

Analisis Item Response Theory Raven's Coloured Progressive Matrices pada sampel Anak Usia Dini

*(Item Response Theory analysis of Raven's Coloured Progressive Matrices
in a sample of Preschool Children)*

Medianta Tarigan dan Fadillah*

Institut Teknologi Bandung, Bandung, Indonesia

*fadillah@itb.ac.id

Abstrak

Raven's Coloured Progressive Matrices merupakan alat tes kognitif anak yang sangat populer digunakan di Indonesia. Bentuk soalnya yang non-verbal dengan jumlah item yang singkat membuat alat ini dapat digunakan oleh individu dari berbagai latar belakang budaya. Kepopuleran alat tes ini di dunia praktis tidak diiringi dengan kebaruan data-data uji psikometrinya, terutama terkait kelayakan butir-butir soalnya. Untuk menguji kelayakan butir soal tersebut, dalam penelitian ini dilakukan analisis butir menggunakan teknik IRT dengan model logistik 3 (tiga) parameter (IRT 3 PL). Partisipan dalam penelitian ini adalah 467 siswa taman kanak-kanak berusia 5-7 tahun di Kota Bandung. Hasil analisis menunjukkan bahwa semua item dalam tes sesuai dengan 3-Parameter Logistic IRT Model yang mencakup daya beda, kesulitan dan peluang menebak. Terdapat 31 item, baik seri A, seri B, dan seri AB atau setara dengan 86% dari jumlah seluruh item CPM yang memiliki tingkat kesulitan di kategori rata-rata. Untuk estimasi parameter peluang menebak (*guessing*) hanya satu item yang dikategorikan sebagai item dengan peluang menebak yang ditolak. Sehingga sebesar 97% item CPM memiliki peluang menebak yang dapat diterima berdasarkan kriteria yang dipilih.

Kata kunci: *Colored Progressive Matrices; Item Response Theory; Kognitif*

Abstract

Raven's Coloured Progressive Matrices is a cognitive test that's used widely In Indonesia. Its short and non-verbal items makes it suitable for individuals of various cultural backgrounds. In practice, the popularity of this test is not accompanied with its up-dated psychometric property, especially in relation to the fitness of the item questions. In this study, a question analysis using the IRT technique examine the item questions, using the three parameter logistic model (3PL). The participants of this study are 467 kindergarten students aged 5-7 in the city of Bandung. All items of the test yield favorable statistics under the 3-Parameter Logistic IRT Model with regards to discrimination, difficulty, and guessing. There are 31 items, in either series A, series B, and series AB or the equivalent of 86% of the total number of CPM items that have a difficulty level in the average category. In estimating the probability parameter for guessing, only one item is categorized as an item with a rejected guessing probability. 97% Therefore, 97% of the CPM items generate an acceptable guessing probability based on the selected criteria.

Keywords: *Colored Progressive Matrices; Item Response Theory; Cognitive*

PENDAHULUAN

Kemampuan kognitif memungkinkan anak usia dini untuk memahami hubungan antara ide-ide, untuk memahami proses sebab dan akibat dan untuk

meningkatkan keterampilan analitisnya (Messick, 1984). Secara keseluruhan, dengan mengetahui kemampuan kognitif maka lingkungan dapat menyiapkan program-program pengembangan

keterampilan kognitif yang sesuai, yang tidak hanya dapat bermanfaat bagi anak di dalam kelas tetapi juga di luar kelas. Kemampuan kognitif dan kompetensi berbahasa anak usia dini memiliki nilai prediksi yang tinggi untuk kesiapan sekolah pada tahap berikutnya (Fisher dkk., 2004).

Raven's Colored Progressive Matrices (CPM) adalah salah satu tes kognitif yang paling banyak digunakan oleh praktisi di bidang pendidikan yang berfungsi mengidentifikasi tingkat kognitif anak. J. C Raven yaitu pembuat dari tes ini, dianggap telah berhasil melakukan terobosan dalam mendesain alat ukur yang secara efektif berhasil meminimalkan empat hambatan budaya yang diidentifikasi oleh Anastasia dan Urbina (1997) yaitu bahasa, konten tes, kemampuan membaca soal, dan kecepatan pengerjaan tes.

Dua komponen utama dari kemampuan kognitif umum (*g factor*) yang diukur oleh Raven adalah, seperti halnya yang disebutkan juga oleh teori Spearman, yakni: 1) *eductive ability*, yaitu kemampuan untuk membuat makna dan menghasilkan sebuah skema dari informasi yang tidak terstruktur, membingungkan, biasanya bersifat non-verbal, dan implisit sehingga kompleksitas yang dihadapi tadi menjadi lebih mudah ditangani; dan 2) *reproductive ability*, yaitu kemampuan untuk menyerap, mengingat, dan mereproduksi informasi yang telah dibuat eksplisit dan dikomunikasikan dari satu orang ke orang lain (Raven, 2008). Walaupun kemudian penelitian menunjukkan bahwa alat ukur yang dibuat Raven tidak sepenuhnya mengukur *g factor* (Gignac, 2015).

Penelitian atas validitas CPM dengan melibatkan 9.929 peserta di negara Brasil menggunakan analisis faktor menunjukkan bahwa CPM cukup memadai untuk menilai penalaran analogi pada anak-anak di Brasil, namun beberapa item dari tes asli tampaknya tidak berguna dan tidak memadai. Namun demikian, kerangka teoretis yang menjadi dasar pembuatan tes ini telah dikonfirmasi validitasnya sejak tes pertama kali diterbitkan (Pasquali dkk., 2002).

