

Måleegenskaper ved de norske versjonene av Ravens matriser [Standard Progressive Matrices (SPM)/Coloured Progressive Matrices (CPM)]

Fredrik Helland-Riise¹ [fredrik.helland-riise@cemo.uio.no]

Monica Martinussen² [monica.martinussen@uit.no]

PsykTestBarn 2017, 2:2

ISSN 1893-9910

Mottatt dato 7. september 2017

Publisert dato 18. desember 2017

Artikkelens URL <https://psyktestbarn.r-bup.no/no/artikler/ravens-matriser>

Som alle artikler i PsykTestBarn, kan denne fagfelleverderte artikkelen lastes ned, skrives ut og distribueres fritt for alle slags formål på følgende vilkår: korrekt referanse skal oppgis (se under), ingen kommersiell bruk og ingen bearbeidelse av tekst eller innhold.

¹ CEMO - Centre for Educational Measurement, Universitetet i Oslo

² RKBUNord, UiT Norges arktiske universitet.

Denne artikkelen skal siteres på følgende måte:

Helland-Riise, F. & Martinussen, M. (2017). Måleegenskaper ved de norske versjonene av Ravens matriser [Standard Progressive Matrices (SPM)/Coloured Progressive Matrices (CPM)]. *PsykTestBarn*, 2:2

Sammendrag

Beskrivelse. Ravens matriser består av flere varianter der Coloured Progressive Matrices (CPM) (5–11 år) og Standard Progressive Matrices (SPM) (8–65 år) kan anvendes for barn og unge. Testen ble utviklet av John C. Raven, og SPM ble først publisert i 1938, mens CPM-versjonen ble publisert i 1998. Testen er ment å måle non-verbal intelligens eller evne til abstrakt resonnering ved at barnet skal finne systemet bak hvordan ulike geometriske figurer er satt opp. Begge versjonene av Raven har oppgaver med økende vanskelighetsgrad organisert i ulike sett. Testen skåres ved at antallet riktige svar summeres opp til en totalskår. Dette kan konverteres til prosentiler basert på normtabellene i manualene. Testen administreres vanligvis uten tidsbegrensning for barn og unge, og det stilles krav om at de som skal anvende testen er sertifisert for å anvende evnetester eller er psykologer. Pearson Assessment har rettighetene til salg og distribusjon av testen i Skandinavia og internasjonalt (pearsonassessment.no).

Litteratursøk. Til sammen ble 15 norske og 24 svenske/danske artikler inkludert i oppsummeringen. Ingen av disse var rene psykometriske studier, men stort sett studier der Raven ble brukt for å måle intelligens enten som utkommevariabel, kontrollvariabel eller for å beskrive gruppen. Litt under halvparten av studiene var basert på kliniske grupper (for eksempel autisme, epilepsi eller døve), men de øvrige studiene var basert på skolebarn fra 5–16 (SPM) og 5–7 (CPM) år.

Psykometri. Ingen av studiene hadde gjennomført adekvate studier av testens reliabilitet og det er heller ikke gjennomført normstudier basert på norske eller svensk/danske utvalg. Resultatene fra studiene støtter begrepsvaliditeten til testen som et godt mål på abstrakt resonneringsevne ved generelt høye korrelasjoner med andre kognitive tester.

Konklusjon. Det er god dokumentasjon på testens begrepsvaliditet, men det mangler studier av testens reliabilitet basert på norske eller svenske/danske utvalg. Det er heller ikke normstudier fra Skandinavia, noe som er problematisk ved klinisk bruk av testen.

Abstract

Description. There are different versions of Ravens Matrices including Coloured Progressive Matrices (CPM) (5–11 years) and Standard Progressive Matrices (SPM) (8–65 years) which may be used for children and adolescents. The test was first developed by John C. Raven, and the SPM was published in 1938, and the CPM in 1998. The test is assumed to measure non-verbal intelligence or abstract reasoning ability with items consisting of geometrical figures where the task is to figure out the system underlying the figures. Both versions of Raven have items with increasing difficulty organized in different sets. The test is scored in terms of the number of correct answers, which is summed to a total score. The total score can be converted to percentiles based on norm tables included in the manuals. The test is usually administered without time constraints for children and adolescents, and the test user is required to be a psychologist or a certified user of ability tests.

Pearson Assessment holds the copyright to Ravens Matrices, and is responsible for sale and distribution of the test in Scandinavia and internationally (pearsonassessment.no).

Literature search. A total of 15 Norwegian and 24 Swedish/Danish publications were included in the review. None of these were psychometric studies, but mainly studies where Raven had been used to measure intelligence either as an outcome variable, control variable or to describe the group. Approximately half of the included studies were based on clinical groups (e.g., autism, epilepsy or deaf), while the remaining studies included school children of 5–16 (SPM) and 5–7 (CPM) years.

Psychometrics. None of the included studies had performed adequate studies of test reliability and there were no norm studies based on Norwegian or Swedish/Danish samples. The findings support the construct validity of the test as a good measure of abstract reasoning ability by in general high correlations with other cognitive tests.

Conclusion. There is solid evidence supporting the construct validity of the test, but studies examining test reliability using Scandinavian samples are lacking. There are no norm studies from Scandinavia, which is problematic for clinical and applied test use.

Innledning

Ravens matriser består av tre ulike tester som alle måler abstrakt resonneringsevne, og refereres ofte til som mål på nonverbale evner, flytende intelligens, eller rett og slett bare intelligens. Testene anvendes klinisk, til forskning og for seleksjon av personell. De tre utgavene er Coloured Progressive Matrices (CPM), Standard Progressive Matrices (SPM) og Advanced Progressive Matrices (APM). De tre versjonene har delvis overlappende aldersintervall for sine målgrupper men i hovedsak er CPM ment for yngre barn (5–11 år), SPM for barn, ungdommer og voksne (8–65 år), mens APM er best egnet for voksne med generelt høyt evnenivå og omtales derfor ikke videre i denne artikkelen. Pearson Assessment har rettighetene til salg og distribusjon av testen i Skandinavia og internasjonalt (pearsonassessment.no). Pearson stiller krav til hvilken kompetanse kjøpere av de ulike testene må ha, og for Raven er det stilt krav om at man må være autorisert psykolog eller sertifisert bruker av evnetester. Hver versjon består av manual på engelsk, testhefte med oppgaver og svarark. Den første versjonen var SPM som ble publisert i 1938 ved Oxford Psychologists Press Ltd. (Raven, 1941) og betegnes som «classic», mens de nye versjonene ble publisert i 1998 (Raven, 2008b). I tillegg til originalversjonene finnes det parallelle utgaver som kan anvendes ved behov, for eksempel om det er behov for å teste et barn flere ganger.

SPM foreligger også en i plussversjon som innebærer at det er lagt til flere og vanskeligere oppgaver. De internasjonale normene som foreligger på de nye versjonene er i hovedsak fra et britisk utvalg. De ulike Raven-testene kan administreres individuelt eller i gruppe, og er såkalte «power-tester» med testmessig tidskontroll, noe som vil si at man primært er interessert i resonneringsferdigheten (ikke responstiden), og tidsbegrensningen varierer avhengig av hvilke og hvor mange sett som administreres. Testen er tradisjonelt antatt som mindre kulturbundet ved at den er non-verbal (se for eksempel Cattell, 1940). Det er 36 oppgaver i både CPM og grunnversjonen av SPM (60 i plussversjonen) med ulik vanskelighetsgrad. Hver oppgave består av flere ruter (for eksempel 3X3) med mønster der én av rutene mangler. Oppgaven består i å finne den manglende ruten blant flere alternativer (6-8 alternativer) ved å finne det logiske systemet bak mønsteret i de aktuelle rutene. Testen antas å måle evne til å sammenligne og resonnerer ved hjelp av analogier og systematisere spatiale persepsjoner. Vanskelighetsgrad er antatt å avhenge av for eksempel antall elementer, antall regler, reglenes kompleksitet og den perseptuelle organiseringen av oppgavene (Primi, 2001), samt personenes evne til målregulering (Carpenter, Just, & Shell, 1990). Antallet riktig besvarte oppgaver summeres til en totalskår som ved klinisk bruk kan

konverteres til evnenivå (i form av prosentiler) via normtabellene i de respektive manualene.

Raven er brukt i store deler av verden siden den første testen ble publisert i 1938 (Raven, 2000b). Den er også mye brukt i Norge, og ifølge Norsk Psykologforening var testen i 2012 den nest mest brukte kognitive funksjonstesten blant norske psykologer med en brukerandel på hele 38 %, kun slått i popularitet av WAIS-III (Vaskinn & Egeland, 2012). I en studie av skolepsykologer fra 64 land var SPM brukt i 40 % av landene og kun WISC var brukt av flere (80 %) (Oakland, Douglas, & Kane, 2016).

