



ANALISIS RASCH TERHADAP INSTRUMEN GENERAL SELF-EFFICACY SCALE-12 VERSI BAHASA INDONESIA: STUDI KOMPARASI RATING SCALE MODEL (RSM) DAN PARTIAL CREDIT MODEL (PCM)

Rasch Analysis of an Indonesian version of the General Self-Efficacy Scale-12: A Comparison of Rating Scale Model (RSM) and Partial Credit Model (PCM)

Muhammad Dwirifqi Kharisma Putra¹ & Heri Retnawati²

¹Universitas Gadjah Mada

Bulaksumur, Caturtunggal, Daerah Istimewa Yogyakarta 55281

²Universitas Negeri Yogyakarta

Jl. Colombo No.1, Karang Malang, Daerah Istimewa Yogyakarta 55281

Email: muhammad.dwirifqi@mail.ugm.ac.id

Naskah Diterima Tanggal 27 Maret 2020 – Direvisi Akhir Tanggal 5 Mei 2020 –

Disetujui Tanggal 25 Mei 2020 – Publikasi Online: 28 Juni 2020

DOI: <https://doi.org/10.26499/ijeav3i1.55>

Abstract. Self-efficacy is one of the most researched topics in education and psychology, which grounded within social cognitive theory. Various instruments are available that can be used to measure general self-efficacy. As a foundational instrument in the measurement of general self-efficacy, the study investigated the psychometric properties of the General Self-Efficacy Scale-12 (GSES-12). Although the instrument has been adapted into *Bahasa*, the contribution of the Rasch measurement model in validating this instrument has never done before. By utilizing various advantages of the Rasch model, the polytomous Rasch model (RSM and PCM) were used. The data used in this study are secondary data from previous research. The sample consisted of 303 students (132 male, 171 female) with an age range 18 - 22 years old (mean age = 19.56, SD = 1.20). The results of Rasch's analysis show that the psychometric characteristics of GSES-12 are adequate. The data fit to the Partial Credit Model compared to Rating Scale Model. All assumptions in applying the Rasch model are fulfilled, and also the item fit to the model. Implications and suggestions for future research are also discussed.

Keywords: Rasch model, psychometric properties, self-efficacy

PENDAHULUAN

Self-efficacy telah berkembang menjadi salah satu topik penting dalam bidang psikologi sejak teori belajar sosial muncul yang dipelopori oleh Albert Bandura (Celik, 2015). Namun di awal perkembangannya, *self-efficacy* merupakan konstruk yang berfokus pada konteks tertentu ataupun tugas khusus (lihat, Bandura, 1977), sehingga para tokoh mengembangkannya menjadi sebuah konstruk yang bersifat umum ataupun global (Sherer et al., 1982). Sejak saat itu, berbagai riset telah menghasilkan kerangka teoritis *self-efficacy* yang dapat berlaku secara umum, yaitu *general self-efficacy* (Shelton, 1990; Sherer et al., 1982), yang dianggap

berbeda dari *self-efficacy* yang berfokus pada konteks tertentu (Chen, Gully, & Eden, 2001). Perkembangan tersebut kini menjadikan *self-efficacy* dapat dikelompokkan sebagai *general*, *specific*, maupun *academic self-efficacy* (Choi, 2004), serta tokoh lain yang mengelompokkan *self-efficacy* yang dikembangkan oleh Bandura sebagai *task-specific* ataupun *general self-efficacy* sebagai *trait* (Chen, Gully, & Eden, 2001).

Sehingga, definisi awal *self-efficacy* yaitu “an individual's perceived capability of producing desired outcomes by taking necessary actions, and as such, it influences one's psychological functioning and performance behavior through choice of activities, the

amount of effort put into chosen activities, and perseverance in one's chosen activities" (Bandura, 1977) telah berkembang menjadi *general self-efficacy* yang didefinisikan sebagai "*a global construct is the composite of all life's successes and failures that are attributed to the self*" (Sherer et al., 1982). Kontribusi yang dilakukan Sherer et al. merupakan kontribusi signifikan terhadap perkembangan *general self-efficacy* (Shelton, 1990).