Walaupun telah mendapatkan pengakuan sebagai tes yang bebas budaya, namun beberapa

peneliti di berbagai negara tergerak untuk menguji kelayakan terhadap uji psikometri CPM untuk kelayakan penggunaannya. Penelitian yang dilakukan oleh Gintiliene dan Butkiene (2005) menunjukkan bahwa perbedaan jenis kelamin tidak berpengaruh secara signifikan terhadap skor CPM, namun tidak demikian dengan tingkat sosial ekonomi status, terutama terkait tingkat pendidikan orang tua serta lokasi demografi (kota atau pinggiran kota), yang ternyata cukup berpengaruh terhadap skor CPM. Penelitian ini kemudian juga menyimpulkan bahwa CPM merupakan alat yang cukup valid dan reliabel untuk digunakan sebagai alat identifikasi kemampuan kognitif pada anak-anak di Lithuania. Penelitian lain juga menunjukkan hasil yang sama yaitu tidak ditemukan perbedaan yang signifikan terkait perolehan skor CPM terhadap perbedaan jenis kelamin (Fajgelj, Bala & Tubic, 2007). Bahkan dalam penelitian ini ditambahkan dari hasil analisa terhadap norma skor CPM, tes ini dianggap terlalu mudah bagi partisipan dengan usia di atas 11 tahun karena adanya *Flynn's effect*, yang menyiratkan bahwa seseorang kemungkinan akan mencapai skor IQ yang lebih tinggi pada versi tes sebelumnya daripada versi saat ini. Pada kebanyakan kasus, nilai tes IQ di seluruh dunia telah meningkat dari waktu ke waktu: generasi yang lebih muda berkinerja lebih baik daripada generasi yang lebih tua.

Penelitian terhadap reliabilitas CPM dilakukan oleh Carlson dan Jensen (1981) dengan mengikutsertakan 783 partisipan penelitian yang terbagi menjadi 3 kelompok umur anak dan 3 etnis kelompok yang berbeda. Hasil penelitian ini menunjukkan bahwa tidak ditemukannya reliabilitas yang setara antara kelompok usia, namun tes ini menunjukkan reliabilitas yang baik pada 3 kelompok etnis yang berbeda. Beberapa tahun kemudian, Valencia (1984) juga melakukan penelitian untuk menyelidiki reliabilitas konsistensi internal CPM untuk anak dengan latar belakang etnis yang berbeda, yaitu Anglo dan Meksiko-Amerika, dengan latar belakang status sosial ekonomi rendah. Hasilnya menunjukkan bahwa reliabilitas untuk kedua kelompok etnis cukup tinggi sehingga

penelitian ini menyimpulkan bahwa CPM dapat dikatakan sebagai alat ukur kecerdasan nonverbal yang menjanjikan ketika digunakan dengan anak-anak dari latar belakang budaya dan bahasa yang beragam. Dengan menggunakan 372 partisipan berusia 5-11 tahun di daerah Omani, Kazem dkk. (2007) juga melakukan studi terhadap properti psikometri CPM. Hasilnya menunjukkan validitas dan reliabilitas yang baik sehingga dapat digunakan untuk mengukur inteligensi non-verbal bagi anak-anak di Omani. Penelitian ini kemudian menyarankan untuk juga dilakukan uji properti psikometri dengan menggunakan latar belakang budaya yang berbeda.

Penelitian terkait uji psikometri CPM di Indonesia masih terbatas, setidaknya untuk sepuluh tahun terakhir. Penelitian pernah dilakukan oleh Wulan (1982), dengan partisipan siswa SD di Yogyakarta. Penelitian ini menguji validitas CPM dengan melihat adanya hubungan yang signifikan antara hasil skor tes CPM dan tes WISC. Sementara itu, penelitian yang dilakukan oleh Masrun (1975) dilakukan untuk menguji reliabilitas menunjukkan nilai reliabilitas CPM bergerak antara 0,29 - 0,548. Sehingga bisa disimpulkan bahwa pada masa tersebut, CPM dianggap sebagai tes yang valid dan reliabel untuk digunakan sebagai alat ukur kognitif non-verbal di Indonesia.

Hidajat (1992) pernah meneliti tentang penyusunan norma tes CPM pada siswa sekolah dasar. Hal ini dilakukan untuk mengadaptasi norma yang lebih layak untuk digunakan di Indonesia. Rizkan, Shintia, dan Hartuti (2009) juga melakukan hal yang sama pada partisipan siswa sekolah dasar di Bengkulu dalam rangka membuat norma yang lebih layak bagi siswa yang berada di Luar Jawa karena adanya asumsi perbedaan tingkat kecerdasan antara siswa pada Pulau Jawa dan siswa di luar Pulau Jawa. Penelitian ini menyebutkan ditemukan perbedaan yang signifikan antara siswa dengan tingkat sosial ekonomi yang berbeda serta lokasi demografi, namun penelitian ini gagal membuat norma CPM yang baru karena terbatasnya jumlah data responden. Hingga saat ini, secara umum norma yang banyak digunakan di Indonesia masih norma asli yang

diberikan oleh Raven. Sementara, penggunaan CPM sebagai alat pengukuran kecerdasan peserta didik di Indonesia sudah cukup sering. Selain pengadministrasiannya yang mudah, CPM juga dapat memberikan hasil yang cepat sehingga alat ini menjadi salah satu rujukan alat ukur yang populer. Hanya saja, tidak seperti pada rangkaian tes *The Raven Progressive Matrices* (RPM) lainnya yang sudah banyak diteliti di Indonesia, masih sedikit yang meneliti terkait properti psikometrinya. Sebagai tes yang telah dikembangkan sejak 1974 dan telah digunakan di Indonesia pada kisaran tahun 1980-an, maka diperlukan pengujian properti psikometri patut dilakukan.

Diagnosis kognitif merupakan tugas mendasar dan krusial dalam dunia pendidikan. *Item Response Theory* (IRT) adalah metode diagnosis kognitif yang dapat memberikan parameter yang dapat diinterpretasikan (yaitu, sifat laten siswa, diskriminasi pertanyaan, dan tingkat kesulitan) sehingga dapat digunakan untuk menganalisis kinerja siswa. Komisi Tes Internasional menyarankan agar IRT digunakan untuk mengevaluasi instrumen psikologis yang ada dan telah banyak digunakan (Muñiz, 2011).