Ulike former for reliabilitet for britiske barn er beregnet som for eksempel split-half reliabilitet på 0,97 for CPM (alder 4-12) og 0,94 for SPM (alder 7-18), samt for SPM test-retest reliabilitet som har blitt rapportert å være på 0,83 (alder 8-18), og for CPM parallell-form-reliabilitet på 0,87 (alder 5-11) (Raven, 2008a, 2008b). Lavest reliabilitet er funnet når testene har vært brukt på svært unge barn. Begrepsvaliditeten er undersøkt gjennom ulike studier der Raven har blitt brukt som kovariat i analyser, hvor andre konstrukter ofte har vært det primære studieobjektet (f.eks. med WISC: r med SPM = 0,70-0,80; r med CPM = 0,50-0,87; Raven, 2008a; Raven, 2008b), og prediktiv validitet ved å undersøke sammenhengen mellom testen og prestasjoner i for eksempel skole og arbeidsliv (se f.eks. Raven, 2000a for en drøftelse). Fra internasjonal forskning er det en allmenn oppfatning at Raven måler sentrale deler av intelligensbegrepet (Marshalek, Lohman, & Snow, 1983). Fordelen med testen er at den er ansett som relativt kulturfri, kan anvendes uavhengig av språk, samt på mange ulike grupper som barn med ulike utviklingsforstyrrelser som autismespekter forstyrrelser, eller barn med spesielle behov, for eksempel døve. Den egner seg ikke som mål på generell intelligens der barnet ikke er i stand til abstrakt resonnering ved hjelp av figurer.

Metode

Bibliotekar Brynhildur Axelsdottir ved Regionsenteret for Barn og Ungdoms Psykisk Helse, Helseregion Øst og Sør, søkte etter dokumentasjon på testens psykometriske egenskaper i databasene, PsycINFO, Medline, Embase, Cochrane Library, Oria

(BIBSYS), Norart, SveMed+, PubMed, CRISStin.no, NORA.no, Forskningsdatabasen.dk og Swepub. Søkedato: 25.02.2016 og oppdateringssøk: 21.03.2017. Søkestrategien er tilgjengelig på <https://psyktestbarn.r-bup.no/no/artikler/ravens-matriser>. Der publikasjoner ikke var tilgjengelige, ble forfatterne forsøkt kontaktet. Da det allerede var mange publikasjoner på Raven, ble det ikke gjort ytterligere forsøk på å kontakte forfatterne.

Vi inkluderte alle publikasjoner av studier som har undersøkt og rapportert minst ett av følgende i skandinaviske utvalg:

- normdata for testen
- reliabilitet: indre konsistens, test-retest, interrater og endringssensitivitet
- validitet: samsvar med liknende testskårer, samsvar med referansestandard eller annet kriterium, og/eller faktorstruktur

I tillegg, og kun for norske versjoner av Raven, inkluderte vi publikasjoner som rapporterte gjennomsnittsskårer og/eller forekomster for henholdsvis generelle populasjoner og kliniske undergrupper.

Begge forfatterne gikk gjennom sammendragene til alle identifiserte publikasjoner, uavhengig av hverandre, etter at dubletter var fjernet. Alle publikasjoner som kunne virke relevante ble bestilt inn i fulltekst, og prosessen over ble gjentatt for fulltekstrapportene.

Begge forfatterne vurderte normering, validitet og reliabilitet, uavhengig av hverandre, ved hjelp av en tilpasset versjon av Test review form and notes for reviewers (European Federation of Psychologists' Association - EFPA, 2013).

Resultater

Litteratursøk

Det norske søket inneholdt totalt 42 artikkelsammendrag. To publikasjoner måtte reklassifiseres som svenske eller danske (hvorav en var duplikat og den andre kunne ekskluderes). Totalt 16 ble ekskludert etter en grovscreening av sammendrag og i noen tilfeller fulltekster. Veldig få studier hadde primærfokus på Raven, og grunnet et

stort antall falske negativer (det sto ingenting i artikkelsammendraget om Raven, selv om artikkelen hadde relevant informasjon om testen), var ikke artikkelsammendraget tilstrekkelig for å ta eksklusjonsbeslutninger. Til sammen 15 publikasjoner ble etter grundigere lesning vurdert til å oppfylle inklusjonskriteriene. De mest vanlige eksklusjonsgrunnene var manglende norsk utvalg ($K = 7$), at målgruppen ikke besto hovedsakelig av barn eller unge ($K = 4$), ikke rapporterte psykometriske data fra Raven ($K = 6$), samt øvrige grunner ($K = 7$). En klausulert masteroppgave lot seg ikke lokalisere.

For det svenske/danske søket ble det funnet til sammen 276 treff. Av disse ble 111 vurdert som aktuelle basert på sammendraget og vurdert i

fulltekst. En gjennomgang av de 111 artiklene i fulltekst resulterte i tilsammen 22 artikler som oppfylte inklusjonskriteriene. Den mest vanlige grunnen for eksklusjon var at artikkelen ikke rapporterte psykometriske data ($K = 36$), deretter at målgruppen ikke var barn og unge ($K = 20$), manglende skandinavisk utvalg ($K = 13$) samt øvrige grunner ($K = 18$) (for eksempel at Raven ikke var brukt, ingen empiriske data eller konferansesammendrag). I tillegg ble ytterligere to artikler (Gustafsson, 1984; Gustafsson & Wolff, 2015) lokalisert via henvisninger i en av de inkluderte artiklene. Dette resulterte sammenlagt i 24 artikler basert på 22 ulike studier. En doktoravhandling lot seg ikke lokalisere.

Tabell 1. Inkluderte norske studier

Referanse	Design	Populasjon	N	Mål	Rapporterte egenskaper, verdier
Bosnes (2005)	Tverrsnitt	Blandet klinisk utvalg	41	SPM (1982/1988/2000)	Middelverdi og spredning, validitet
Elbro and Buch-Iversen (2013)	RCT	6. klassinger, normale barn	236	SPM (1958)	Middelverdi og spredning
Fasting and Lyster (2005)	Kvasiekperiment	Barn med selvrapperte lese- og staveproblemer.	52	SPM (1983)	Middelverdi og spredning
Frostad (1995)	Tverrsnitt	Hørselshemmede barn	268	SPM (1992)	Middelverdier og spredning, normtabeller, kumulative distribusjoner, testdiagnostikk, deskriptiv statistikk.
Green, Tønnessen, Tambs, Thoresen, and Bjertness (2009)	Kvasiekperiment	10.klassinger, normale barn vs grader av lese og skrivevansker	308	SPM (2003)	Middelverdi og spredning, begrepsvaliditet
Høie, et al. (2005)	Tverrsnitt/eksperiment	Barn med epilepsi	309	CPM (1965)	Forekomster i klinisk gruppe
Høie, Sommerfelt, et al. (2006) ^a	Tverrsnitt/eksperiment	Barn med epilepsi	250	CPM (1965)	Begrepsvaliditet (r med psykososiale vansker)
Høie, Mykletun, Waaler, Skeidsvoll, and Sommerfelt (2006) ^a	Tverrsnitt/eksperiment	Barn med epilepsi	241	CPM (1965)	Begrepsvaliditet (regressjon med eksekutive funksjoner)
(Høie, et al., 2008)) ^a	Tverrsnitt/eksperiment	Barn med epilepsi	269	CPM (1965)	Begrepsvaliditet (r med WISC-R)
Landmark and Grinde (1962)	Tverrsnitt	Normale barn	44	SPM set A, AB, B & revised order (1956)	Middelverdi og spredning, begrepsvaliditet, split half reliabilitet
Lervåg, Bråten, and Hulme (2009) ^b	Longitudinell	Normale barn, 1. – 2. klasse	228	CPM (1990)	Middelverdi og spredning, begrepsvaliditet, Cronbachs alpha
Lervåg and Hulme (2010) ^b	Longitudinell	Normale barn, 1. – 2. klasse	228	CPM (1990)	Middelverdi og spredning, validitet
Lyster, Lervåg, and Hulme (2016)	Kontrollgruppe fra eksperiment.	Normale barn, 1. klasse	30	SPM (1988)	Middelverdi og spredning

Solheim (2011)	Tverrsnitt	Normale barn, 5. klasse	217	SPM (1958)	Middelverdi og spredning, begrepsvaliditet
Söderqvist, Nutley, Ottersen, Grill, and Klingberg (2012)	Kvasiekseptiment	Barn med IQ <70	41	CPM (1998)	Middelverdi og spredning

Note. SPM = standard progressive matrices. CPM = coloured progressive matrices. ^a Baserer seg delvis på samme utvalg som Høie, et al. (2005). Metoder og utvalg kan variere mht. tilgjengelige måleinstrumenter, nye deltagere etc., derfor kan tallene variere noe mellom studiene. ^b Baserer seg på samme utvalg som Lervåg, et al. (2009).