Sejalan dengan perkembangan definisi tersebut, instrumen yang dapat digunakan untuk mengukur *general self-efficacy* juga telah dikembangkan, seperti General Self-Efficacy Scale (GSES; Sherer et al., 1982), Generalized Self-efficacy Scale (GSE; Schwarzer & Jerusalem, 1995), serta General Self-Efficacy Scale 12 (GSES-12; Bosscher & Smit, 1998), maupun sejumlah besar instrumen lain yang dikembangkan secara khusus pada konteks tertentu seperti misalnya Career decision-making self-efficacy (CDMSE; Peterson, 1993) maupun Multidimensional Scales of Perceived Self-Efficacy (MSPSE; Bandura, 1990). Instrumen pertama yaitu GSES memang didesain untuk mengukur konstruk *self-efficacy* secara umum. Meskipun instrumen tersebut telah tersedia, struktur faktor instrumen tersebut baru dihasilkan oleh riset bertahun-tahun kemudian yang menemukan bahwa GSES yang berisi 17 item terdiri atas tiga aspek: *initiative*, *effort* dan *persistence* (Woodruff & Cashman, 1993). Lalu, modifikasi dilakukan terhadap GSES dengan mengurangi jumlah item menjadi 12 dan memindahkan item dari aspek lama ke aspek yang baru, dihasilkanlah GSES-12 (Bosscher & Smit, 1998).

Instrumen GSES-12 memiliki ciri khas tersendiri, hal ini disebabkan oleh fleksibilitas dari aspek-aspek yang diukur. Inisiatif, usaha, dan ketekunan dapat dengan mudah dipahami oleh para peneliti serta dapat diterapkan pada berbagai kondisi. Artinya, instrumen ini tidak berfokus pada konsep *self-efficacy* klasik yang terfokus pada tugas tertentu (lihat, Shelton, 1990). Meskipun *self-efficacy* merupakan konstruk yang erat dengan bidang pendidikan, dengan aspek GSES-12 yang secara operasional mudah dipahami, GSES-12 memiliki kemungkinan untuk dikembangkan agar dapat mencakup kondisi lain seperti dalam konteks perubahan lingkungan pembelajaran di tengah pandemic Covid-19 yang mengharuskan siswa belajar di rumah. Meskipun belum terdapat riset tersebut, jelas bahwa aspek inisiatif, usaha, dan ketekunan memiliki peluang besar untuk dikembangkan.

GSES-12 telah diadaptasi ke Bahasa Indonesia dalam studi yang dilakukan oleh Putra dan Tresniasari (2015), namun dikarenakan terdapat kalimat dalam item yang kurang tepat, dilakukan perbaikan terhadap item hasil adaptasi tersebut dan dilakukan uji validitas konstruk dengan metode *Bayesian Confirmatory Factor*

Analysis. Instrumen tersebut terbukti memiliki validitas konstruk yang baik serta membuktikan struktur faktor unidimensional (Putra, Rahayu, & Umar, 2019).

Dalam perspektif metodologis, GSES-12 dikembangkan dengan metode berbasis analisis faktor (Bosscher & Smit, 1998). Meskipun analisis faktor telah memberikan bukti validitas konstruk serta struktur faktor dari GSES-12, terdapat metodologi yang dapat digunakan untuk menggali aspek yang belum menjadi fokus analisis faktor, misalnya tingkat kesukaran (lokasi) item yang dibandingkan secara langsung dengan person trait level pada skala yang sama yang dapat diungkap oleh *latent trait theory* (misal, IRT ataupun Rasch model). Meskipun IRT, Rasch model, dan analisis faktor secara matematis ekuivalen, terdapat perbedaan secara filosofis (Putra, Hayat, & Suryadi, 2018). Namun, sejauh ini kontribusi Rasch model dalam evaluasi karakteristik psikometris GSES-12 secara khusus belum ditemukan, baik versi asli maupun versi Bahasa Indonesia. Rasch model telah diaplikasikan pada instrumen pengukuran *self-efficacy* lain seperti GSE (Bonsaksen et al., 2013) maupun Career Decision Self-efficacy Scale (Makransky, Rogers, & Creed, 2014) yang mengukur aspek *self-efficacy* yang berbeda dari GSES-12.