IRT merupakan *framework* pengukuran yang digunakan dalam proses mendesain dan menganalisis sebuah bentuk asesmen pendidikan atau psikologi (tes prestasi, skala penilaian, *inventory*, atau instrumen lain) yang mengukur *mental traits*. Pada dasarnya, IRT dapat digunakan untuk pengukuran kognitif maupun non-kognitif. Namun memang, IRT menjadi lebih populer di antara para ahli pengukuran, terutama terkait pengukuran kognitif. Di dalam pengukuran kognitif, analisis butir soal menjadi hal yang penting. Analisis butir soal merupakan proses melakukan telaah terkait kualitas butir soal pada tes. Tujuannya tentu saja untuk menentukan kelayakan atau validitas alat tes tersebut.

Dalam pengukuran kognitif, terdapat dua pendekatan yang sering digunakan untuk melakukan analisis butir soal, yaitu *Classical Test Theory* (CTT) dan *Item Response Theory* (IRT). IRT merupakan metode yang lebih terbaru sebagai usaha menangani kelemahan-kelemahan yang ada

pada CTT. Dalam perjalanan penggunaannya di bidang statistika, CTT memiliki beberapa kelemahan lain misalnya, daya beda butir soal yang tergantung pada kelompok subyek yang mengerjakan, karakteristik butir berubah seiring waktu berjalan, skor tes yang linear, tidak didasari teori yang kuat untuk menentukan bagaimana peserta memperoleh tes yang sesuai dengan kemampuan peserta yang bersangkutan serta Standard Error Measurement (SEM) berlaku pada seluruh peserta tes. Sementara itu, IRT menggunakan model probabilitas, di mana probabilitas subjek menjawab butir dengan benar bergantung pada kemampuan subjek dan karakteristik butir. Dengan kata lain, peserta tes yang memiliki kemampuan tinggi mempunyai probabilitas menjawab benar lebih besar dibandingkan peserta tes yang berkemampuan rendah (Magno 2009).

Oleh karena itu, penelitian ini menelaah kualitas kelayakan item alat ukur kecerdasan CPM. Analisis akan dilakukan melalui teknik analisis butir soal *Item Response Theory* (IRT). Penelitian ini diharapkan mampu memberi manfaat berupa informasi terkini mengenai kualitas item CPM dan menjadi acuan bagi penelitian yang memiliki kaitan dengan CPM selanjutnya.

1.1. *Raven's Colored Progressive Matrices* (CPM)

Colored Progressive Matrices (CPM) adalah salah satu rangkaian tes dari *The Raven Progressive Matrices* (RPM) yang dibuat oleh J. C. Raven (1938). Tes ini digunakan untuk mengukur kemampuan kognitif umum. Terdapat perbedaan dari rangkaian tes RPM, yaitu (a) *Standard Progressive Matrices* (SPM), berisi item asli yang dibuat untuk mengukur kemampuan umum pada orang dewasa; (b) *Advanced Progressive Matrices* (APM), versi yang lebih sulit dari RPM yang digunakan untuk mahasiswa; sementara (c) *Colored Progressive Matrices* (CPM), merupakan versi sederhana untuk anak-anak dan orang dengan kemampuan intelektual terbatas. Seperti penelitian yang pernah dilakukan pada anak sekolah di Sabah, Malaysia untuk mengidentifikasi adanya anak-anak dengan kesulitan belajar terkait dengan kapasitas kecerdasannya (Hashmi & Gang, 2015).

CPM adalah tes Raven untuk anak-anak yang mirip dengan *Standard Progressive Matrices* (SPM), namun tes disajikan dengan dicetak tinta berwarna sehingga lebih mengandalkan visual daripada kemampuan penalaran (Farah dkk., 2015). Tes CPM biasanya digunakan untuk mengukur kecerdasan umum, di mana tes ini terdiri dari rangsangan nonverbal berupa bentuk-bentuk yang disajikan dalam matriks 3 x 3 dengan bidang kanan bawah yang sengaja dikosongkan kemudian individu diminta untuk memilih jawaban yang benar dari beberapa opsi jawaban yang diberikan (Raudies & Hasselmo, 2017). Terkait dengan model pilihan jawaban berupa pilihan ganda, disebutkan bahwa item dengan pilihan ganda bisa menyebabkan adanya efek menebak jawaban walaupun belum diketahui berapa besar efeknya dengan skor akhir sebuah tes (Jančařík & Kostelecká, 2015).

Sebagian besar rangsangan dicetak dengan latar belakang berwarna, sehingga tes ini dinamai *Colored Progressive Matrices* (CPM). Warna-warna tersebut hanya digunakan sebagai elemen yang menarik dan memotivasi untuk mempertahankan perhatian anak saat tes berlangsung (Muniz dkk., 2016). Pasquali dkk. (2002) menyatakan bahwa Raven menggunakan tiga teori untuk mengembangkan matriks termasuk CPM. Adapun teori yang dimaksud adalah (1) teori dua faktor Spearman (faktor g); (2) Teori Gestalt; dan (3) teori perkembangan kognitif. Terdapat beberapa kemampuan yang akan diukur pada tes CPM, yaitu seri A yang mengukur kemampuan *visuoperceptual*, Seri AB mengukur kemampuan simetris, dan Seri B mengukur kemampuan berpikir konseptual dan analogis (Ivano dkk., 2015).

