

Endurprófunaréiðanleiki WPPSI-R^{IS}¹

Eva Gunnarsdóttir, Einar Guðmundsson
Háskóli Íslands

Endurprófunaréiðanleiki WPPSI-R^{IS} var athugaður í hópi 30 íslenskra leikskólabarna á aldrinum fjögurra ára og sex mánaða til fimm ára og tíu mánaða. WPPSI-R^{IS} var að meðaltali lagt fyrir með 24 daga millibili. Stöðugleiki verklegrar greindartölu mældist 0,79, munnlegrar greindartölu 0,91 og heildartölu greindar 0,86. Stöðugleiki undirprófa mældist á bilinu 0,48 til 0,83. Meðalhækkun mælitalna á einstökum undirprófum við endurprófun var minnst 0,1 stig (Myndfletir) en mest 1,5 stig (Litafletir). Meðaltal heildartölu greindar hækkaði um 5,8 stig á milli tveggja fyrirhlagna prófsins. Við endurprófun nam hækkun á verklegri greindartölu 6,3 stigum en 3,6 stigum á munnlegri greindartölu. Niðurstöðurnar benda til þess að WPPSI-R^{IS} veiti nákvæma og stöðuga mælingu á greindarþroska íslenskra barna yfir nokkurra vikna tímabil og svipar til niðurstaðna sams konar athugunar á stöðugleika WPPSI-R í Bandaríkjunum.

Mikilvægar ákvarðanir eru oft teknar á grundvelli niðurstaðna greindarprófa. Jafnframt er algengt að sama barn sé prófað oftar en einu sinni í tilteknum greiningaraðstæðum. Notagildi niðurstaðna tveggja fyrirhlagna sama prófs byggjast á því að mælingarnar séu hvort tveggja í senn nákvæmar og stöðugar.

Endurprófunaréiðanleiki (*test-retest reliability*) er reiknaður til að meta stöðugleika mælitækis frá einum tíma til annars. Hin hefðbundna leið við að afla upplýsinga um endurprófunaréiðanleika er að leggja sama mælitæki fyrir sömu einstaklinga tvisvar sinnum. Yfirleitt líður ekki langur tími á milli fyrirhlagna, algengt er að hann sé á bilinu tvær til fjórar vikur (Kamphaus, 1993; Sattler, 2001). Að báðum fyrirhlögnum loknum er fylgni reiknuð á milli heildartölu, prófþátta og undirprófa tiltekens prófs. Fylgnistuðullinn gefur til kynna stöðugleika niðurstöðu prófsins frá einum tíma til annars. Því hærri sem fylgnin er þeim mun ónæmari eru niðurstöður mælitækis fyrir áhrifum villuþátta (Anastasi og Urbina, 1997).

Ýmsir þættir geta haft áhrif á stöðugleika mælinga frá einum tíma til annars. Allar breytingar, sem hafa áhrif á frammistöðu próftaka við eina prófun en ekki aðra, hafa áhrif á stöðugleika endurtekinna mælinga. Dæmigerðir þættir af þessu tagi eru breytingar á tilhögun fyrirhlagnar, til dæmis þegar nýr prófandi hefur umsjón með seinni fyrirhlögn, prófað er á mismunandi stöðum og á ólíkum tímum dags. Breytilegir þættir sem snerta próftaka, eins og til dæmis einbeiting, þreyta, feimni og kvíði, hafa einnig áhrif. Að auki hefur lengd tímabils milli fyrirhlagna og aldur próftaka áhrif (Sattler, 2001).

Almennt á við að eftir því sem tímabilið, sem líður á milli fyrirhlagna, því hærri verður endurprófunaréiðanleikinn. Rannsókn Schuerger og Witt (1989) leiddi til að mynda í ljós hækkanði endurprófunaréiðanleika einstaklingsgreind arprófa eftir því sem hlé á milli fyrirhlagna var styttra. Annar þáttur sem hefur áhrif á endurprófunaréiðanleika er aldur próftaka. Meiri líkur eru á hærri fylgni þegar próftakar eru eldri en yngri (Anastasi og Urbina, 1997). Þessu til

¹Fyrirspurnum um efni greinarinnar og óskum um sérprent skal beina til Evu Gunnarsdóttur (evagunna@hotmail.com) og Einars Guðmundssonar (eing@hi.is).

stuðnings má aftur nefna rannsókn Schuerger og Witt (1989) þar sem fram kom meiri stöðugleiki prófnidurstaðna hjá unglingum og fullorðnum en börnum.