Tabell 2. Inkluderte svenske og danske studier

Referanse	Design	Populasjon	N	Mål	Rapporterte egenskaper, verdier
Andersson (2007)	Tverrsnitt	Skolebarn, Sverige ($M = 10$ år, $SD = 0,76$ år)	69	SPM (1976), sett B, C & D	Begrepsvaliditet (r med andre kunnskaps- og kognitive tester)
Andersson (2008)	Tverrsnitt	Skolebarn, Sverige, ($M = 10,3$ år, $SD = 0,58$ år)	141	SPM (1976), sett B, C & D	Begrepsvaliditet (r med andre kunnskaps- og kognitive tester)
Nutley, Darki, and Klingberg (2014)	Longitudinell	Barn, by i Sverige, 6–25 år ($M = 12,25$ år), tilfeldig utvalgt	339	SPM (2003), sett A til D (6 år), A-E for resten av utvalget	Begrepsvaliditet (forskjell mellom to grupper)
Brynskov, et al. (2017)	Tverrsnitt	To grupper danske barn (ASD, $M = 66$ mnd, $SD = 7,2$ mnd) (Vanlige barn, $M = 74,1$ mnd, $SD = 6,7$ mnd)		CPM (Raven, 2008a)(2008)	Begrepsvaliditet (forskjell mellom grupper)
Bylund, et al. (2000)	Longitudinell	Premature barn og kontrollgruppe, Sverige (9 år)	142	SPM (1983)	Begrepsvaliditet (forskjell mellom grupper)
Dahlin (2011)	Tverrsnitt	Skolebarn m/lærevansker, Sverige (9–12 år)	57	CPM (ingen dato)	Begrepsvaliditet (r med andre kunnskaps- og kognitive tester)
Dumontheil and Klingberg (2012)	Longitudinell	Barn og unge (6–16 år) fra by i Sverige	246	SPM (1998), sett A-D (yngste) og A-E (eldste)	Begrepsvaliditet (sammenheng med matematiske ferdigheter)
Fischbein, Guttman, and Nathan (1999)	Tvillingstudie	Barn i Stockholm (12 år)	73	SPM (ingen dato), sett A-E	Begrepsvaliditet (r med andre kognitive tester)
Forssman, Eninger, Tillman, Rodriguez, and Bohlin (2012)	Tverrsnitt	Ungdommer, tilfeldig utvalg, 3 byer i Sverige ($M = 14,8$ år, $SD = 0,89$)	120	SPM (1977), sett A-E	Begrepsvaliditet (r med andre kognitive tester)
Gellert and Elbro (2013)	Tverrsnitt/longit.	Skolebarn fra København ($M = 9,6$ år, $SD = 0,3$ år)	99	CPM, sett A, AB og B (1990)	Begrepsvaliditet (r med andre kognitive tester)
Gustafson, Samuelsson, Johansson, and Wallmann (2013)	Tverrsnitt	Skolebarn, Sverige med og uten lesevaner (4.klasse)	72	CPM, sett A, AB og B (1990)	Begrepsvaliditet (r med andre kognitive tester). + Alpha er rapportert (0.72)
Gustafsson (1984)	Tverrsnitt	Skolebarn Sverige (12 år)	1224	SPM (ingen dato)	Begrepsvaliditet (faktoranalyse)

Gustafsson and Wolff (2015)	Tverrsnitt	Barn, Sverige (4 år)	364	CPM, sett A, AB og B (1990)	Begrepsvaliditet (faktoranalyse)
Nordberg, Dahlgren Sandberg, and Miniscalco (2015)	Tverrsnitt	Barn med CP og talevansker ($M = 11$ år, $SD = 1,4$ år)	15	CPM, sett A, AB og B (1990)	Begrepsvaliditet (r med andre kognitive tester) samt sml med normer
Nordberg, Miniscalco, and Lohmander (2014)	Tverrsnitt	Barn med CP ($M = 11,2$ år)	19	CPM, sett A, AB og B (1990)	Begrepsvaliditet (r med andre kognitive tester) samt sml med normer
Olsson, Ostergren, and Traff (2016)	Tverrsnitt	Barn, Sverige med og uten dyskalkuli	72	SPM (1976), sett B, C & D	Cronbachs alpha rapportert (0,80)
Samuelsson, et al. (2006)	Longitudinell	Premature barn med lav fødselsvekt + kontroll (9 år)	142	SPM (1983)	Begrepsvaliditet (gruppeforskjeller)
Skagerlund and Traff (2016b)	Tverrsnitt	Skolebarn, Sverige (9,7 år) inkl gruppe med spesialundervisning	133	SPM (1976), sett B, C & D	Begrepsvaliditet (r med andre kognitive tester)
Skagerlund and Traff (2016a)	Tverrsnitt	Skolebarn med og uten dyskalkuli, Sverige ($M = 11,75$ år)	77	SPM (1976), sett B, C & D	Cronbachs alpha rapportert (0,87)
Svensson and Jacobson (2006)	Longitudinell	Skolebarn, Sverige med og uten lesevaner (8–9 år)	70	CPM (1965). Ravens SPM (1995), sett B, C & D	Begrepsvaliditet (forskjeller mellom grupper)
Thirus, Starbrink, and Jansson (2016)	Tverrsnitt	Videregående elever, Stockholm ($M = 17,3$, $SD = 0,67$)	21	SPM, sett A-E	Begrepsvaliditet (forskjeller mellom grupper)
Träff (2013)	Tverrsnitt	Skolebarn, Sverige ($M = 142$ mnd, $SD = 8,58$)	134	SPM (1995), sett B, C & D	Begrepsvaliditet (r med andre kognitive tester)
Vejleskov (1968)	Tverrsnitt	Skolebarn, Danmark (10–11 år)	628	SPM (1958), sett B, C, D & E	Item-analyse
Wolff and Gustafsson (2015)	Tverrsnitt	Barn, Sverige (4 år)	364	CPM, sett A, AB & B (1990).	Begrepsvaliditet (faktoranalyse)

Note. SPM = standard progressive matrices. CPM = coloured progressive matrices. ASD = autism spectrum disorders. ^aRapporterer samme resultater for deltakerne ved 9 år som Bylund, 2000. ^bBasert på data fra samme studie som Gustafsson and Wolff (2015). Studiene varierer mht hvilken informasjon som er oppgitt i forhold til utvalget og bruk av Raven, og dermed kan innholdet i tabellen variere noe mellom studiene.

Normer og middelverdier

Det ble ikke identifisert noen normstudier på norske barn fra normalpopulasjonen. Det ble imidlertid funnet fem studier som oppga middelverdier og spredning for normale barn fra utvalgte skoler i eller rundt norske storbyer for SPM, samt en for CPM.

Internasjonale normer (hovedsaklig britiske; Raven, 2008a; Raven, 2008b) for forventet gjennomsnittlig skår for aldersspennet er supplert i tabellen (Tabell 3 og 4), for å gi et sammenligningsgrunnlag, med

tanke på å belyse om normene kan anvendes i Norge. I fem av seks studier av SMP avviker gjennomsnittet i den norske populasjonen fra den internasjonale normen, hvorav en eldre studie ligger under, og fire nyere ligger over. For CPM ligger den internasjonale normen innenfor utvalgets konfidensintervall ved begge observasjonstidspunktene. Selv om denne studien har et solid utvalg, er det imidlertid vanskelig å konkludere med at de internasjonale normene kan anvendes da studien er begrenset til første- og andreklassinger.

Tabell 3. Middelerverdier (*M*) og standardavvik (*SD*) for ulike normgrupper – SPM

Referanse	Utvalg/gruppe	Alder	N	Skala		Forventet skåre jf. britiske normer
				M	SD	
Elbro and Buch-Iversen (2013)	6. klassinger, pretest kontrollgruppe	10–11	85	38,9	10,2	> 32-33
	6. klassinger, pretest eksperimentgruppe	10–11	151	39,5	8,91	> 21-33
Green, et al. (2009)	10. klassinger, normale barn	15–16	233	45,4	8,6	> 35-36
Landmark and Grinde (1962)	Normale barn fra Bærum, 8-9 år (2. klasse 1962, 5 skoler, loddtrekning, representativ for bærum)	8–9	44	22,1	5,9	< 26-31
Lyster, et al. (2016)	Kontrollgruppe, 1. klasse	5–6	30	21,9	5,7	= 22-23 ^a
Solheim (2011)	5. klassinger, nasjonalt representativt, middelklasse	9–10	217	40,6	8,0	> 30-32

Note. Normene som brukes som sammenligningsgrunnlag er hentet fra en britiske standardiseringsstudie publisert i manualen til plussversjonen av SPM (Raven, 2008b), og er den forventede skåren for en person som ligger på medianen (50. prosentil) i sin aldersgruppe. (Raven, 2008b). > betyr at normen ligger under studiens konfidensintervall, = betyr innenfor og < betyr over. ^a Normer ikke tilgjengelig for aldersspenn, de supplerte er for 7-åring. Tallene er hentet fra artiklene, så formateringen kan variere noe. Originaltall som ikke hadde én desimal har for uniformitetens skyld enten blitt avrundet eller fått en ekstra null.