Studi validitas mengenai CPM di antaranya dilakukan oleh Qiu dkk., (2020) yang meneliti penggunaan SPM dan CPM pada partisipan usia 8-16 tahun di China. Hasil penelitian ini menunjukkan adanya korelasi yang signifikan antara tes SPM dan CPM. Hanya saja, ada catatan tambahan yang diterima dari penelitian ini yaitu tidak disarankan untuk mentransformasi hasil CPM menjadi hasil IQ untuk partisipan usia di atas 10 tahun. Hal ini

dikarenakan bentuk item CPM yang lebih sederhana dan kecepatan pengerjaan yang berbeda dari SPM sehingga menyebabkan adanya lonjakan nilai persentil yang berbeda. Sehingga, sangat disarankan agar anak usia di bawah 10 tahun menggunakan CPM dan anak usia di atas 10 tahun menggunakan SPM. Sementara pada tahun 2016, Muniz dkk. melakukan penelitian yang bertujuan memverifikasi struktur faktor CPM. Penelitian dengan menggunakan sampel berukuran 1,279 anak menunjukkan bahwa faktor-faktor yang menyusun konstruk CPM dapat digunakan sebagai alat ukur kognitif anak-anak. Adapun reliabilitas dan validitas CPM telah diuji sebelumnya oleh Bildiren (2017) dengan mengambil sampel anak-anak usia 3-9 tahun. Reliabilitas CPM dalam penelitian ini diuji dengan menggunakan uji reliabilitas *two split-half* dan *test retest* menggunakan rumus Kuder Richardson-20 (KR-20) dan Spearman-Brown. Adapun uji validitas CPM dilakukan dengan menghitung korelasi Pearson Product Moment antara CPM dengan Bender-Gestalt Visual Motor Perception Test dan TONI-3. Berdasarkan penelitian ini, diperoleh kesimpulan bahwa CPM adalah alat ukur yang valid dan reliabel dalam mengukur *general skill* dan *mental capacity*. Penelitian lain menguji koefisien validitas prediktif CPM dengan skor tes matematika dan pemahaman bacaan di mana hasilnya menunjukkan koefisien validitas tergolong sedang dan hasil menyimpulkan bahwa tes CPM yang biasa digunakan untuk menilai kemampuan kecerdasan umum ternyata juga dapat memprediksi akademik siswa (Balboni dkk., 2010). Sejauh ini, tidak seperti rangkaian tes Raven yang lainnya, penelitian terkait kelayakan butir soal belum banyak diteliti.

1.2. Item Response Theory

Isu-isu yang muncul dalam analisis kuantitatif terkait pengukuran kognitif dalam era modern semakin menonjol. Metode pengukuran terbaru, termasuk metode *Item Response Theory* (IRT), memungkinkan estimasi kemampuan kognitif anak menjadi lebih efisien dengan tetap memastikan

bahwa anak diberikan item pengukuran dengan tingkat kesulitan yang tepat dan skor yang dihasilkan lebih sensitif sesuai dengan teori dan model pengukuran (Jacob & Rothstein, 2016). IRT adalah teori pengukuran yang diperkenalkan oleh D. N. Lawley melalui sebuah makalah pada tahun 1943. IRT adalah suatu prosedur untuk mengevaluasi kemampuan subjek tanpa bergantung pada item yang sama dalam suatu tes (Zanon dkk., 2016). IRT menghasilkan estimasi terhadap properti item, yaitu tingkat kesulitan dan diskriminasi (daya beda) item, selain mengestimasi tingkat kemampuan subjek tes. Model IRT juga menempatkan tingkat kemampuan subjek tes dan parameter kesulitan item pada skala yang sama sehingga memungkinkan untuk perbandingan antara kemampuan sifat laten dan karakteristik item (Allen dkk., 2014). IRT fokus pada bagaimana suatu item dapat menilai suatu konstruksi (Thorpe & Favia, 2012). Adapun 'kinerja' (*performance*) setiap item dapat dievaluasi dengan menggunakan parameter item yang digambarkan dalam suatu kurva (Yang & Kao, 2014).

IRT mensyaratkan tiga asumsi yang harus dipenuhi, yaitu unidimensi, independensi lokal, dan invariansi parameter (Rahmawati, 2014; Sijtsma & Junker, 2006; Yang & Kao, 2014). Asumsi unidimensi adalah asumsi yang mensyaratkan hanya terdapat satu faktor atau konstruk laten yang diukur oleh sejumlah item dalam suatu tes. Adapun independensi lokal merupakan kondisi di mana jawaban terhadap item dari setiap subjek secara statistik tidak saling berhubungan dengan jawaban untuk item lain. Asumsi invariansi parameter adalah item bersifat independen dalam berbagai karakteristik sampel dalam suatu populasi.

Dengan melihat tingkatan jawaban item, IRT dikelompokkan menjadi IRT dikotomus dan politemus. Kategori dikotomus adalah bentuk tes yang memiliki jawaban benar-salah, meskipun pilihan jawaban berbentuk pilihan ganda dengan lima opsi, akan tetapi untuk model benar-salah ini tetap dikategorikan sebagai dikotomus. Sedangkan politemus adalah bentuk item dengan jawaban tidak hanya benar atau salah (Ostini & Nering, 2012).

Sedangkan jika dilihat dari jumlah parameter yang diestimasi, model IRT dikelompokkan menjadi beberapa jenis, yaitu IRT model logistik 1, 2, dan 3 parameter. Tabel 1 menampilkan ringkasan ketiga model tersebut, dengan e adalah konstanta 2.718, L adalah penyimpangan logistik (logit) (Baker, 2001).

IRT banyak digunakan di antaranya dalam bidang pendidikan sebagai prosedur kalibrasi dan evaluasi item dalam tes, kuesioner, dan instrumen lainnya untuk menilai subjek pada kemampuan, sikap, atau sifat laten lainnya (An & Yung, 2014).

Oleh karena itu, IRT ini adalah pendekatan yang sesuai untuk mengukur kelayakan item CPM.

Adapun analisis kelayakan item ini dilakukan dengan menggunakan teknik *Item Response Theory* (IRT) model logistik 3 (tiga) variabel. Dalam penelitian ini hipotesis uji model adalah:

1. H_0 : item CPM mengikuti model IRT 3PLM
2. H_1 : item tidak mengikuti model IRT 3PLM

H_0 diterima jika nilai *p-value* lebih dari atau sama dengan taraf signifikan ($\alpha = 0,05$). Dengan kata lain, item sesuai dengan model IRT 3PLM merupakan item yang memiliki nilai *p-value* $\geq 0,05$.

