka kategóriák kezelésében is segíthet, ha összevonunk kategóriákat. Ahogy látni fogjuk, a kategóriák összevonása esetünkben is a legtöbb kategória kapcsán elősegítette a KALPHA-érték emelkedését, de ez nem tekinthető univerzálisan működő stratégiának (De Swert, 2012).

EREDMÉNYEK

Kutatásunk célja egy megbízható kódrendszer kialakítása volt, ennek érdekében az elsődleges kategóriarendszer átalakítására volt szükség. Az átalakításhoz felhasználtuk a teszt kódolás során született KALPHA-értékeket (ld. 2. táblázat). Ezen kívül értelmes kategóriaösszevonásokat is végeztünk, és ezen összevont kategóriákra is kiszámoltuk a KALPHA-értékeket (ld. 3. táblázat). A 2. és 3. táblázatban a megbízható kategóriákat szürke kiemeléssel jeleztük. A hiányzó értékek esetében a kategóriák előfordulási gyakorisága vagy nulla vagy alacsony volt.

A megfigyelők közötti egyetértés vizsgálata során megjelenő alacsony KALPHA-értékek háttérben az is állhat, hogy egy-egy megfigyelő a többiekétől nagyban eltérően kódolja a megfigyelt eseményeket. Ilyenkor érdemes szisztematikusan megvizsgálni, hogy melyik kódoló az, aki

„kilóg”, melyik az az egy vagy esetleg több kódoló, akinek a kizárása után jelentősen javulnak a KALPHA-értékek (De Swert, 2012).

A teszt kódolásban résztvevő 8 fő közül egy fő kizárásával a KALPHA-értékek több kategória esetében is emelkedtek (ld. 2. és 3. táblázat megjegyzései között). Megfigyelhettük, hogy a kizárt kódoló a fő kategóriákban a többi kódolóhoz hasonlóan értelmezte a videón látott eseményeket. A tevékenységek és a társas szintek felismerésében sem tért el jelentősen a többi megfigyelőtől. Azonban az interakció kategórián belül ez már nem mondható el. A kizárt kódoló tehát a fő kategóriákban ugyan jól, a többi kódolóhoz hasonlóan kódolt, de az alkategóriáknál már nem. Úgy döntöttünk, hogy a fő kategóriák KALPHA-értékeinél megtartjuk a kódolását és csak az alkategóriák KALPHA-értékeinek kiszámításánál zárjuk ki a feldolgozásból.

Az alacsony KALPHA-értékeket a kategóriák viszonylag ritka előfordulása is okozhatja. Feltűnő, hogy az interakció alkategóriák esetében alacsony KALPHA-értékkel találkozhatunk. Tehát az interakció alkategóriáin belüli cízellálás a legtöbb kódolónak nehézséget okozhatott. Ugyanakkor csak egy kódoló volt, akinek a kizárásával a KALPHA-értékek jelentősen növekedtek. Megállapíthatjuk, hogy az interakciók nehezen kódolhatók, az interakció típusai közötti differenciálás vélhetően több megfigyelésben szerzett tapasztalatot igényel.

Az eredeti kategóriarendszer fejlesztésének egy fontos lépése volt egyes kategóriák összevonása. A kategóriaösszevonások célja az volt, hogy a ritkán megjelenő, de a vizsgálat szempontjából fontos tartalmat hordozó kategóriákat, ha nem is az eredeti részletezettségükben, de megtarthassuk.