Tabell 4. Middelerverdier (*M*) og standardavvik (*SD*) for ulike normgrupper – CPM

Referanse	Utvalg/gruppe	Alder	N	Skala		Forventet skåre jf. britiske normer
				M	SD	
Lervåg, et al. (2009)	Normale barn, 1. klassinger	5–6	228	21,1	6,1	= 18-21
	Normale barn, 2. klassinger et år etter	6–7	228	25,7	5,4	= 22-27

Note. Normene som brukes som sammenligningsgrunnlag er hentet fra en britiske standardiseringsstudie publisert i manualen til plussversjonen av SPM (Raven, 2008b), og er den forventede skåren for en person som ligger på medianen (50. prosentil) i sin aldersgruppe. (Raven, 2008a). > betyr at normen ligger under studiens konfidensintervall, = betyr innenfor og < betyr over. Originaltall som ikke hadde én desimal har for uniformitetens skyld enten blitt avrundet eller fått en ekstra null.

Middelerverdier og/eller forekomster i kliniske undergrupper

For SPM ble det funnet én studie med rapporterte normer og kvantiler for populasjonen av døve barn (Frostad, 1995), samt to som brukte den som kovariat eller til å beskrive utvalget med et urelatert forskningsspørsmål. Disse oppga estimater for 18 forskjellige (relativt små) utvalg, delt på alder eller diagnose. Tabell 5 gir en oversikt over middelerverdier og spredning i de kliniske undergruppene for SPM.

(Frostad, 1995) presenterer en mer utfyllende analyse av SPM for relativt små utvalg av døve i ulike aldre, blant annet en tabell med fordeling av skårer på delkomponenter for ulike evnenivåer, hvor deres utvalg er sammenlignet med amerikanske og tyske normer for normale barn. I tillegg presenteres korrelasjoner mellom oppgave-rangering og aldersgruppe, tabell med kumulativ distribusjon (sammenlignet med amerikanske normdata), og parvise sammenligninger av forskjeller mellom aldersgruppene gjennomsnitt. Konklusjoner som

kan trekkes fra (Frostad, 1995) er at SPM fungerer relativt godt for døve, med noen avvik fra de internasjonale normene for normale barn.

Internasjonale normer (hovedsaklig britiske; Raven, 2008a; Raven, 2008b) for forventet gjennomsnittlig skår for det aktuelle aldersspennet er supplert i tabellene (5 og 6) for å gi et sammenligningsgrunnlag. De fleste studiene i denne kunnskapsoppsummeringen har relativt lav *n*, noe som gjør konfidensintervallene (95 %) store. I alt åtte av 19 studier kunne vise til gjennomsnitt som avviker fra den internasjonale normen, hvorav to ligger under og seks ligger over.

Det ble funnet en studie som brukte CPM. Studien til Söderqvist, et al. (2012) omhandler barn i et betydelig aldersspekter, med en gruppe som

mottok et kognitivt treningsprogram samt en kontrollgruppe. Utvalget er lite og har et bredt spekter av diagnoser, med svært få deltakere for hver diagnose, med det til felles at de skal være klassifisert med en intellektuell utviklingsforstyrrelse [intellectual disability], dvs. en intelligenskvotient (IQ) på under 70.

En sammenligning av så varierte og spesielle utvalg med en forventet medianskår må tas med en klype salt. Hvorvidt skårene burde ligge innenfor normalen avhenger også av om hva som forventes for den enkelte gruppen. Det faktum at det er lite entydighet i forholdet til medianen for sammenlignbare utvalg taler ikke til fordel for de britiske normenes brukbarhet. Studiene har imidlertid forholdsvis lav statistisk styrke, og betydningen av funnene bør derfor ikke overdrives.

Tabell 5. Middelerverdier (*M*) og standardavvik (*SD*) for ulike undergrupper - SPM

Referanse	Utvalg/gruppe	Alder	<i>n</i>	Skala		Forventet skåre jf. britiske normer
				<i>M</i>	<i>SD</i>	
Fasting and Lyster (2005)	5. klassinger, "struggling readers and spellers", eksperimentgruppe pretest	9–10	4	35,4	16,9	= 30-32
	6. klassinger, "struggling readers and spellers", eksperimentgruppe pretest	10–11	16	23,1	14,2	< 32-33
	7. klassinger, "struggling readers and spellers", eksperimentgruppe pretest	11–12	6	25,2	11,6	= 32-33
	5. klassinger, "struggling readers and spellers", kontrollgruppe pretest	9–10	4	29,6	16,9	= 30-32
	6. klassinger, "struggling readers and spellers", kontrollgruppe pretest	10–11	15	24,8	14,3	= 32-33
	7. klassinger, "struggling readers and spellers", kontrollgruppe pretest	11–12	7	18	6,6	< 32-33
Frostad (1995)	Hørselshemmede elever	7	11	25,8	9,9	= 22-23
	Hørselshemmede elever	8	24	27,9	6,8	= 26-27
	Hørselshemmede elever	9	33	27,2	10,8	= 30-31
	Hørselshemmede elever	10	22	33,3	12,9	= 32
	Hørselshemmede elever	11	24	39,4	9,9	> 32-33
	Hørselshemmede elever	12	31	37,3	8,6	> 33

	Hørselshemmede elever	13	43	39,4	7,9	> 34-35
	Hørselshemmede elever	14	30	41,6	9,6	> 34-35
	Hørselshemmede elever	15	50	43,7	9,8	> 35-36
Green, et al. (2009)	10. klassinger med lese og skrivevansker (selvrapportert) type mild	15–16	43	45,0	6,7	> 35-36
	10. klassinger med lese- og skrivevansker (selvrapportert) type moderat	15–16	28	40,1	11,5	= 35-36
	10. klassinger med lese- og skrivevansker (selvrapportert) type alvorlig (severe)	15–16	4	42,5	8,5	= 35-36

Note. Normene som brukes som sammenligningsgrunnlag er hentet fra en britiske standardiseringsstudie publisert i manualen til plussversjonen av SPM (Raven, 2008b), og er den forventede skåren for en person som ligger på medianen (50. prosentil) i sin aldersgruppe. > betyr at normen ligger under studiens konfidensintervall, = betyr innenfor og < betyr over. Originaltall som ikke hadde én desimal har for uniformitetens skyld enten blitt avrundet eller fått en ekstra null.

Tabell 6. Middelerverdier (*M*) og standardavvik (*SD*) for ulike undergrupper - CPM

Referanse	Utvalg/gruppe	Alder	<i>n</i>	Skala		Forventet skåre jf. britiske normer
				<i>M</i>	<i>SD</i>	
Söderqvist, et al. (2012)	Intellectual disability (IQ under 70), blandede diagnoser: ADHD, Downs, epilepsi eller andre grupper. Pretest gruppe 1	6–12	22	9,0	3,9	< 20-33
	Intellectual disability (IQ under 70) blandede diagnoser: ADHD, Downs, epilepsi eller andre grupper. Pretest gruppe 2	6–12	19	8,0	4,2	< 20-33

Note. Normene som brukes som sammenligningsgrunnlag er hentet fra en britiske standardiseringsstudie publisert i manualen til plussversjonen av SPM (Raven, 2008b), og er den forventede skåren for en person som ligger på medianen (50. prosentil) i sin aldersgruppe. > betyr at normen ligger under studiens konfidensintervall, = betyr innenfor og < betyr over. Tallene er hentet fra artiklene, så antall desimaler kan variere noe. Originaltall som ikke hadde én desimal har for uniformitetens skyld enten blitt avrundet eller fått en ekstra null.

Reliabilitet

I det norske søket fantes det to artikler som rapporterte reliabilitetsmål, begge på testens interne konsistens. Landmark and Grinde (1962) rapporterte for SPM en split-half-reliabilitet på 0,89. Lervåg, et al. (2009) rapporterte for CPM alpha-koeffisienter (Cronbach's alpha; Cronbach, 1951) på 0,80 samt 0,85 når barna ble testet ett år senere. I de aller fleste artiklene basert på svenske/danske utvalg var ikke reliabilitet undersøkt for det aktuelle utvalget, men man hadde henvist til reliabilitet oppgitt i manualen. Tre av studiene hadde rapportert Cronbachs alpha for testen, med 0,80 (Olsson, et al., 2016) og 0,87 (Skagerlund & Traff,

2016a) for SPM, og 0,72 for CPM (Gustafson, et al., 2013). Disse verdiene betraktes som adekvate til gode i forhold til EFPA-kriteriene (EFPA, 2013).