Þegar tímabilið, sem líður á milli fyrirlagna, er stutt aukast líkur á að þjálfun hafi áhrif á niðurstöður við endurprófun. Þjálfunaráhrif vísa til breytinga á frammistöðu sem verða þegar próftaki hefur reynslu af prófverkefnum (Kantowitz, Roediger III og Elmes, 1994; Sattler, 2001). Þjálfunaráhrif geta torveldað túlkun á niðurstöðum endurtekinnna mælinga því erfitt er að segja til um hvort hærri einkunn við endurprófun er afleiðing íhlutunar eða náms af einhverju tagi eða hvort hún stafi einfaldlega af þjálfun próftaka í að leysa verkefni prófsins. Þjálfunaráhrif aukast venjulega eftir því sem tímabilið, sem líður á milli fyrirlagna, er styttra. Einnig geta þjálfunaráhrif verið breytileg eftir verkefnum í prófi, aldri og frammistöðu próftaka (Sattler, 2001). Börn sem til dæmis fá lága einkunn við fyrstu prófun hafa tilhneigingu til að hækka sig í einkunn við endurprófun en á móti hafa börn sem hljóta háa einkunn við fyrri prófun tilhneigingu til að lækka í einkunn við endurprófun. Annar þáttur sem getur haft áhrif á niðurstöður endurprófunar, þegar stuttur tími líður á milli fyrirlagna, er minnisstarf. Ef endurprófunarhlé er stutt aukast líkur á að próftaki muni svör sín frá fyrri prófun og þar af leiðandi er líklegra að sama svarmynstur komi fram við seinni prófun. Það veldur því að niðurtöður mælinga verða háðar innbyrðis og fylgnin milli þeirra þar af leiðandi óeðlilega há (Anastasi og Urbina, 1997).

Endurskoðuð útgáfa á greindarprófi David Wechsler (1989, 1990) handa börnum á leikskóla- og grunnskólaaldri (WPPSI-R) er mikið notuð við greiningu þroskafrávika barna á Íslandi og erlendis. Endurprófunar-áreiðanleiki WPPSI-R var kannaður þegar prófið var staðlað í Bandaríkjunum. Í handbók með prófinu kemur fram að 175 börn voru prófuð tvisvar með þriggja til sjö vikna millibili. Fylgni endurtekinnna fyrirlagna var 0,87 fyrir verklega greindartölu, 0,89 fyrir munnlega greindartölu og 0,91 fyrir heildartölu greindar.

Fylgni undirprófa var á bilinu 0,53 (Hlutaröðun) til 0,81 (Ófullgerðar myndir) (Wechsler, 1990). Heildartala greindar hækkaði að meðaltali um 5,1 stig milli fyrirlagna og reyndist hækkunin tölfraðilega marktæk. Skýringin á hækkun meðaltals heildartölnnar við síðari fyrirlögn á sennilega rætur að rekja til þjálfunaráhrifa. Breytingar á mælitölum verklegs og munnlegs þáttar við endurprófun voru einnig tölfraðilega marktækar. Hækkunin er meiri á verkþætti eða 6,3 stig að meðaltali en munnlegum þætti þar sem hækkunin nam að meðaltali 2,8 stigum.

Aðrar rannsóknir á endurprófunar-áreiðanleika WPPSI-R eru af skornum skammti erlendis. Fáeinir rannsóknir hafa þó verið gerðar á stöðugleika fyrstu útgáfu prófsins, WPPSI (Wechsler, 1967). Razavieh og Shahim (1990) gerðu athugun á endurprófunar-áreiðanleika WPPSI í tengslum við stöðlun prófsins í Íran. Prófið var lagt fyrir 30 börn á aldrinum 4:6 til 6:0 ára með 39 daga hléi að meðaltali. Niðurstöður rannsóknarinnar leiddu í ljós fylgnina 0,87 milli verklegra greindartalna við endurteknar prófanir, 0,76 fyrir munnlega greindartölu og 0,83 fyrir heildartölu greindar. Endurprófunar-áreiðanleiki undirprófa var lægstur 0,27 (Líkingar) en hæstur 0,87 (Völundarhús).