Tabel 1. Model *Item Response Theory*

Model IRT	Persamaan	Parameter yang Diestimasi
1 PL	$P(\theta) = \frac{1}{1 + e^{-1(\theta-b)}}$	<ul style="list-style-type: none"> • tingkat kesulitan (b)
2 PL	$P(\theta) = \frac{1}{1 + e^{-L}} = \frac{1}{1 + e^{-a(\theta-b)}}$	<ul style="list-style-type: none"> • tingkat kesulitan (b) • daya beda (a)
3 PL	$P(\theta) = c + (1-c) \frac{1}{1 + e^{-a(\theta-b)}}$	<ul style="list-style-type: none"> • tingkat kesulitan • daya beda • <i>guessing</i> (c)

METODE

Penelitian ini menggunakan metode penelitian kuantitatif, yaitu penelitian yang melibatkan sejumlah prosedur statistik yang berperan penting dalam perhitungan faktor yang diukur melalui suatu alat tes (Fischer dkk., 2014). Adapun analisis item CPM ini menggunakan teknik IRT dengan model logistik 3 (tiga) parameter (IRT 3 PL). Analisis item dibantu oleh perangkat jMetrik versi 4.1.

Partisipan dalam penelitian ini adalah 467 siswa taman kanak-kanak berusia 5-7 tahun di Kota Bandung. Administrasi tes CPM dilakukan oleh dua orang psikolog dan didampingi oleh satu fasilitator guru kelas. Satu sesi kelompok terdiri dari maksimal 5 siswa untuk menjaga kelayakan proses pengambilan

Instrumen yang diuji dalam penelitian ini adalah alat ukur kognitif *Colored Progressive Matrices* (CPM), yang terdiri dari 36 item yang mengukur kemampuan penalaran figural. Tes ini tidak memiliki

batas waktu pengerjaan. Tes terdiri dari tiga set dari dua belas desain matriks (Bass, 2000).

HASIL

Hasil estimasi secara detail ditampilkan pada tabel yang telah dilampirkan. Berdasarkan hasil analisis diketahui bahwa rata-rata skor seri A paling tinggi daripada seri AB dan B meskipun nilai maksimal pada seri A ini adalah yang paling rendah jika dibandingkan dengan kedua set lainnya.

Selanjutnya, hasil penelitian yang terlampir dikategorikan berdasarkan kriteria estimasi parameternya (daya beda, tingkat kesulitan, dan *guessing*) dan dirangkum dalam beberapa tabel berikut ini. Kriteria pengategorian estimasi parameter berdasarkan kriteria menurut Baker (2001) (daya beda dan peluang menebak) dan De Ayala (2009) (tingkat kesulitan) yang tertera pada bagian lampiran.

Tabel 2. Rekapitulasi Daya Beda

Subtes	Seri A	Seri AB	Seri B
Tidak dapat membedakan	-	-	-
Sangat rendah	-	1	-
Rendah	1	-	-
Sedang	6	1	5
Tinggi	3	8	1
Sangat tinggi	2	2	6
Sempurna	-	-	-

Tabel 2 memberikan informasi distribusi seluruh item tes CPM yang telah terbagi menjadi beberapa kriteria bila ditinjau dari hasil estimasi daya bedanya. Item pada Seri A memiliki item dengan estimasi daya beda paling banyak kategori sedang, sedangkan item

pada Seri AB terbanyak pada kelompok tinggi, dan Seri B paling banyak pada kategori sangat tinggi.

Selanjutnya adalah Tabel 3 yang merangkum rekapitulasi item CPM berdasarkan estimasi tingkat kesulitannya berdasarkan model IRT 3 PL.

Tabel 3. Rekapitulasi Tingkat Kesulitan

Kategori	Seri A	Seri AB	Seri B
Mudah	-	-	-
Rata-rata	10	11	10
Sulit	2	1	2

Dari tabel 3 diketahui bahwa sebagian besar item tes CPM memiliki tingkat kesulitan rata-rata. Tabel 4 menunjukkan rekapitulasi item berdasarkan

pengategorian estimasi peluang menebaknya. Hanya terdapat satu item yang ditolak pada tes CPM ini, yaitu item pada Seri A.

Tabel 4. Rekapitulasi Peluang Menebak

Subtes	Peluang Menebak (c)	
	Diterima	Ditolak
Seri A	11	1
Seri AB	12	-
Seri B	12	-

DISKUSI

Selain estimasi parameter, informasi yang bisa didapatkan dari IRT adalah melihat kualitas butir dengan model atau kesesuaian butir (*fit* item). *Fit* item menjelaskan apakah butir soal berfungsi normal, melakukan pengukuran atau tidak (Sumintono & Widhiarso, 2015). Butir soal yang sesuai (item *fit*) berarti soal tersebut 'berperilaku' secara konsisten dengan apa yang diharapkan oleh model. Sedangkan butir soal tidak *fit* merupakan indikasi bahwa terjadi miskonsepsi pada partisipan tes terhadap butir

tersebut. Informasi miskonsepsi pada partisipan sangat berguna untuk memperbaiki kualitas butir sehingga miskonsepsi bisa dihindari kembali.

Dalam penelitian ini, sebagaimana yang tertera pada tabel hasil estimasi parameter pada bagian lampiran, diketahui bahwa sebanyak 5 (lima) item yang tidak sesuai dengan model IRT 3 PL, yaitu item nomor 5 dan 6 pada seri A, 2 dan 11 pada seri AB, dan nomor 5 pada seri B. Berdasarkan hasil ini diduga terjadi miskonsepsi pada subjek tes terhadap item tersebut. Sementara itu terdapat 31 item yang

secara statistik sesuai dengan model IRT 3PL (signifikan). Hasil ini menunjukkan bahwa CPM sudah cukup baik karena hanya sedikit item yang tidak sesuai dengan model IRT 3 PL.