Alpha-koeffisienter er kun ett begrenset mål på den nedre grensen for dette aspektet ved reliabilitet (mål som for eksempel test-retest og parallell-form adresserer andre viktige aspekter ved reliabilitet, men disse ble ikke rapportert i noen av studiene), og er bl.a. avhengig av en sentral antagelse om at oppgavene er likeverdige (Schmitt, 1996; Sijtsma, 2009). I en test som Raven, hvor oppgavene varierer i vanskelighetsgrad eller hvor det finnes tidspress, vil reliabiliteten for eksempel kunne variere sterkt avhengig av hvor på skalaen en person eller

oppgave befinner seg samt hvor mange som befinner seg på et gitt punkt på skalaen og hvor godt de differensierer, og dette fanges ikke opp i enkeltstatistikker som alpha-koeffisienten. Her er latent variabel-modellering nødvendig (item response theory - se f.eks. Lord, Novic, & Birnbaum, 2008).

Validitet

Validitet - Norske studier

Fra det norske søket var det fire studier som belyste validiteten til SPM, samt to som belyste CPM, og da i hovedsak begrepsvaliditet.

En vanlig måte å undersøke begrepsvaliditeten til en test er å identifisere dens plass i det nomologiske nettverket (Cronbach, 1955), det vil si hvordan den relaterer seg til andre konstrukter som den teoretisk sett skal konvergere med eller divergere fra. Det ble ikke identifisert noen dedikerte valideringsstudier av Ravens matriser på den norske populasjonen, men noen artikler rapporterte relasjoner til andre konstrukter. Raven skal være en av de mest sentrale målene på flytende intelligens (*Gf* – eng: fluid intelligence) (Marshalek, et al., 1983). Dette betyr at mål på *Gf* burde være nærmere i testens nomologiske nettverk enn for eksempel mål på krystallisert intelligens (*Gc* – eng: crystallized intelligence) (McGrew, 2005), og relativt fjernt fra konstrukter som ikke har mye med intelligens å gjøre. Siden vi vet at *Gf* lader høyt på den overordnede *g*-faktoren (Kvist & Gustafsson, 2008), burde vi også anta at Raven i det minste vil korrelere noe med de fleste evnetester uavhengig av domene.

Her ser resultatene for SPM (Tabell 7) ut til å korrespondere godt med disse antakelsene for den norske populasjonen, ved at den er tettere relatert til visse intelligenstester enn til språkkonstrukter, mens holdningskonstrukter er relativt urelaterte til Raven. Ut ifra klassiske tommelfingerregler (se f.eks. Cohen, Cohen, West, & Aiken, 2003) kan vi si at SPM stort sett korrelerer middels ($r = 0,30$) til relativt høyt ($r = 0,50$) med andre intelligenstester

for barn fra normalutvalg, og korrelerer tilsynelatende også høyt med tilsvarende matrisebaserte tester på et blandet klinisk utvalg. Det vises til lave ($r = 0,10$) til middels høye korrelasjoner med språklige konstrukter, samt en screeningtest for dysleksi. Holdningsbaserte konstrukter som reading self-efficacy og task value ser ut til å korrelere svært svakt med testen. Bosnes (2005) rapporterte også en korrelasjon på 0,81 mellom SPM og subtesten *matriser* fra WAIS III, men det er uklart om dette resultatet er fra den aktuelle studien. For ordens skyld, tilsvarende korrelasjonene som er observert her også noenlunde det som rapporteres for andre lands populasjoner (mer om internasjonale valideringsstudier kan bli funnet i kunnskapsoppsummeringen til John Rust og kolleger fra manualen til SPM; Raven, 2008b).

De norske studiene i kunnskapsoppsummeringen belyser det nomologiske nettverket til CPM (Tabell 8) viser en lav til middels relasjon mellom CPM og diverse språkkonstrukter (Lervåg, et al., 2009) samt en noe sterkere relasjon mellom CPM og WISC-R (Høie, et al., 2008), og er her på linje med SPM. Studien til Lervåg og kollegaer viser også en forventet utvikling i CPM-skår ettersom barna blir eldre. Når det gjelder forskjeller mellom grupper, viser Høie, et al. (2005) med et stort utvalg av barn med epilepsi ($N = 309$; alder = 8-12 år), at disse er sterkt overrepresentert i de lavere prosentiler av CPM, mens kontrollgruppen er sterkt overrepresentert i de øvre prosentiler. I en studie på den samme populasjonen fant Høie, Sommerfelt, et al. (2006) en negativ korrelasjon mellom CPM og lærer/foreldrerapporterte psykososiale vansker ($r = -0,37$), stort sett urelatert til om barnet hadde epilepsi eller ikke. De fant også i en annen studie (Høie, Mykletun, et al., 2006) at omtrent 20 % av variasjonen i et mål på eksekutive funksjoner kunne forklares av skår på CPM, etter å ha kontrollert for depresjon og sosioøkonomisk status. De sistnevnte studiene belyser det nomologiske nettverket til intelligenskonstruktet ved å vise at testen måler aspekter ved sentrale kognitive funksjoner, samt at lav intelligens-skår kan være en psykososial risikofaktor.

Tabell 7. Korrelasjonskoeffisienter mellom SPM og andre mål (norske studier)

Forfatter	Utvalg	Alder(år)	N	Skala	r
Bosnes (2005)	Blandet klinisk utvalg: spesifikke lærevansker, lettere psykiske vansker, psykisk utviklingshemming, traumatisk hjerneskade, ADHD, alvorlig psykisk lidelse, autisme, blødning i sentralnervesystemet og demens.	9–68	41	WASI matriser	0,69
Green, et al. (2009)	10-klassinger (15–16 år), normale barn	15–16	233	Duvan Dyslexia Screening test	0,32
Landmark and Grinde (1962)	Normale barn, 8–9 år (2. klasse 1962), 5 skoler hvor elever er valgt etter loddtrekning, representativ for Bærum, SPM sets A, AB, B, revised order (Raven, 1956)	8–9	44	Stanford-Binet, norsk standard	0,75
				Leiter international performance scale, Arthur adaption	0,69
				Columbia Mental Maturity scale	0,42
				Goodenough "draw-a-man"	0,53
Solheim (2011)	5. klassinger, nasjonalt representativt, middelklasse	9–10	217	Word reading ability	0,25
				Listening comprehension	0,27
				Reading self-efficacy	0,10 n.s.
				Reading task value	-0,15
				Multiple choice (MC) reading comprehension	0,48
				Constructed response (CR) reading comprehension	0,39

Note. Samtlige r (unntatt en merket n.s.) var signifikante. r = bivariat korrelasjon.

Tabell 8. Korrelasjonskoeffisienter mellom CPM og andre mål (norske studier)

Forfatter	Utvalg	Alder(år)	N	Skala	r
(Høie, et al., 2008))	Barn med epilepsi	6–12	269	WISC-R	0,63
Lervåg, et al. (2009)	Normale barn, 1. klassinger	5–6	228	Phoneme awareness	0,43
For korrelasjoner med substester, se Lervåg and Hulme (2010)				Paired-associated learning (PAL)	0,36
				Non-alphanumeric rapid automatized naming (RAN)	0,29
				Verbal short-term memory	0,39
				Letter knowledge	0,44
				Verbal abilities	0,47

Normale barn, 2. klassinger, observasjon ett år etter	6-7	228	Phoneme awareness	0,47
			Paired-associated learning (PAL)	0,34
			Non-alphanumeric rapid automatized naming (RAN)	0,26
			Verbal short-term memory	0,26
			Letter knowledge	0,31
			Verbal abilities	0,23

Note. Samtlige *r* var signifikante. *r* = bivariat korrelasjon.

Validitet – Svenske og Danske studier

Fra det svenske/danske søket var det til sammen 24 studier (Tabell 2) som i hovedsak belyste begrepsvaliditeten til testen på litt ulike måter. Disse kan grupperes avhengig av hvordan de kan sies å belyse begrepsvaliditeten enten i form av korrelasjoner med andre tester, forskjeller mellom grupper, intervensjoner for å øke intelligens eller faktoranalysestudier.

Korrelasjoner med andre tester.

Litt flere korrelasjonsstudier var gjennomført med SPM sammenlignet med CPM. Hovedfunn fra disse studiene i form av korrelasjoner er presentert i Tabell 9 og 10. De fleste studiene for Raven var

gjennomført på skolebarn og viste generelt noe høyere korrelasjoner mellom Raven og tester som målte problemløsning, spatiale og matematiske evner sammenlignet med ulike verbale og språktester (Tabell 9). Dette er i tråd med antakelser om at Raven er et mål på non-verbal intelligens. I en prospektiv svensk studie ble det undersøkt om alder, SPM og arbeidshukommelse predikerte matematiske ferdigheter to år senere hos en gruppe tilfeldig valgte barn i alderen 6-16 år (*N* = 246) (Dumontheil & Klingberg, 2012). Alle prediktorene var signifikante ($R^2 = .58$), inklusive Raven. De fleste studiene for CPM sammenlignet testen med mål på dekodning, hukommelse og ulike språktester der korrelasjonene varierte fra små og ikke signifikante til middels store (Tabell 10).