2. táblázat. A teszt kódolás KALPHA-értékei az eredeti kategóriákban (Böddi, 2017: 197)

Kategória neve		KALPHA értéke	Megjegyzés
Tevékenység	Eszik-iszik	0,96	
	Beszélget	0,28	alacsony korrelációk a kódolók között – nem értették meg?
	Nézeget	0,61	
	Alkot	0,89	
	Játszik	0,77	
	Önkiszolgálás	–	ez a kód csak nagyon ritkán jelent meg
	Rakodás	0,26	
	Konfliktus	0,24	alacsony korrelációk a kódolók között és ritkán jelenik meg
	Egyéb	0,09	alacsony korrelációk a kódolók között és ritkán jelenik meg
Társas szint	Teljesen_magányos	0,71	
	Magányos__társas jelenléttel	0,11	alacsony korrelációk a kódolók között – nem értették meg?
	Parhuzamos	0,35	alacsony korrelációk a kódolók között
	Szemlélődés	0,15	alacsony korrelációk a kódolók között és ritkán jelenik meg
	Interakció	0,72	
Interakció jellege	Kétszemélyes__gyerek	0,64	kihagyott kódoló nélkül: 0,7
	Kétszemélyes_felnőtt	0,67	kihagyott kódoló nélkül: 0,68
	Gyerekcsoport	0,49	alacsony korrelációk – nem értették meg?
	Felnőtt-gyerek_csoport	0,45	alacsony korrelációk – nem értették meg?
Kétszemélyes felnőtt interakció – a felnőtt interakció kategóriái	Beszél	0,54	alacsony korrelációk – nem értették meg?
	Ad valamit	0,39	alacsony korrelációk a kódolók között és ritkán jelenik meg
	Mutat	0,39	alacsony korrelációk a kódolók között és ritkán jelenik meg
	Odafordul	0,4	alacsony korrelációk a kódolók között és ritkán jelenik meg
	Bólint	-	ez a kód csak nagyon ritkán jelent meg
	Mosolyog	-	ez a kód csak nagyon ritkán jelent meg
	Kinyújtja a kezét	-	ez a kód csak nagyon ritkán jelent meg
	Odahajol	0,49	alacsony korrelációk a kódolók között – nem értették meg?
	Testi kontaktus	0,71	kihagyott kódoló nélkül: 0,74

Például a *kétszemélyes gyermek* és a *gyermekcsoport interakció* kategóriák összevonása esetében a 0,64 és 0,49 KALPHA-érték az összevont kategóriában 0,66-ra emelkedik, valamint egy kódoló kihagyásával a KALPHA-érték 0,71 lesz. A létrejött kategória tehát megbízhatóan mér, és a csoportszobai interakciók szempontjából egy fontos komponens, hogy a megfigyelt gyerek egykorú társaival (másik gyerekkel vagy gyermekcsoporttal) került-e interakcióba a felnőtt jelenléte nélkül. Így a kategóriának megtartott jelentéstartalma marad az összevonás után is.

Egy másik példa az összevonásra az eredeti kategóriarendszerben két kategó-

riaként megjelenő *odahajol* és *odafordul* kategóriák összevonása. A két kategória összevonásával nem veszítettünk volna sok információt, azonban a kategória előfordulási gyakorisága az összevonás után is alacsony maradt és az összevont kategória értéke sem érte el a kívánt 0,66 körüli értéket. Így végül mindkét kategóriát kihagytuk a végleges kategóriarendszerekből.

A végleges kategóriarendszer

A fentiek alapján az eredeti 28 kategóriából az alacsony megfigyelők közötti egyetértéssel jellemezhető kategóriák kizárása és az értelmes kategóriaösszevonások után

3. táblázat. A teszt kódolás KALPHA-értékei az összevont kategóriákban (Böddi, 2017: 197–198)

Kategória neve		KALPHA értéke	Megjegyzés
Társas szint	Magányos_társas + Szemlélődés + Párhuzamos	0,58	
	Interakció + Párhuzamos	0,49	
	Párhuzamos + Szemlélődés	0,41	
	Párhuzamos + Szemlélődés + Interakció	0,45	
	Teljesen_magányos + Magányos_társas	0,44	
Interakció jellege	Kétszemélyes (kétszem_gyerek + kétszem_felnőtt)	0,63	
	Csoport (gyermekcsoport + felnőtt-gyerek_csoport)	0,48	alacsony korrelációk (0,4 körüliek)
	Gyerek (kétszem_gy + gyermekcsoport)	0,66	kihagyott kódoló nélkül: 0,71
	Felnőtt (felnőtt-gyerek_csop + kétszem_felnőtt)	0,72	kihagyott kódoló nélkül: 0,74
	Odahajol + Odafordul	0,54	
Kétszemélyes felnőtt interakció – a felnőtt interakció kategóriái	Ad_valamit + mutat	0,41	nagyon ritka