Í annari athugun á stöðugleika WPPSI var 21 þýskt barn á aldrinum 4:4 til 5:4 ára prófað samtals þrisvar sinnum. Tími milli fyrirlagna var um það bil sex mánuðir. Stöðugleiki verklegrar, munnlegrar og heildartölu greindar reyndist vera 0,79, 0,69 og 0,84. Endurprófunar-áreiðanleiki undirprófa mældist frá 0,20 (Líkingar) til 0,78 (Litafletir) (Neyens og Aldenkamp, 1996).

Nýlega kom út íslensk stöðlun á endurskoðaðri útgáfu af greindarprófi David Wechsler handa börnum á leikskóla- og grunnskólaaldri (Einar Guðmundsson og Hólmfríður Ólafsdóttir, 2003). Í handbók með íslenskri útgáfu prófsins koma fram ýmsir eiginleikar prófsins hérlendis. Almennt á við að próffræðilegir eiginleikar WPPSI-R^{IS} eru hliðstæðir eiginleikum prófsins erlendis. Á þessu eru þó nokkrar undantekningar. Það á til dæmis almennt við um helmingunar-áreiðanleika

(*split-half reliability*) prófsins að hann er lægri héraendis en í Bandaríkjunum. Aftur á móti er þáttbygging prófsins sú sama á Íslandi og í Bandaríkjunum.

Í ljósi fyrirbyggjandi gagna um helmingunaráreiðanleika prófsins á Íslandi og í Bandaríkjunum þótti ástæða til að athuga sérstaklega héraendis hvort það sama ætti við um endurpröfunaráreiðanleika prófsins. Þessar upplýsingar skipta máli fyrir notkun og túlkun niðurstöðu prófsins héraendis. Jafnframt er mikilvægt að auka þekkingu héraendis á því hversu vel próffræðingin leikar, sem gefnir eru upp í erlendum handbókum matstækja, eiga við héraendis.

Aðferð

Þátttakendur

Börn í fjórum leikskólum í Reykjavík, sem valin voru af handahófi úr hverfi 101, tóku þátt í rannsókninni. Samtals tóku 30 börn, 15 stúlkur og 15 drengir, þátt. Yngsti þátttakandinn var fæddur í ágúst 1998 og sá elsti í maí 1997. Meðalaldur stúlkna og drengja var fimm ár og einn mánuður við fyrri prófun, staðalfrávik fjórir mánuðir. Meðalaldur stúlkna við seinni prófun var fimm ár og einn mánuður en drengja fimm ár og tveir mánuðir, staðalfrávik fjórir mánuðir.

Mælitæki

Í íslenskri handbók WPPSI-R^{IS} eru ýtarlegar upplýsingar um uppbyggingu prófsins, fyrirlagnar- og matsreglur, áreiðanleika og réttmæti þess héraendis (Einar Guðmundsson og Hólmfríður Ólafsdóttir, 2003). Samtals eru 12 undirpróf í WPPSI-R^{IS}. Tvö þeirra (Setningar og Dýrahús) eru varaskeifur sem ekki eru lögð fyrir reglulega. Eftirtalin fimm munnleg undirpróf voru lögð fyrir í rannsókninni: Þekking, Skilningur, Reikningur, Orðskilningur og Líkingar. Fimm verkleg undirpróf voru lögð fyrir: Hlutaröðun, Myndfletir, Litafletir, Völundarhús, Ófullgerðar myndir.

Meðaltal helmingunaráreiðanleika fyrir undirpróf er á bilinu 0,55 (Hlutaröðun) til 0,82

(Orðskilningur). Meðaltal áreiðanleikastuðla fyrir prófhluta og heildartölu greindar er á bilinu 0,88 (verkleg greindartala) til 0,92 (heildartala greindar).