Tabell 9. Korrelasjonskoeffisienter mellom SPM og andre mål (svenske/danske studier)

Forfatter	Utvalg	Alder(år)	<i>N</i>	Skala/test	<i>r</i>
(Andersson, 2007)	Skolebarn, Sverige	10	69	Mathematical word problem solving	0,43
				Arithmetical calculation	0,44
				Reading ability	0,34
				Visual matrix task	0,23
				Animal dual-task	0,28
				Verbal fluency	0,27
				Trail-making	0,41
				Digit span	0,22
(Andersson, 2008)	Skolebarn, Sverige	10,3	121	Arithmetical composite measure	0,58
				Reading task	0,35
				Counting span task	0,36

				Visual matrix span	0,41	
				Verbal fluency	0,27	
				Trail-making	0,61	
				Stroop task	0,18	
				Digit span task	0,23	
				Corsi-block span	0,29	
(Fischbein, et al., 1999)	Skolebarn, Sverige	9–15	73	Verbal test	0,42	
				Swedish	0,41	
				Arithmetic	0,52	
(Forssman, et al., 2012)	Ungdommer, Sverige	15	120	ADHD/ODD symptoms behavior	0,09–0,22	
				Working memory	0,37	
				Interference control	0,20	
				Response inhibition	0,00	n.s.
				Reaction time	0,29	
(Skagerlund & Traff, 2016b)	Skolebarn, Sverige	9,7	133	Math score	0,43	
				Processing speed	0,22	
				Trail making	0,25	
				Visuospatial working memory	0,15	
				Verbal working memory	0,16	
				Symbolic number comparison	0,05	n.s.
				Non-symbolic number discrim.	0,17	
				Mental rotation	0,39	
				Paper folding	0,55	
				Time discrimination	0,37	
(Träff, 2013)	Skolebarn, Sverige	10–13	134	Reading task	0,35	
				Arithmetic fact retrieval	0,29	
				Arithmetic calculation	0,43	
				Word problem solving	0,51	
				Subitizing and dot counting	0,29	
				Number line estimation errors	0,37	
				Symbolic number comparisons	0,14	
				Visual matrix span	0,30	
				Listening span	0,27	

Verbal fluency	0,35
Color naming	0,23

Note: Alle korrelasjonene var i forventet retning og er derfor presentert uten fortegn i tabellen. Samtlige r (unntatt to merket n.s.) var signifikante. r = bivariat korrelasjon.

Tabell 10. Korrelasjonskoeffisienter mellom CPM og andre mål (svenske/danske studier)

Forfatter	Utvalg	Alder(år)	N	Skala	r
(Gellert & Elbro, 2013)	Skolebarn, Danmark	9,6	68–90	Vocabulary training	0,25
				Definition knowledge	0,05 n.s.
				Immediate recall	0,19 n.s.
				Immediate recognition	0,29
				Delayed recall	0,08 n.s.
				Delayed recognition	0,08 n.s.
				Receptive vocabulary	0,30
				Expressive vocabulary	0,38
(Dahlin, 2011)	Barn med lærevansker	9–12	57	Word decoding	0,39
				Orthographic verification	0,37
				Reading comprehension	0,52
				Working memory training scores	0,41
(Gustafson, et al., 2013)	Skolebarn med og uten lesevansker	4 klasse	36x2	Reading ability	0,36
				Decoding	0,06 n.s.
				Phonological decoding	0,16 n.s.
				Orthographic decoding	0,10 n.s.
				Language comprehension	0,39
				Word comprehension	0,48
				Receptive grammar	0,43
(Nordberg, et al., 2014)	Barn med CP og språkvansker	11	15	Narrative assessment profile	0,06 n.s.
				Expressive language	0,49 n.s.
				Receptive language	0,30 n.s.
				Visual memory	0,54
				Auditory memory	0,06 n.s.
(Nordberg, et al., 2014)	Barn med CP og språkvansker	11,2	19	Articulation score	0,56

Note: * Studien presenterte r for begge grupper av barn. Korrelasjoner for vanlige skolebarn er presentert i tabellen. Alle korrelasjonene var i forventet retning og er derfor presentert uten fortegn i tabellen. Samtlige r (unntatt de som er merket n.s.) var signifikante. r = bivariat korrelasjon.

Forskjeller mellom grupper

En studie av Nutley et al. (2014) viste forskjeller i SPM mellom grupper av barn der de barna som spilte et instrument, skåret omtrent ett standardavvik over gruppen som ikke spilte musikk ved første måletidspunkt. Forskjellen mellom gruppene var fremdeles signifikant etter at man hadde kontrollert for en del demografiske og familiefaktorer som foreldres utdanningsnivå.

I en liten klinisk studie av barn med autisme ble gruppen sammenlignet med en gruppe med vanlige barn på ulike språktester samt CPM. Det var ikke signifikante forskjeller i Raven mellom disse gruppene, selv om tendensen gikk i forventet retning. Det vil si at barna diagnostisert med autisme skåret lavere enn sammenligningsgruppen. Studien hadde imidlertid lav statistisk styrke til å avdekke forskjeller siden det kun var 21 barn i hver gruppe (Brynskov, et al., 2017).

I en studie av premature barn med lav fødselsvekt (< 1500gr) ble ulike kognitive funksjoner undersøkt ved 9-årsalderen og resultatene sammenlignet med en kontrollgruppe med fulltermin barn med normalvekt (Bylund, et al., 2000). Resultatene viste signifikante forskjeller i skårer på SPM der gruppen av barn med lav fødselsvekt skåret i 3/4 standardavvik lavere enn kontrollgruppen (Hedges $g = 0,72$).

I en longitudinell studie av barn med lesevansker ($N = 70$) (Svensson & Jacobson, 2006) ble gruppen sammenlignet med barn uten lesevansker fra samme klassetrinn på en rekke kognitive tester først i 2. klasse og deretter 10 år senere. Det var ingen signifikant forskjell mellom gruppene på T1 på CPM, mens på T2 skåret gruppen med lesevansker signifikant lavere på CPM enn sammenligningsgruppen. Forskjellen mellom gruppene på T2 var over et standardavvik – noe som må betegnes som en stor forskjell (Svensson & Jacobson, 2006).

Studier for å øke intelligens

En intervensjonsstudie (Thirus, et al., 2016) ble inkludert i oppsummeringen der formålet var å undersøke om det var mulig å øke matematiske evner og generell intelligens målt med SPM blant elever i videregående skole. Studien var

eksperimentell med 10 elever i intervensjonsgruppen og 11 i kontrollgruppen. Intervensjonsgruppen mottok et nettbasert program SMART (Strengthening Mental Abilities with Relational Training) som gikk over 8-12 uker. Resultatene indikerte at programmet hadde hatt en liten effekt på prestasjoner på SPM etter at man hadde kontrollert for alder og pre-test skårer på Raven. Studien er imidlertid svært liten og med betydelig frafall underveis.

Faktoranalysestudier

I flere studier har Gustafsson et al. undersøkt strukturen til intellektuelle evner, og ulike grupper er testet med en rekke kognitive og kunnskapstester, deriblant SPM. Den tidligste studien (Gustafsson, 1984) ble gjennomført på en stor gruppe skolebarn ($N = 1224$) i tolv-års alderen. En hierarkisk modell med tre nivå av faktorer, inklusive en g -faktor på det tredje nivået. Modellen fikk støtte og Raven ladet på en spatial faktor (Figural Relations) som igjen ladet på faktoren flytende intelligens (Gf) – denne ladet i sin tur på g -faktoren. I en nyere lignende studie av fireåringer ble også konfirmatorisk faktoranalyse benyttet for å undersøke hvilken struktur kognitive evner har (Gustafsson & Wolff, 2015). Studien viste at CPM ladet på faktoren for flytende visuospatuell intelligens (Gf) sammen med to andre Wechsler tester (Block Design og Wechsler Non-verbal Scale for Intelligence) som er antatt å måle den samme faktoren. I en annen artikkel basert på de samme barna (Wolff & Gustafsson, 2015) der formålet var å undersøke faktorstrukturen til fonologiske evner, ble også CPM inkludert i faktoranalysen. I disse analysene ble en tre-faktor modell funnet der CPM ladet på faktoren flytende visuell intelligens.

Øvrig

En gammel studie (Vejleskov, 1968) undersøkte vanskelighetsgraden for de ulike oppgavene i de ulike settene for SPM, og i tillegg kjønns- og demografiske forskjeller. Studien gav støtte til antakelsen om at settene har økende vanskelighetsgrad og tilsvarende at de tolv oppgavene innen hvert sett også øker i vanskelighetsgrad. Gjennomsnittstallene fra de to

danske skolene var svært like gjennomsnittene i manualen fra 1958 (Vejleskov, 1968).