4. táblázat. A végleges kategóriarendszer kategóriái és a megbízhatósági értékek 8, illetve 2 kódoló esetén

Kategória neve	KALPHA értéke 8 megfigyelővel számított reliabilitás a tesztfelvételeken	Cohen-kappa értéke 2 megfigyelővel számított reliabilitás a teljes ingeranyagon
Eszik-iszik	0,96	0,92
Alkot	0,89	0,92
Játszik	0,77	0,66
Magányos (korábban Teljesen magányos)	0,71	0,5
Interakció	0,72	0,67
Kétszemélyes gyerek	0,70	0,60
Kétszemélyes felnőtt	0,68	0,69
Testi kontaktus a felnőttel	0,74	0,61
Gyerek (Kétszemélyes gyerek + Gyerekcsoport)	0,71	0,62
Felnőtt (Felnőtt_gyerek csoport + Kétszemélyes felnőtt)	0,74	0,74

a végleges kategóriarendszert 10 kategória alkotta (ld. 4. táblázat). A tevékenységek közül az *eszik-iszik*, *alkot*, *játszik*; a társas szintből a *magányos* és az *interakció* kategóriákat tartottuk meg. Az interakció jellegének megragadására a *kétszemélyes gyerek*, *kétszemélyes felnőtt*, *gyerek* (kétszemélyes gyerek + gyerekcsoport) és a *felnőtt* (kétszemélyes felnőtt + felnőtt_gyerek csoport) kategóriákat használtuk. A kétszemélyes felnőtt interakciónál megjelenő kategóriák közül az eredmények alapján a testi kontaktus kategóriát tarthattuk meg. A végleges kategóriarendszerben szereplő kategóriák leírásai az 1. mellékletben olvashatók.

A végleges kategóriarendszer reliabilitása

Célunk volt annak bizonyítása, hogy a 8 kódoló eredményei alapján átalakított kódrendszer jól használható a kutatásban. Ezért a vizsgálatban kapott eredmények megbízha-

tóságának alátámasztására a végleges ingeranyagon két független kódoló között is kiszámoltunk az egyetértés mértékét.

A végleges kategóriarendszert két független kódoló használta. A megfigyelés ebben az esetben már az óvodai csoportba járó sajátos nevelési igényű gyermek viselkedésére vonatkozott (szemben a próbakódolással, ahol ez nem volt elsődleges szempont). A kódolás a már ismertetett formában zajlott. Összesen 5 csoportban felvett, csoportonként 3 darab, megszakítás nélküli, 90 perces videófelvételt 6 darab 15 perces egységre bontottuk, melyeknek az első 5 perce képezte a kódolás tárgyát. Így a kódolók összesen 450 percnyi anyag 10 mp-enkénti elemzését végezték el. A végleges ingeranyagon a kutatói társadalom által elfogadott és két megfigyelő közötti egyetértésre gyakran használt Cohen-kappa-értékeket közöljük (lásd. 4. táblázat), amelynek kiszámításához szintén az SPSS 23 programot használtuk.

Érdekes megfigyelni, hogy ugyan az értékek többnyire a megfelelő vagy kifejezetten jó értékek közé sorolhatók, vannak olyan kategóriák, amelyekben alacsonyabb reliabilitásértékek figyelhetők meg, mint 8 kódoló esetén. Ez is mutathatja, hogy a megfigyelők egy jól működő kódrendszer használatakor is könnyen eltérhetnek egymástól, az egyéni eltéréseket két fő esetben pedig nehéz kiszűrni.