Fjöldi erlendra réttmætrisrannsókna liggur fyrir um WPPSI-R (Sattler, 2001). Í íslenskri handbók kemur fram að þáttbygging prófsins er sú sama héraendis og erlendis. Meðaltal mælitalna undirprófa á WPPSI-R^{IS} er 10 og staðalfrávik 3. Meðaltal munnlegrar, verklegrar og heildartölu greindar er 100 og staðalfrávik 15. Í íslenskri handbók með prófinu kemur fram að almennt eru gólf- og rjáfuráhrif í prófinu minni héraendis en í Bandaríkjunum.

Framkvæmd

Foreldrum barna sem fædd voru seinni hluta ársins 1997 og fyrri hluta ársins 1998 í fjórum leikskólum í Reykjavík, sem valdir voru að handhófi í hverfi 101, voru send bréf þar sem óskað var eftir skriflegu leyfi fyrir þátttöku barnanna í rannsókninni. Starfsfólk leikskólanna sá um dreifingu bréfanna og innheimtur undirskrifta. Börn þeirra foreldra, sem skiluðu undirrituðu bréfi til starfsmanna leikskólans, voru skráð sem þátttakendur í rannsókninni.

WPPSI-R^{IS} var tvisvar lagt fyrir hvert barn í rannsókninni. Prófið var lagt fyrir samkvæmt íslenskri handbók sem var notuð við stöðun prófsins. Að meðaltali liðu 24 dagar á milli fyrirlagna. Styst leið 21 dagur en lengst liðu 35 dagar milli fyrirlagna. Cand. psych. nemi í sálfræði sá um allar fyrirlagnir sem fóru fram á leikskóla barnanna. Að lokinni fyrirlögn fengu börnin verðlaun (límmiða).

Niðurstöður

Reiknuð voru meðaltöl og staðalfrávik verklegrar, munnlegrar og heildartölu greindar við báðar prófanir. Við fyrri prófun var meðaltal munnlegrar greindartölu hærra en meðaltal verklegrar greindartölu (1. tafla). Við endurpröfun var munnleg greindartala einnig hærri en verkleg greindartala þó munurinn væri minni en við fyrri prófun. Meðaltal heildartölu greindar við fyrri prófun var 104,0 en 109,8 við

1. Tafla. Meðaltal og staðalfrávik greindartalna WPPSI-R^{IS} í tveimur prófunum með 24 daga millibili að meðaltali ($N = 30$).

	Fyrri prófun (FP)		Seinni prófun (SP)		Mismunur (SP – FP)
	M	sf	M	sf	
Verkleg greindartala	101,2	16,6	107,5	15,7	+ 6,3 **
Munnleg greindartala	105,4	16,6	109,0	15,1	+ 3,6 *
Heildartala greindar	104,0	16,6	109,8	15,1	+ 5,8 ***

* $p < 0,05$, ** $p < 0,005$; *** $p < 0,001$.

seinni prófun. Meðaltal heildartölu greindar hækkaði því um 5,8 stig á milli prófana ($t = 3,7$, $p < 0,001$). Við endurprófun nam hækkun á verklegri greindartölu 6,3 stigum ($t = 3,3$, $p < 0,01$) en 3,6 stigum á munnlegri greindartölu ($t = 2,9$, $p < 0,05$).

Meðaltöl og staðalfrávik mælitalna undirprófa við báðar prófanir voru einnig reiknuð. Mælitölur undirprófa við fyrri prófun voru að meðaltali á bilinu 8,8 til 11,5 en 9,8 til 12,3 við

seinni prófun (2. tafla). Meðalhækkun mælitalna á einstökum undirprófum við endurprófun var minnst 0,1 (Myndfletir) en mest 1,5 (Litafletir). Hækkunin var að meðaltali meiri á verklegum en mállegum undirprófum. Tölfræðilega marktækur munur var á meðaltölum mælitalna þriggja undirprófa: Litafletir ($t = 3,6$, $p < 0,001$), Ófullgerðar myndir ($t = 2,3$, $p < 0,05$) og Þekking ($t = 2,6$, $p < 0,05$).

2. Tafla. Meðaltal og staðalfrávik mælitalna undirprófa WPPSI-R^{IS} í tveimur prófunum með 24 daga millibili að meðaltali ($N = 30$).