Berdasarkan hasil yang disajikan pada Tabel 2, Tabel 3, dan Tabel 4 berikut diketahui bahwa sebagian besar item CPM memiliki item dengan daya beda yang masuk pada kategori "Sedang" dan "Tinggi", yaitu masing-masing 12 item sehingga keseluruhannya adalah 24 item ataupun setara dengan 67% dari seluruh item CPM. Adapun untuk tingkat kesulitan, baik seri A, seri B, dan seri AB yang memiliki item dengan kategori rata-rata adalah sebanyak 31 item CPM atau setara dengan 86% dari jumlah seluruh item CPM. Serta untuk estimasi parameter peluang menebak (*guessing*) hanya satu item yang dikategorikan sebagai item dengan peluang menebak yang ditolak. Sehingga sebesar 97% item CPM memiliki peluang menebak yang dapat diterima berdasarkan kriteria yang dipilih.

Apabila dibandingkan dengan penelitian-penelitian kualitas butir dengan menggunakan analisis IRT pada alat tes kecerdasan lain yang digunakan di Indonesia, maka hasil ini menunjukkan tingkat kelayakan yang cukup baik. Misalnya saja analisis butir untuk tes IST di mana hanya sekitar 50% saja pada butir soal di tiap subtes yang dianggap layak (Agustin & Sirodj, 2018), kualitas butir *Wonderlic Personnel Test* (WPT) yang menunjukkan hanya 62% butir yang dinyatakan layak (Tarigan & Fadillah, 2019), kualitas butir Tes Bakat Skolastik dimana sekitar 30% butirnya, baik verbal dan numerik memiliki tingkat peluang menebak yang tinggi (Setiawati dkk., 2018). Dengan melihat hasil pemeriksaan kualitas butir tersebut maka implikasinya adalah CPM, sebagai tes kognitif untuk anak yang masa ciptanya sudah sangat lama, masih dianggap layak untuk digunakan. Secara praktis, hasil penelitian ini dapat membantu memberikan keyakinan bagi para psikolog untuk menggunakannya sebagai salah satu alat untuk mengidentifikasi kondisi kecerdasan anak sesuai dengan konstruksi dan definisi kecerdasan yang dimaksudkan pada alat tes ini.

SIMPULAN

Berdasarkan analisis dan pembahasan pada bagian sebelumnya, dapat disimpulkan bahwa studi kelayakan item alat ukur kecerdasan CPM dapat dikatakan telah layak dan memenuhi model IRT 3 PL ditinjau dari estimasi parameter daya beda, tingkat kesulitan, dan peluang menebak item. Walaupun dinyatakan layak, perlu diteliti lebih lanjut terkait model soal dan pengaruhnya terhadap fungsi kognitif tertentu. Studi terbaru menunjukkan bahwa item pada CPM bukan merupakan satu set tugas yang homogen, artinya skor yang sama mungkin diperoleh dari dukungan kemampuan kognitif yang berbeda sehingga bisa saja memiliki nilai kualitatif tertentu (Smirni, 2020). Perlu ditelaah lebih lanjut pada kelompok item mana saja yang memiliki kelayakan yang lebih baik dibandingkan kelompok item yang lain.

DAFTAR PUSTAKA

- Agustin, D., & Sirodj, N. (2018). Analisis kualitas item Intelligence Structure Test (IST) melalui metode Item Response Theory (IRT). *In schema Journal of Psychological Research*, 4(2).
- Allen, V. D., Weissman, A., Hellwig, S., Maccann, C., & Roberts, R. D. (2014). Development of the situational test of emotional understanding - brief (STEU-B) using item response theory. *Personality and Individual Differences*, 65, 3–7. <https://doi.org/10.1016/j.paid.2014.01.051>
- An, X., & Yung, Y. (2014). Item Response Theory : What It Is and How You Can Use the IRT Procedure to Apply It. *SAS Institute Inc.*, 1–14.
- Anastasia, A & Urbina, S. (1997). *Psychological Testing – Seventh Edition*. New Jersey: Prentice Hall. Inc.
- Baker, F. B. (2001). *Item Response Theory the Basics of Item Response Theory the* (C. Boston & L. Rudner (eds.). ERIC Clearinghouse on Assessment and Evaluation.
- Balboni, G., Naglieri, J. A., & Cubelli, R. (2010). Concurrent and predictive validity of the

- Raven Progressive Matrices and the Naglieri Nonverbal Ability Test. *Journal of Psychoeducational Assessment*. <https://doi.org/10.1177/0734282909343763>
- Bass, N. T. (2000). *The Raven's Coloured Progressive Matrices Test: A pilot study for the establishment of normative data for Xhosa-speaking Primary School pupils in the Grahamstown region*. January, 156.
- Bildiren, A. (2017). *Reliability and Validity Study for the Coloured Progressive Matrices Test between the Ages of 3-9 for Determining Gifted Children in the Pre-School Period*. 5(11), 13–20.
- Carlson, J. S., & Jensen, C. M. (1981). Reliability of the Raven Colored Progressive Matrices Test: Age and ethnic group comparisons. *Journal of Consulting and Clinical Psychology*, 49(3), 320–322.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Publications.
- Fajgelj, S., Bala, G., & Tubic, T. (2007). Raven's colored progressive matrices (CPM) - basic metric characteristics and norms. *Psihologija*, 40(2), 293–308.
- Farah, S., Bakhiet, A., Said, N., Abdelrasheed, G., Lynn, R., & Essa, P. (2015). *A Study of the Colored Progressive Matrices in Kuwait A study of the Colored Progressive Matrices in Kuwait*. January.
- Fischer, H. E., & Boone, W. J. (2014). *Quantitative Research Designs and Approaches Quantitative Research Designs and Approaches University of Duisburg-Essen , Essen , Germany Knut Neumann Leibniz-Institute for Science and Mathematics Education (IPN), Kiel , Germany*. October 2015.
- Fisher, Laurel & Spencer, Fiona & Fisher, Ian & Ouvrier, Robert. (2004). Cognitive Screening for Young Children: Development and Diversity in Learning Contexts. *Journal of child neurology*. 19. 313-7.
- Gignac, Gilles. (2015). Raven's is not a pure measure of general intelligence: Implications for g factor theory and the brief measurement of g. *Intelligence*. 52. 71-79. 10.1016/j.intell.2015.07.006.
- Gintilienė, G., & Butkienė, D. (2005). Raveno Spalvotų Progresuojančių Matricių Standartizacija Lietuvoje. = *A Lithuanian standardization of the Raven's Coloured Progressive Matrices*. *Psichologija*, 32, 22–34. Retrieved from <http://ebsco.com/>.
- Hashmi, S. I., & Gang, C. A. G. (2015). No Title Performance of Students' with Learning Disabilities (LD) on Ravens' Coloured Progressive Matrices. *Southeast Asia Psychology Journal*, 3, 23–33.
- Hidajat, S. (1992). *Studi Tentang Penyusunan Norma Tes Coloured Progressive Matrices Pada Siswa Sekolah Dasar Usia 7-11 Tahun Di Kotamadya Surabaya Tahun 1991*. Universitas Surabaya.
- Ivano, F., Alessandro, A., Bruno, I., Chieffi, S., Carnevale, G., Iaccarino, L., Cimminella, F., Chiavazzo, A., & Garofalo, E. (2015). Qualitative patterns at Raven's colored progressive matrices in mild cognitive impairment and Alzheimer's disease. *Aging Clinical and Experimental Research*. <https://doi.org/10.1007/s40520-015-0438-9>
- Jacob, B., & Rothstein, J. (2016). The measurement of student ability in modern assessment systems. *Journal of Economic Perspectives*. <https://doi.org/10.1257/jep.30.3.85>
- Jančařík, A., & Kostelecká, Y. (2015). The scoring of matching questions tests: A closer look. *Electronic Journal of E-Learning*, 13, 270-276
- Kazem, A. M., Alzubiadi, A. S., Yousif, Y. H., Aljamali, F. A., Al-Mashhdany, S. I., Alkharusi, H. A., ... Al-Fori, S. M. (2007). Psychometric properties of Raven's Colored progressive matrices for Omani children AGED 5 through 11 years. *Social Behavior and Personality*, 35(10), 1385–1398. <https://doi.org/10.2224/sbp.2007.35.10.1385>