Felmerül a kérdés, hogy milyen megoldásokkal, alternatívákkal próbálkozhatunk, ha a kódolók között alacsony egyetértéssel találkozhatunk. A nehezen felismerhető, differenciálható kategóriák vizsgálatakor egy lehetséges módszer az orvostudományokban használt gyakorlat átvétele. Ott a reliabilitásvizsgálatoknál a szakértő véleményét vetik össze a kevesebb tapasztalattal rendelkező megfigyelők kódjaival, és a szakértők véleményét tekintik etalonnak (Fletcher és mtsai, 2011). Természetesen az, hogy mikor tekinthetjük a megfigyelők közötti egyetértés mértékét elfogadhatónak vagy kifejezetten jónak, minden egyes kutatás esetében eltérő.

KONKLÚZIÓ

Tanulmányunkban egy óvodai interakciókat vizsgáló kutatás segítségével ismertettük a megfigyelők közötti egyetértés mérésének egy lehetséges statisztikai módszerét. A vizsgálati elrendezés (8 megfigyelő, dichotóm kategóriák) azonban csak egy statisztikai eljárás alkalmazását tették lehetővé, a Krippendorff-alfa (KALPHA) használatát. Tanulmányunkban egy – az óvodai csoportokban megjelenő integrációs szociális aspektusát vizsgáló – kutatáson keresztül bemutattuk a megfigyelők közötti egyetérté-

sek tükrében azt, hogy miként volt érdemes átalakítani a kódrendszerünket.

Az eredeti 28 kategóriából álló kódrendszert a megfigyelők közötti egyetértéseket figyelembe véve egy könnyebben használható, 10 kategóriából álló kódrendszerré alakítottuk. Ezt sikeresen használtuk az óvodai csoportszobákban zajló tevékenységek, társas folyamatok és interakciók feltárására és leírására.

Az egyes kategóriák, illetve az összevont kategóriák előfordulási frekvenciáját kiszámítva vizsgálhatóvá vált, hogy a megfigyelt sajátos nevelési igényű gyermekek a megfigyelt időben milyen arányban foglalták el magukat a csoportszobában egyes tevékenységekkel, s hogy ezek a tevékenységek milyen társas jellemzőkkel bírtak. Tovább bontva a kategóriákat elemezhetővé vált az is, hogy – az óvodáskorban oly kiemelt jelentőségű – játék esetében milyen interakciók jellemezték azt.

A kutatásunkban létrehozott kategóriarendszer jól használható, és lehetővé teszi, hogy a kategóriarendszert, illetve a kódolással nyert adatokat hipotéziselemzésre is felhasználjuk. A kódrendszer jól használható más óvodai jelenségek vizsgálatára is, például szociális kapcsolatok feltárására, illetve akár idősebb gyermekcsoportnál is használható.

Természetesen elengedhetetlen a kategóriarendszerek folyamatos fejlesztése, és az esetlegesen nem megbízhatóan mérő kategóriák kapcsán mind a kategória definíciójának, mind a betanítás pontos instrukciójának az újragondolása. Jelen kutatás eredményei alapján a Magányosság kategóriája szorul további részletezésre és vizsgálatra.

A Krippendorff-alfa (KALPHA) az SPSS-be könnyen importálható statisztikai eljárás, amely sok olyan problémát jól kezel, melyeket más statisztikai próbák nem vagy nem ilyen

mértékben tudnak megoldani. A KALPHA lehetővé teszi, hogy egy statisztikai eljárást alkalmazva bármilyen vizsgálati elrendezésből származó megfigyelők közötti egyetértés mértékét ki tudjuk számolni, függetlenül a kódolók számától, a skála fajtájától, a minta méretétől és a hiányzó adatoktól.

A KALPHA alkalmazásának lehetőségét bemutatva szeretnénk támogatni azt a törekvést, hogy a kvalitatív és kvantitatív módszerek egymást kiegészítve, egymást segítve használjuk. Ennek egyik kézenfekvő eszköze, ha a megfigyelési módszerek

reliabilitását szakszerűen mérjük, publikáljuk és így módszertanilag is egyre magasabb színvonalú vizsgálatokat végzünk.