	Fyrri prófun (FP)		Seinni prófun (SP)		Mismunur (SP – FP)
	M	sf	M	sf	
Hlutaröðun	8,8	3,4	9,8	3,6	+ 1,0
Myndfletir	11,0	2,6	11,1	2,9	+ 0,1
Litafletir	10,8	3,3	12,3	2,5	+ 1,5**
Völundarhús	9,1	3,5	10,1	3,1	+ 1,0
Ófullgerðar myndir	10,9	2,5	11,6	2,3	+ 0,7*
Þekking	10,2	2,9	11,2	2,9	+ 1,0*
Skilningur	11,1	3,2	11,3	2,8	+ 0,2
Reikningur	10,2	3,4	10,6	3,5	+ 0,4
Orðskilningur	11,5	3,1	11,9	2,6	+ 0,4
Líkingar	10,9	2,6	11,5	3,0	+ 0,6

* $p < 0,05$; ** $p < 0,001$.

3. Tafla. Endurprófunaráreiðanleiki (Pearson fylgni) við endurteknar prófanir milli einstakra undirprófa og greindartalna WPPSI-R á Íslandi (N = 30) og í Bandaríkjunum (N = 175).

	WPPSI-R ^{IS} Ísland	WPPSI-R USA
Hlutaröðun	0,62	0,53
Myndfletir	0,67	0,66
Litafletir	0,74	0,77
Völundarhús	0,48	0,55
Ófullgerðar myndir	0,77	0,81
Pekking	0,73	0,78
Skilningur	0,76	0,77
Reikningur	0,70	0,71
Orðskilningur	0,80	0,71
Líkingar	0,83	0,73
Verklegr greindartala	0,79	0,87
Munnleg greindartala	0,91	0,89
Heildartala greindar	0,86	0,91

Endurprófunaráreiðanleiki var reiknaður með Pearson fylgni. Í 3. töflu koma fram upplýsingar um fylgni milli greindartalna og undirprófa við tvær prófanir sömu barna. Til samanburðar eru birtar í töflunni hliðstæðar upplýsingar fyrir WPPSI-R í Bandaríkjunum. Fylgnin mældist 0,79 fyrir verklega greindartölu, 0,91 fyrir munnlega greindartölu og 0,86 fyrir heildartölu greindar. Fylgni einstakra undirprófa var lægst 0,48 (Völundarhús) en hæst 0,83 (Líkingar).

Umræða

Athugun á endurprófunaráreiðanleika WPPSI-R á Íslandi leiddi í ljós að stöðugleiki verklegrar greindartölu nam 0,79, munnlegrar greindartölu 0,91 og heildartölu greindar 0,86. Heildartölur prófsins eru því almennt stöðugar við endurprófun. Samanborið við endurprófunaráreiðanleika greindartalna WPPSI-R í Bandaríkjunum er hann heldur lægri á Íslandi. Hugsanleg skýring á þessu er munur á fjölda barna í úrtökunum tveimur þar sem bandarískt úrtak innihélt tæplega sex

sinum fleiri þátttakendur en úrtakið hérlendis. Þessum niðurstöðum svipar til samanburðar á helmingunaráreiðanleika WPPSI-R hérlendis og í Bandaríkjunum sem sýndi almennt lægri áreiðanleika á Íslandi en í Bandaríkjunum.

Endurprófunaráreiðanleiki einstakra undirprófa á Íslandi dreifist frá 0,48 til 0,83. Hann er því allt frá því að vera óviðunandi til þess að vera viðunandi. Hann er lægstur á undirprófinu Völundarhús, líkt og í Bandaríkjunum, en hæstur á undirprófinu Líkingar. Alls mælast þrjú undirpróf með endurprófunaráreiðanleika undir 0,70 (Hlutaröðun, Myndfletir, Völundarhús). Að meðaltali er endurprófunaráreiðanleiki verklegra undirprófa lægri en munnlegra undirprófa. Þessar niðurstöður samræmast vel bandarískum niðurstöðum sama efnis.