Diskusjon og konklusjon

Reliabilitetsmessig ser de få resultatene som finnes ut til å sammenfalle noenlunde med de internasjonale forventningene. At det mangler statistikk på test retest-reliabilitet og parallelle former for norske eller skandinaviske populasjoner er en betydelig svakhet, og dette betyr at brukere av testen essensielt sett må ta en beslutning om hvorvidt de internasjonale reliabilitetsresultater kan antas å være gyldige også for den norske populasjonen.

Validitet har heller ikke vært overveldende undersøkt i Norge, men de studiene som er gjennomført ser ut til å støtte oppunder Ravens plass i det nomologiske nettverket. I Sverige har man funnet at testen også ligger i kjernen av konstruktet flytende intelligens både for barn (Gustafsson, 1984) og voksne (Kvist & Gustafsson, 2008), resultater som også har blitt påvist i Norge for lignende abstrakte resonneringstester (som Cattell's culture free test og WISC matriser; Undheim, 1981; Undheim & Gustafsson, 1987). Validiteten til Raven i Norge og Norden later følgelig til å sammenfalle med det vi vet fra internasjonal forskning på testen (Raven, 2000a), hvor den ansees som et godt mål på abstrakt resonneringsevne.

Selv om validiteten er god ser ikke testens normgrunnlag ut til å ha blitt målrettet undersøkt hverken i Norge, Sverige eller Danmark, med unntak av studien på hørselshemmede (Frostdad, 1995). I publikasjonene som ble gjennomgått i kunnskapsoppsummeringen er det en del usikkerhet knyttet til forholdene testene er administrert under. Det er lite standardisert eller rapportert når det for eksempel gjelder hvilke testsett som ble brukt, og få om noen har oppgitt hvilket tidspress barna var satt under når de tok testen. Dette er en betydelig systematisk feilkilde som sår tvil om sammenligningsgrunnlaget til middelverdiene. Fra de rapporterte middelverdiene for SPM, ser det ut til at en god del av studiene ikke

samsvarer med normene fra manualen (Raven, 2008b). Hvorvidt dette skyldes at norske barn reagerer annerledes på oppgavens stimuli enn barn fra andre kulturer, om det er en del av den internasjonale trenden med populasjonsomfattende økninger i skår på lignende tester (se f.eks. Flynn, 2011), eller om det kun er tilfeldige variasjoner, kan vi kun spekulere i. En islandsk studie av et relativt stort utvalg av tilfeldig valgte barn ($N = 550$) viste for eksempel at disse skåret to-tre poeng bedre på SPM enn normene i manualen noe forfatterne tilskrev Flynn-effekten ved at normene var omtrent 20 år gamle på det tidspunkt studien ble gjennomført (Pind, Gunnarsdóttir, & Jóhannesson, 2003). Mer forskning på den norske populasjonen er helt klart nødvendig, og en norsk standardisering av testen kunne adressert en del av usikkerhetsmomentene. En kunnskapsoppsummering av studier basert på Raven på tvers av kulturer og tidsrom viste betydelige økning i totalskår over tid, noe som understreker betydningen av å anvende oppdaterte normer (Raven, 2000b). Dette ble også funnet i en meta-analyse av studier med Raven basert på hele 48 land og over en periode på 64 år (Wongupparaj, Kumari, & Morris, 2015). Disse studiene understreker betydningen av å anvende oppdaterte normer.

Både SPM og CPM er generelt fri for språklige stimulus, og skal ideelt sett være relativt kulturfrie, men før det er foretatt noen ordentlig normstudie av testene i Norge, kan vi kun støtte oss på teoretiske antagelser om at den fungerer som den skal. Dette burde være forsvarlig i forskning hvor man er interessert i et rent mål på intelligens, med referanse til fordelingen i eget utvalg (noe som er tilfelle for majoriteten av studiene i denne kunnskapsoppsummeringen, men ikke alle). Det blir imidlertid problematisk å for eksempel bruke den i kliniske settinger hvor klassifikasjonsbeslutninger er basert på fordelingen i den norske normalpopulasjonen. Raven burde være et godt mål på intelligens, også i Norge, selv om de britiske normene kan ha begrenset overføringsverdi. Det vil derfor være svært nødvendig å få etablert norske normer for testen med tanke på klinisk bruk.

Referanser

- Andersson, U. (2007). The contribution of working memory to children's mathematical word problem solving. *Applied Cognitive Psychology, 21*(9), 1201-1216.
- Andersson, U. (2008). Working memory as a predictor of written arithmetical skills in children: The importance of central executive functions. *British Journal of Educational Psychology, 78*(2), 181-203.
- Bosnes, O. (2005). Comparison of Wechsler Adult Intelligence Scale/Wechsler Intelligence Scale for Children-Revised, and Wechsler Abbreviated Scale of Intelligence in a Norwegian clinical sample. *Tidsskrift for Norsk Psykologforening, 42*(7), 598-602.
- Brynskov, C., Eigsti, I.-M., Jorgensen, M., Lemcke, S., Bohn, O.-S. & Krojgaard, P. (2017). Syntax and morphology in Danish-speaking children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 47*(2), 373-383. doi: <http://dx.doi.org/10.1007/s10803-016-2962-7>
- Bylund, B., Cervin, T., Finnström, O., Gäddlin, P.-O., Leijon, I., Mård, S. et al. (2000). Very low birth weight children at 9 years : School performance and behaviour in relation to risk factors. *Prenatal and Neonatal Medicine, 5*(2), 124-133.
- Carpenter, P. A., Just, M. A. & Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological review, 97*(3), 404-431.
- Cattell, R. (1940). A culture-free intelligence test I. *Journal of Educational Psychology, 31*(3), 161-179.
- Cohen, J., Cohen, P., West, S. G. & Aiken, L. (2003). *Applied Multiple Regression / Correlation Analysis for the Behavioral Sciences* (3rd utg.). London: Lawrence Erlbaum associates.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 74*(1), 145-154.
- Cronbach, L. J. (1955). Construct validity in psychological tests. *Psychological bulletin, 52*(4), 281-302.
- Dahlin, K. I. (2011). Effects of working memory training on reading in children with special needs. *Reading and Writing, 24*(4), 479-491. doi: 10.1007/s11145-010-9238-y
- Dumontheil, I. & Klingberg, T. (2012). Brain activity during a visuospatial working memory task predicts arithmetical performance 2 years later. *Cerebral Cortex, 22*(5), 1078-1085.
- Elbro, C. & Buch-Iversen, I. (2013). Activation of background knowledge for inference making: Effects on reading comprehension. *Scientific Studies of Reading, 17*(6), 435-452.
- European Federation of Psychologists' Association - EFPA. (2013). EFPA Review model for the description and evaluation of psychological tests: Test review form and notes for reviewers, v 4.2.6: European Federation of Psychologists' Association (EFPA).
- Fasting, R. B. & Lyster, S.-A. H. (2005). The effects of computer technology in assisting the development of literacy in young struggling readers and spellers. *European Journal of Special Needs Education, 20*(1), 21-40.
- Fischbein, S., Guttman, R. & Nathan, M. (1999). Genetic and environmental influences on pupil performances. *Twin Research, 2*(3), 183-195.
- Flynn, J. R. (2011). Secular changes in intelligence. I R. J. Sternberg & S. B. Kaufman (Red.), *The Cambridge handbook of intelligence* (s. 647-663). New York: Cambridge University press.
- Forssman, L., Eninger, L., Tillman, C. M., Rodriguez, A. & Bohlin, G. (2012). Cognitive functioning and family risk factors in relation to symptom behaviors of ADHD and ODD in adolescents. *Journal of Attention Disorders, 16*(4), 284-294. doi: 10.1177/1087054710385065
- Frostad, P. (1995). *Ravens progressive matriser brukt på et utvalg hørselshemmede elever : notat*. Trondheim: P. Frostad.
- Gellert, A. S. & Elbro, C. (2013). Do experimental measures of word learning predict vocabulary development over time? A study of children from grade 3 to 4. *Learning and Individual Differences, 1*-8.
- Green, K., Tønnessen, F., Tambs, K., Thoresen, M. & Bjertness, E. (2009). Dyslexia: Group screening among 15-16-year-olds in Oslo, Norway. *Scandinavian Journal of Educational Research, 53*(3), 217-227.
- Gustafson, S., Samuelsson, C., Johansson, E. & Wallmann, J. (2013). How simple is the simple view of reading? *Scandinavian Journal of Educational Research, 57*(3), 292-308.
- Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*(3), 179-203. doi: [http://dx.doi.org/10.1016/0160-2896\(84\)90008-4](http://dx.doi.org/10.1016/0160-2896(84)90008-4)