KÖSZÖNETNYILVÁNÍTÁS

Köszönettel tartozunk a próba- és tesztködésben résztvevő hallgatóknak, valamint Szathmári Editnek a végleges kódolásban való értékes együttműködéséért, továbbá Hunyady Györgynének a kutatás szakmai támogatásáért.

SUMMARY

PRACTICAL APPLICATION OF KRIPPENDORFF 'S ALPHA (KALPHA): TESTING INTER-RATER RELIABILITY BETWEEN MORE THAN TWO CODERS FOR DICHOTOMOUS VARIABLES

Background, Objectives: Qualitative studies have become increasingly appreciated by psychological research since the 1990's. The cooperative use of qualitative and quantitative methods can enhance this process. A perfect example can be a quantitative verification of reliability calculating inter-rater reliability when coding observation. One of the less published topics is the problem of reliability in the case of multiple coders and dichotomous categories. The statistical procedure to be used in this case is not part of the basic statistical toolkit of researchers. In the present study – used as an example – our goal was to develop a reliable categorization system to describe interactions between children.

Methods: Natural observation was conducted in kindergartens. The children's behavior was recorded on video. The purpose of elaborating a code system was to grab children's activities, the social level and the characteristics of interactions. The primary code system was used by eight independent coders, tested on a 53-minute video sample.

Results: Krippendorff's Alpha (KALPHA) was used to calculate inter-rater reliability between the coders. Currently, this is the only statistical process that is easily implementable into SPSS. It also manages calculating reliability even in the case of more than two coders, dichotomous and binomial variables as well. Analyzing KALPHA values contributed to the development of the final category system. Two independent coders successfully used the newly created code system proved by Cohen's kappa coefficients.

Conclusions: Using KALPHA can be easily integrated into Hungarian research and statistical practice. KALPHA coefficient fills a niche in the examination of inter-rater reliability.

Keywords: reliability, inter-rater reliability, KALPHA, dichotomous variables, statistical test, coding

IRODALOM

- ANTOINE, J., VILLANEAU, J., LEFEUVRE, A. (2014): Weighted Krippendorff's alpha is a more reliable metrics for multi-coders ordinal annotations: experimental studies on emotion, opinion and coreference annotation. *EACL 2014*. 1–10. <https://aclweb.org/anthology/E14-1058> (Letöltés ideje: 2019. március 28.)
- BEAM, R. A. (2003): Content differences between daily newspapers with strong and weak market orientations. *Journalism & Mass Communication Quarterly*, 80(2). 368–390.
- BÖDDI Zs. (2017). *Az óvodai integrált nevelés vizsgálata: Az eredményességet növelő tényezők és az óvóképzés fejlesztésének fókuszával*. Doktori disszertáció. Témavezető: Hunyady Györgyné. ELTE PPK, Budapest.
- BÖDDI Zs., KESZEI B., SERFŐZŐ M., DÜLL A. (2015): A megfigyelés kutatómódszertana – Interakciók megfigyelése integrált és inkluzív óvodai környezetben. *Gyermeknevelés*, 3(2). 29–50.
- DE SWERT, K. (2012): *Calculating inter-coder reliability in media content analysis using Krippendorff's Alpha*. Center for Politics and Communication, Amsterdam. 1–15.
- DEDOUIT, F., BINDEL, S., GAINZA, D., BLANC, A., JOFFRE, F., ROUGÉ, D., TELMON, N. (2008): Application of the Iscan method to two- and three-dimensional imaging of the sternal end of the right fourth rib. *Journal of Forensic Sciences*, 53(2). 288–295.
- FALUS I. (2004). A megfigyelés. In Falus I. (szerk.): *Bevezetés a pedagógiai kutatás módszereibe*. Műszaki Könyvkiadó, Budapest. 125–170.
- FILE, N. (1994): Children's Play, Teacher-Child Interactions, and Teacher Beliefs in Integrated Early Childhood Programs. *Early Childhood Research Quarterly*, 9(2). 223–240.
- FLETCHER, I., MAZZI, M., NUEBLING, M. (2011): When coders are reliable: The application of three measures to assess inter-rater reliability/agreement with doctor–patient communication data coded with the VR-CoDES. *Patient Education and Counseling*, 82(3). 341–345
- GOLAFSHANI, 2003): Understanding reliability and validity in qualitative research. *The Qualitative Report*, 8(4). 597–607.
- HALLGREN, K. A. (2012): Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1). 23–34.
- HARPER, L. V., MCCLUSKEY, K. S. (2002): Caregiver and peer responses to children with language and motor disabilities in inclusive preschool programs. *Early Childhood Research Quarterly*, 17(2). 148–166.
- HARPER, L. V., MCCLUSKEY, K. S. (2003): Teacher–child and child–child interactions in inclusive preschool settings: do adults inhibit peer interactions? *Early Childhood Research Quarterly*, 18(2). 163–184.
- HAYES, A. F. (é. n.). *KALPHA macro*. <http://www.afhayes.com/spss-sas-and-mplus-macros-and-code.html> (Letöltés ideje: 2019. január 11.)
- HAYES, A. F., KRIPPENDORFF, K. (2007): Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1). 77–89.