- Magno, Carlo. (2009). Demonstrating the Difference between Classical Test Theory and Item Response Theory Using Derived Test Data. *The International Journal of Educational and Psychological Assessment*, 1(1), 1-11.
- Masrun. (1975). Validitas dan Reliabilitas Tes Raven sebagai alat pengukuran kecerdasan anak di Indonesia. *Jurnal Psikologi*, 5(1), 1-13.
- Messick, S. (1984). Abilities and Knowledge in Educational Achievement Testing: The Assessment of Dynamic Cognitive Structure. *Social and Technical Issues in Testing: Implications for Test Construction and Usage*. 11.
- Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement*, 21(3), 215-237.
- Muniz, M., Federal, U., Carlos, D. S., Carlos, S., Regina, S., Universidade, P., Paulo, D. S., & Preto, R. (2016). *Factor structure of Raven's Coloured Progressive Matrices*. 259-272.
- Muñiz J., (2011), International strategies to improve tests and testing. Paper presented at the 12th European Congress of Psychology, Istanbul.
- Ostini, R., & Nering, M. (2012). Polytomous Item Response Theory Models. In *Polytomous Item Response Theory Models*. <https://doi.org/10.4135/9781412985413>
- Pasquali, L. (2002). Raven's Colored Progressive Matrices for Children: A validation study for Brazil. *Avaliação Psicológica*, 2, 95-110.
- Qiu, C., Hatton, R., & Hou, M. (2020). Variations in Raven's Progressive Matrices scores among Chinese children and adolescents. *Personality and Individual Differences*, 164, 110064. <https://doi.org/10.1016/j.paid.2020.110064>
- Rahmawati, E. (2014). Evaluasi Karakteristik Psikometri Intelligenz Struktur Test (IST). *Proceeding Seminar Nasional Psikometri*, 270-282.
- Raudies, F., & Hasselmo, M. E. (2017). A model of symbolic processing in Raven's progressive matrices. *Biologically Inspired Cognitive Architectures*. <https://doi.org/10.1016/j.bica.2017.07.003>
- Raven, John. (2008). *The Raven Progressive Matrices Tests: Their Theoretical Basis and Measurement Model. Uses and Abuses of Intelligence: Studies Advancing Spearman and Raven's Quest for Non-arbitrary Metrics*. Unionville, New York: Royal Fireworks Press.
- Rizkan, Shintia R., Hartuti P. (2009). Pengembangan Norma Tes Kecerdasan Caloured Progressive Matrices (CPM) pada siswa Sekolah Dasar di Provinsi Bengkulu. Laporan Penelitian. Universitas Bengkulu.
- Setiawati, F. A., Izzaty, R. E., & Hidayat, V. (2018). Analisis respons butir pada Tes Bakat Skolastik. *Jurnal Psikologi*, 17(1). <https://doi.org/10.14710/jp.17.1.1-17>
- Sijtsma, K., & Junker, B. W. (2006). Item Response Theory: Past Performance, Present Developments, and Future Expectations. *Behaviormetrika*, 33(1), 75-102. <https://doi.org/10.2333/bhmk.33.75>
- Smirni, D. (2020). The Raven's coloured progressive matrices in healthy children: A qualitative approach. *Brain Sciences*. <https://doi.org/10.3390/brainsci10110877>
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. September.
- Tarigan, M., & Fadillah, F. (2019). Analisa Item Response Theory Wonderlic Personnel Test (WPT). *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia*, 8(1).
- Thorpe, G. L., & Favia, A. (2012). Data Analysis Using Item Response Theory Methodology: An Introduction to Selected Programs and Applications. *Psychology*, 20, 1-34.
- Valencia, R. R. (1984). Reliability of the Raven coloured progressive matrices for Anglo and for Mexican-American children. *Psychology in the Schools*, 21(1), 49-52.
- Wulan, R. (1982). Laporan Penelitian tentang

Hubungan antara Tes Inteligensi Coloured Progressive Matrices dengan Wechsler Intelligence Scalea for Children di Kotamadya Yogyakarta. Proyek Pengembangan Ilmu dan Teknologi. Fakultas Psikologi UGM.