- Gustafsson, J.-E. & Wolff, U. (2015). Measuring fluid intelligence at age four. *Intelligence*, 50, 175-185. doi: <http://dx.doi.org/10.1016/j.intell.2015.04.008>
- Høie, B., Mykletun, A., Sommerfelt, K., Bjornæs, H., Skeidsvoll, H. & Waaler, P. E. (2005). Seizure-related factors and non-verbal intelligence in children with epilepsy a population-based study from Western Norway. *Seizure*, 14(4), 223-231.
- Høie, B., Mykletun, A., Waaler, P., Skeidsvoll, H. & Sommerfelt, K. (2006). Executive functions and seizure-related factors in children with epilepsy in western Norway. *Developmental Medicine & Child Neurology*, 48(6), 519-525.
- Høie, B., Sommerfelt, K., Waaler, P., Alsaker, F., Skeidsvoll, H. & Mykletun, A. (2006). Psychosocial problems and seizure-related factors in children with epilepsy. *Developmental Medicine & Child Neurology*, 48(3), 213-219.
- Høie, B., Sommerfelt, K., Waaler, P. E., Alsaker, F. D., Skeidsvoll, H. & Mykletun, A. (2008). The combined burden of cognitive, executive function, and psychosocial problems in children with epilepsy: A population-based study. *Developmental Medicine and Child Neurology*, 50(7), 530-536.
- Kvist, A. V. & Gustafsson, J.-E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's Investment theory. *Intelligence*, 36(5), 422-436. doi: [10.1016/j.intell.2007.08.004](http://dx.doi.org/10.1016/j.intell.2007.08.004)
- Landmark, M. & Grinde, T. (1962). A group of normal children tested with five intelligence tests. *Nordisk Psykologi*, 14(4), 171-185.
- Lervåg, A., Bråten, I. & Hulme, C. (2009). The cognitive and linguistic foundations of early reading development: A Norwegian latent variable longitudinal study. *Developmental Psychology*, 45(3), 764-781.
- Lervåg, A. & Hulme, C. (2010). Predicting the growth of early spelling skills: Are there heterogeneous developmental trajectories? *Scientific Studies of Reading*, 14(6), 485-513.
- Lord, F. M., Novic, M. R. & Birnbaum, A. (2008). *Statistical theories of mental test scores*. USA: Information age publishing inc.
- Lyster, S.-A. H., Lervåg, A. O. & Hulme, C. (2016). Preschool morphological training produces long-term improvements in reading comprehension. *Reading and Writing*, 29(6), 1269-1288. doi: <http://dx.doi.org/10.1007/s11145-016-9636-x>
- Marshalek, B., Lohman, D. & Snow, R. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, 7(2), 107-127.
- McGrew, K. S. (2005). The Cattell-Horn-Carroll Theory of Cognitive Abilities: Past, Present, and Future. I P. Flanagan & P. L. Harrison (Red.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues*. New York: Guilford press.
- Nordberg, A., Dahlgren Sandberg, A. & Miniscalco, C. (2015). Story retelling and language ability in school-aged children with cerebral palsy and speech impairment. *International Journal of Language & Communication Disorders / Royal College of Speech & Language Therapists*, 50(6), 801-813. doi: <http://dx.doi.org/10.1111/1460-6984.12177>
- Nordberg, A., Miniscalco, C. & Lohmander, A. (2014). Consonant production and overall speech characteristics in school-aged children with cerebral palsy and speech impairment. *International Journal of Speech-Language Pathology*, 16(4), 386-395.
- Nutley, S. B., Darki, F. & Klingberg, T. (2014). Music practice is associated with development of working memory during childhood and adolescence. *Frontiers in Human Neuroscience*, 7, 926. doi: <https://dx.doi.org/10.3389/fnhum.2013.00926>
- Oakland, T., Douglas, S. & Kane, H. (2016). The top ten standardized tests used with children and youth by school psychologists in 64 countries: A 24 year follow-up study. *Journal of Psychoeducational Assessment*, 34, 166-176.
- Olsson, L., Ostergren, R. & Traff, U. (2016). Developmental dyscalculia: A deficit in the approximate number system or an access deficit? *Cognitive Development*, 39, 154-167. doi: <http://dx.doi.org/10.1016/j.cogdev.2016.04.006>
- Pind, J., Gunnarsdóttir, E. K. & Jóhannesson, H. S. (2003). Raven's Standard Progressive Matrices: new school age norms and a study of the test's validity. *Personality and Individual Differences*, 34(3), 375-386. doi: [10.1016/S0191-8869\(02\)00058-2](http://dx.doi.org/10.1016/S0191-8869(02)00058-2)
- Primi, R. (2001). Complexity of geometric inductive reasoning tasks contribution to the

- understanding of fluid intelligence. *Intelligence*, 30(1), 41-70.
- Raven, J. (1941). Standardization of progressive matrices, 1938. *British Journal of Medical Psychology*, 19, 137-150.
- Raven, J. (2000a). Psychometrics, cognitive ability, and occupational performance. *Review of Psychology*, 7(1-2), 51-74.
- Raven, J. (2000b). The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology*, 41(1), 1-48.
- Raven, J. (2008a). *Manual for the coloured progressive matrices and Crichton vocabulary scale* (UK utg.). London: Pearson education.
- Raven, J. (2008b). *Manual for the standard progressive matrices - plus version and Mill Hill vocabulary scale* (UK utg.). London: Pearson education.
- Samuelsson, S., Finnstrom, O., Flodmark, O., Gaddlin, P.-O., Leijon, I. & Wadsby, M. (2006). A Longitudinal Study of Reading Skills Among Very-Low-Birthweight Children: Is There a Catch-up? *Journal of Pediatric Psychology*, 31(9), 967-977.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological assessment*, 8(4), 350-353.
- Sijtsma, K. (2009). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika*, 74(1), 107-120. doi: 10.1007/s11336-008-9101-0
- Skagerlund, K. & Traff, U. (2016a). Number processing and heterogeneity of developmental dyscalculia: Subtypes with different cognitive profiles and deficits. *Journal of Learning Disabilities*, 49(1), 36-50. doi: <http://dx.doi.org/10.1177/002221941452707>
- Skagerlund, K. & Traff, U. (2016b). Processing of space, time, and number contributes to mathematical abilities above and beyond domain-general cognitive abilities. *Journal of Experimental Child Psychology*, 143, 85-101. doi: <http://dx.doi.org/10.1016/j.jecp.2015.10.016>
- Solheim, O. J. (2011). The impact of reading self-efficacy and task value on reading comprehension scores in different item formats. *Reading Psychology*, 32(1), 1-27.
- Svensson, I. & Jacobson, C. (2006). How persistent are phonological difficulties? A longitudinal study of reading retarded children. *Dyslexia: An International Journal of Research and Practice*, 12(1), 3-20.
- Söderqvist, S., Nutley, S. B., Ottersen, J., Grill, K. M. & Klingberg, T. (2012). Computerized training of non-verbal reasoning and working memory in children with intellectual disability. *Frontiers in Human Neuroscience Vol 6 Oct 2012, ArtID 271*.
- Thirus, J., Starbrink, M. & Jansson, B. (2016). Relational frame theory, mathematical, and logical skills: A multiple exemplar training intervention to enhance intellectual performance. *International Journal of Psychology & Psychological Therapy*, 16(2), 141-155.
- Träff, U. (2013). The contribution of general cognitive abilities and number abilities to different aspects of mathematics in children (Vol. 116:2, s. 139-156): *Journal of experimental child psychology* (Print) Elsevier. (Lokalisert på <http://dx.doi.org/10.1016/j.jecp.2013.04.007>).
- Undheim, J. O. (1981). On intelligence I: Broad ability factors in 15-year-old children and Cattell's theory of fluid and crystallized intelligence. *Scandinavian Journal of Psychology*, 22, 171-179. doi: 10.1111/j.1467-9450.1981.tb00391.x
- Undheim, J. O. & Gustafsson, J. E. (1987). The Hierarchical Organization of Cognitive Abilities: Restoring General Intelligence Through the Use of Linear Structural Relations (LISREL). *Multivariate Behavioral Research*, Vol. 22 (2), 149-171. doi: 10.1207/s15327906mbr2202_2
- Vaskinn, A. & Egeland, J. (2012). Testbruksundersøkelsen: En oversikt over tester brukt av norske psykologer. *Tidsskrift for norsk psykologforening*, 49, 658-665.
- Vejleskov, H. (1968). An analysis of Raven matrix responses in fifth grade children: I. *Scandinavian Journal of Psychology*, 9(3), 177-186.
- Wolff, U. & Gustafsson, J.-E. (2015). Structure of phonological ability at age four. *Intelligence*, 53, 108-117. doi: <http://dx.doi.org/10.1016/j.intell.2015.09.003>
- Wongupparaj, P., Kumari, V. & Morris, R. G. (2015). A Cross-Temporal Meta-Analysis of Raven's Progressive Matrices: Age groups and developing versus developed countries. *Intelligence*, 49, 1-9. doi: 10.1016/j.intell.2014.11.008