- HUGHES, M. A., GARRETT, D. E. (1990): Intercoder reliability estimation approaches in marketing: A generalizability theory framework for quantitative data. *Journal of Marketing Research*, 27(2). 185–195.
- KESZEI B., BÖDDI Zs., DÚLL A. (2016): Lehetőségek és nehézségek a természetes megfigyelésben: Egy óvodai vizsgálat tapasztalatai. *Prosperitas*, 3(2). 63–86.
- KETSKEMÉTY, L., IZSÓ, L., KÖNYVES TÓTH, E. (2011): *Bevezetés az IBM SPSS Statistics programrendszerbe*. Artéria Stúdió Kft., Budapest.
- KOTTNER, J., AUDIGÉ, L., BRORSON, S., DONNER, A., GAJEWSKI, B. J., HRÓBJARTSSON, A., ROBERTS, C., SHOUKRI, M., STREINER, D. L. (2011): Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Journal of Clinical Epidemiology*, 64(1). 96–106.
- KOO, T. K., LI, M. Y. (2016): A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2). 155–163.
- KRIPPENDORFF, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, 30(3). 411–433.
- LANDERS, R. (2015). *Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS*, The Winnower 2: e143518. 81744. doi: 10.15200/winn. 143518.81744 (Letöltés ideje: 2019. január 11.)
- LOMBARD, M., SNYDER-DUCH, J., BRACKEN, C. C. (2002): Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, 28(4). 587–604.
- MÉREI, F. (1989). *Társ és csoport*. Akadémiai Kiadó, Budapest.
- MORETTI, F., VAN VLIET, L., BENSING, J., DELEDDA, G., MAZZI, M., RIMONDINI, M., ZIMMERMANN, C., FLETCHER, I. (2011): A standardized approach to qualitative content analysis of focus group discussions from different countries. *Patient Education and Counseling*, 82(3). 420–428.
- PARTEN, M.B. (1932): Social participation among pre-school children. *The Journal of Abnormal and Social Psychology*, 27(3). 243–269.
- POWELL, A. C., TOROUS, J., CHAN, S., RAYNOR, G. S., SHWARTS, E., SHANAHAN, M., LANDMAN, A. B. (2016): Interrater reliability of mHealth app rating measures: analysis of top depression and smoking cessation apps. *JMIR mHealth and uHealth*, 4(1). e15. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4766362> (Letöltés ideje: 2018. január 25.)
- SÁNTHA, K. (2012): Numerikus problémák a kvalitatív megbízhatósági mutatók meghatározásánál. *Iskolakultúra*, 12(3). 64–73.
- SZOKOLSZKY, Á. (2004): *Kutatómunka a pszichológiában*. Osiris Kiadó, Budapest.
- UPTON, P. (2011): *Developmental Psychology*. SAGE, Los Angeles – London – New Delhi – Singapore – Washington DC.