Marktæk hækkun er á meðaltali allra greindartalna við endurprófun. Meiri hækkun er á meðaltali verklegrar greindartölu en meðaltali munnlegrar greindartölu. Mælitölur verklegra undirprófa hækka sömuleiðis meira

að meðaltali en mælitölur munnlegra undirprófa við endurprófun. Tölfræðilega marktækur munur reyndist þó aðeins vera á meðaltölum þriggja undirprófa. Þessar niðurstöður eru samhljóma bandarískri athugun á stöðugleika WPPSI-R sem einnig leiddi í ljós að verkleg undirpróf eru næmari en munnleg undirpróf fyrir þjálfunarhrifum þegar skammur tími líður á milli fyrirllagna.

Úrtak rannsóknarinnar voru börn á aldri- num fjögurra til fimm ára. Strangt til tekið ættu niðurstöður þessarar rannsóknar aðeins að gilda fyrir þann aldurshóp. Þegar horft er til helmingunaráreiðanleika WPPSI-R^{IS} eftir aldurshópum kemur í ljós að áreiðanleikastuðlarnir dreifast mismunandi eftir því hvort litið er til heildartalna eða undirprófa (Einar Guðmundsson og Hólfríður Ólafsdóttir, 2003). Helmingunaráreiðanleiki heildartalna er svipaður eftir aldurshópum. Aftur á móti getur breytileiki verið nokkur eftir undirprófum. Ekki er ósennilegt að sama eigi við um endurprófunaráreiðanleika prófsins. Um þetta er þó ekki hægt að fullyrða. Varfærnasta leiðin við alhæfingu þessara niðurstaðna á allt aldursbil prófsins (3 – 7 ára) er að gera ráð fyrir að endurprófunaráreiðanleiki heildartalna sé svipaður eftir aldursbilum en breytilegur fyrir undirprófum.

Abstract

Test-retest reliability of the Icelandic standardized version of WPPSI-R was investigated in a sample of 30 five years, one month old Icelandic children. On average 24 days elapsed between administrations of the test. Stability coefficient of Performance IQ was 0,79, Verbal IQ 0,91 and IQ 0,86. Stability coefficients of performance and verbal subtests ranged from 0,48 to 0,83. Lowest mean increase of scaled scores between administrations on individual subtest was 0,1 (Geometric Design) and highest 1,5 scaled scores (Block Design). Mean scaled score increase between testings was 5,8 (IQ), 6,3 (Performance IQ) and 3,6 (Verbal IQ). The results indicate that the Icelandic standardization of WPPSI-R has

sufficient test-retest reliability, similar to that found in the USA.

Heimildaskrá

- Anastasi, A. og Urbina, S. (1997). *Psychological testing* (7. útgáfa). New Jersey: Prentice Hall.
- Einar Guðmundsson og Hólfríður Ólafsdóttir (2003). *Greindarpróf David Wechsler handa börnum á leikskóla- og grunnskólaaldri. Endurskoðuð útgáfa*. Reykjavík: Námsmatsstofnun.
- Kamphaus, R.W. (1993). *Clinical assessment of children's intelligence*. Boston: Allyn and Bacon.
- Kantowitz, B.H., Roediger III, H.L. og Elmes, D.V. (1994). *Experimental psychology* (5. útgáfa). New York: West Publishing Company.
- Neyens, L.G.J. og Aldenkamp, A.P. (1996). Stability of cognitive measures in children of average ability. *Child Neuropsychology*, 2, 161-170.
- Razavieh, A. og Shahim, S. (1990). Retest reliability of the Wechsler Preschool and Primary Scale of Intelligence restandardized in Iran. *Psychological Reports*, 66, 865-866.
- Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4. útgáfa). San Diego: Jerome M. Sattler, Publisher.
- Schuerger, J.M. og Witt, A.C. (1989). The temporal stability of individually tested intelligence. *Journal of Clinical Psychology*, 45, 294-302.

- Wechsler, D. (1967). *Manual for the Wechsler Preschool and Primary Scale of Intelligence*. New York: The Psychological Corporation.
- Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence-Revised*. San Antonio: The Psychological Corporation.

- Wechsler, D. (1990). WPPSI-R^{UK}. *Wechsler Preschool and Primary Scale of Intelligence-Revised*. London: The Psychological Corporation.