Yang, F. M., & Kao, S. T. (2014). Item response theory for measurement validity. *Shanghai Archives of Psychiatry*.

Zanon, C., Hutz, C. S., Yoo, H., & Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexao e Critica*. <https://doi.org/10.1186/s41155-016-00>

Naskah masuk : 21 Januari 2021

Naskah diterima: 22 Oktober 2021

Lampiran

Tabel L.1. Kriteria Daya Beda

Sumber: (Baker, 2001)	Daya Beda (a)	Kategori
Estimasi Daya Beda	0	Tidak dapat membedakan
	0.01–0.34	Sangat rendah
	0.35–0.64	Rendah
	0.65–1.34	Sedang
	1.35–1.69	Tinggi
	>1.7	Sangat tinggi
+ infinity (tak terhingga)	Sempurna	

Tabel L.2. Kriteria Tingkat Kesulitan

Sumber De Ayala (2009)	Tingkat Kesulitan	Kategori
Estimasi Tingkat Kesulitan	-infinity sampai +infinity	Rentang teoritis
	-3 sampai 3	Rentang khas
	kurang dari -2	Mudah
	-2 sampai 2	Rata-rata
	lebih dari 2	Sulit

Tabel L.3. Kriteria Peluang Menebak

Sumber Baker (2001)	Peluang Menebak	Kategori
Estimasi Peluang Menebak	0 sampai 0.35	Dapat diterima
	di atas 0.35	Tidak dapat diterima

Tabel L.4. Parameter IRT Model 3 PL CPM

No	Analisis Parameter Item						Statistik <i>Fit</i> Item	
	a	Ket. a	b	Ket. b	c	Ket. c	<i>P</i> -value	Ket
1a	1,2	Sedang	-0,88	Rata-rata	0,5	Ditolak	0,50	Signifikan
2a	2,21	Sangat tinggi	-1,16	Rata-rata	0,1	Diterima	0,85	Signifikan
3a	1,78	Sangat tinggi	-1,23	Rata-rata	0,0	Diterima	0,63	Signifikan
4a	1,38	Tinggi	-1,05	Rata-rata	0,1	Diterima	0,52	Signifikan
5a	1,51	Tinggi	-0,84	Rata-rata	0,1	Diterima	0,00	Tidak Signifikan

Tabel L.4. Parameter IRT Model 3 PL CPM (*lanjutan*)

<i>No</i>	Analisis Parameter Item						Statistik <i>Fit</i> Item	
	<i>a</i>	Ket. <i>a</i>	<i>b</i>	Ket. <i>b</i>	<i>c</i>	Ket. <i>c</i>	<i>P</i> - <i>value</i>	Ket
6a	1,62	Tinggi	-0,71	Rata-rata	0,08	Diterima	0,04	Tidak Signifikan
7a	0,81	Sedang	0,47	Rata-rata	0,12	Diterima	0,16	Signifikan
8a	0,53	Rendah	0,47	Rata-rata	0,11	Diterima	0,34	Signifikan
9a	0,72	Sedang	0,96	Rata-rata	0,18	Diterima	0,09	Signifikan
10a	0,7	Sedang	0,78	Rata-rata	0,12	Diterima	0,36	Signifikan
11a	0,79	Sedang	3,79	Sulit	0,08	Diterima	0,49	Signifikan
12a	1,23	Sedang	3,03	Sulit	0,07	Diterima	0,50	Signifikan
1ab	1,79	Sangat tinggi	-0,84	Rata-rata	0,09	Diterima	0,34	Signifikan
2ab	1,67	Tinggi	-0,57	Rata-rata	0,06	Diterima	0,01	Tidak Signifikan
3ab	1,37	Tinggi	-0,59	Rata-rata	0,12	Diterima	0,21	Signifikan
4ab	1,43	Tinggi	0,24	Rata-rata	0,11	Diterima	0,58	Signifikan
5ab	1,34	Tinggi	0,51	Rata-rata	0,16	Diterima	0,78	Signifikan
6ab	1,7	Tinggi	0,68	Rata-rata	0,07	Diterima	0,61	Signifikan
7ab	1,4	Tinggi	0,66	Rata-rata	0,15	Diterima	0,84	Signifikan
8ab	1,42	Tinggi	1,6	Rata-rata	0,08	Diterima	0,36	Signifikan
9ab	2,12	Sangat tinggi	1,14	Rata-rata	0,13	Diterima	0,82	Signifikan
10ab	1,69	Tinggi	1,22	Rata-rata	0,22	Diterima	0,83	Signifikan
11ab	0,84	Sedang	1,47	Rata-rata	0,12	Diterima	0,04	Tidak Signifikan
12ab	0,34	Sangat rendah	4,48	Sulit	0,16	Diterima	0,16	Signifikan
1b	1,76	Sangat tinggi	-1,01	Rata-rata	0,16	Diterima	0,24	Signifikan
2b	0,84	Sedang	-0,02	Rata-rata	0,1	Diterima	0,10	Signifikan
3b	1,76	Sangat tinggi	-0,18	Rata-rata	0,09	Diterima	0,68	Signifikan
4b	1,18	Sedang	0,14	Rata-rata	0,1	Diterima	0,50	Signifikan
5b	1,21	Sedang	0,82	Rata-rata	0,12	Diterima	0,00	Tidak Signifikan
6b	0,66	Sedang	1,22	Rata-rata	0,13	Diterima	0,44	Signifikan
7b	1,01	Sedang	1,53	Rata-rata	0,15	Diterima	0,25	Signifikan
8b	2,13	Sangat tinggi	2,12	Sulit	0,08	Diterima	0,33	Signifikan
9b	2,14	Sangat tinggi	1,75	Rata-rata	0,05	Diterima	0,16	Signifikan
10b	2,25	Sangat tinggi	1,66	Rata-rata	0,13	Diterima	0,13	Signifikan
11b	2,4	Sangat tinggi	1,73	Rata-rata	0,05	Diterima	0,12	Signifikan
12b	1,46	Tinggi	2,88	Sulit	0,07	Diterima	0,94	Signifikan