MELLÉKLETEK

1. melléklet. A végleges kategóriarendszer kategóriáinak leírása (Böddi, 2017: 73–74)

„1. Tevékenység

Eszik-iszik

Alkot: rajzol, fest, kézműveskedik, valamilyen tárgyat készít

Játszik: gyakorló (funkcionális) játékot játszik (egyszerű, ismétlődő mozdulatokat végez), konstruáló játékot játszik (tárgyakkal manipulál, hogy létrehozzon valamit, építsen, pl. kockákkal épít, legózik), fantáziajátékot (pl. mintha-játékot, szerepjátékot), szabályjátékot játszik.

2. A tevékenység társas szintje (Parten, 1932, idézi Harper és McCluskey, 2002, 2003 nyomán):

Magányos tevékenység: A gyermek másoktól jóval távolabb egyedül van, egyedül folytat valamilyen tevékenységet, amely eltér a viszonylagos közelségben látható gyermekek tevékenységeitől. Játék esetében egyedül játszik játékokkal, amelyek eltérnek a viszonylagos közelben látható társak játékaitól. Nincs interakcióban mással, sem párhuzamos játékkal (hasonló vagy ugyanolyan tevékenység, játék, mint más gyermeké, de nincsenek interakcióban), sem szemmel láthatóan mások játékának figyelésével nem foglalja el magát. Beleértendő a magányos ábrándozás, elgondolkodás, csakúgy, mint más magányos tevékenységek, amelyek során másoktól távol egyedül van. Az egyik tevékenységből másikba váltás nem tekintendő magányos tevékenységnek, pusztán azért, mert azt egyedül teszi a gyermek (pl. átmegy valahova).

Interakció, interaktív tevékenység: A gyermek olyan viselkedése, amely egy felnőtt vagy egy társ felé irányul és választ, reakciót vált ki, vagy ez a célja. Ilyen reakció például a beszéd, a másik megérintése, egy tárgy odaadása, valami megmutatása neki, figyelmének felkeltése, kooperáció egy közös tevékenységben. Illetve idetartozik, ha a megfigyelt gyermek reagál, választ ad a másik személy ilyen jellegű kezdeményezésére. Tehát verbális és nem verbális kapcsolat látható a személyek között. Játék esetében hasonló vagy ugyanazon játékokkal játszanak, egy játékban vesznek részt. Ebbe a kategóriába értendő, ha a gyermek a többiekkel párhuzamos tevékenységet végez (pl. asztalnál külön-külön asztali játékkal játszanak, rajzolnak, egyénileg készítenek valamit), de az egyéni tevékenysége mellett periodikusan megjelennek a társas interakció jelei (pl. a tevékenység közbeni beszélgetés, testi kontaktus, közös fókuszú figyelem).

Alkategória: az Interakció jellege:

- kétszemélyes interakció gyermekkel
- kétszemélyes interakció felnőttel
- testi kontaktus a felnőttel
- **csak gyermekekkel van interakcióban van (Gyerek):** ebbe a kategóriába tartozik, ha *felnőtt jelenléte nélkül*, csak gyermekekkel van interakcióban a megfigyelt gyermek. Ide

(is) tartozik az is, ha kétszemélyes interakcióban van egy gyermekkel és az is, ha gyermekcsoporttal van interakcióban (kettő vagy kettőnél több gyermekkel)

- **felőtt van az interakcióban (Felőtt):** ebbe a kategóriába tartozik, ha *felőtt jelenlétben* van interakcióban a megfigyelt gyermek. Ide (is) tartozik az is, ha kétszemélyes interakcióban van egy felőtttel és az is, ha felőtt-gyermekcsoporttal van interakcióban (azaz az interakcióban részt vesz a felőttön és a megfigyelt gyermekén kívül még egy vagy több gyermek)”