Penggunaan Rasch model (Rasch, 1960; Wright, 1968) terhadap GSES-12 versi Bahasa Indonesia memanfaatkan keunggulan yang dimiliki Rasch model seperti *specific objectivity*, *additivity*, dan *parameter separation* (Fisher, 1987; Perline, Wright, & Wainer, 1979; Rasch, 1966). Ketiga sifat khusus tersebut tidak dimiliki model lain (misal, teori tes klasik) sehingga, meskipun secara matematis sederhana, Rasch model dapat menghasilkan hasil estimasi yang memenuhi standar tinggi dalam pengukuran (Mair, 2018). Meskipun pada awalnya Rasch model dikembangkan untuk analisis terhadap data dikotomi, seiring perkembangannya telah tersedia model politomus, seperti Rating Scale Model (RSM; Andrich, 1978) dan Partial Credit Model (PCM; Masters, 1982). Kedua model ini dapat digunakan pada instrumen GSES-12 versi Bahasa Indonesia yang memiliki skor item politomus, serta memperpanjang temuan studi terdahulu yang terbatas pada pengujian struktur faktor.

Selain itu, banyak studi penerapan Rasch model di Indonesia yang "memilih" model RSM maupun PCM tanpa bukti statistik *goodness-of-fit*. Ada juga beberapa studi yang mengabaikan pengujian asumsi *local independence* (Putra, Suryadi, & Hayat, 2018). Studi ini memberikan perspektif lain tentang pemilihan model berdasarkan *goodness-of-fit* serta pengujian asumsi-asumsi dari Rasch model. Oleh karena itu, studi ini bertujuan untuk mengevaluasi karakteristik psikometris instrumen GSES-12 versi Bahasa Indonesia menggunakan

Polytomous Rasch Model (PRM). Hasil analisis diharapkan memperkaya temuan studi terdahulu terkait parameter item menggunakan perspektif model pengukuran.

METODE

Data

Data yang digunakan pada penelitian ini adalah data sekunder dari riset terdahulu yang dilakukan oleh Putra, Rahayu, dan Umar (2019). Adapun kaidah-kaidah penggunaan data sekunder dilakukan dengan mengacu pada tata cara yang dijelaskan oleh American Psychological Association (Trzesniewski, Donnellan, & Lucas, 2011), yakni mengharuskan pelaporan bahwa data yang digunakan merupakan data sekunder meskipun ketika pengguna data sekunder merupakan penulis yang sama. Berdasarkan data sekunder yang digunakan, partisipan dalam penelitian ini berjumlah 303 orang mahasiswa (132 laki-laki, 171 perempuan) dengan rentang usia 18-22 tahun (rata-rata usia = 19.56, SD = 1.20) yang berlokasi di Jakarta, Indonesia.

Instrumen GSES-12 versi Bahasa Indonesia

The General Self-efficacy Scale-12 (GSES-12; Bosscher & Smit, 1998) dikembangkan untuk mengukur *general self-efficacy*. Pada awalnya, Sherer et al. (1982) mengembangkan GSES versi 17 item. Lalu instrumen tersebut diuji kembali struktur faktornya oleh Woodruff dan Cashman (1993) yang menghasilkan informasi bahwa aspek yang diukur dalam instrumen ini adalah *initiative* (inisiatif), *effort* (usaha), dan *persistence* (ketekunan). Ketiga aspek tersebut berkorelasi positif satu sama lain. Bosscher dan Smit (1998) melakukan modifikasi terhadap struktur faktor dengan memindahkan satu item yang mengukur '*persistence*' menjadi '*initiative*'. Dalam versi terakhir, instrumen ini terdiri atas 12 item dengan format respons skala peringkat Likert 5-poin. Namun, adaptasi yang dilakukan dalam studi terdahulu telah melakukan perubahan skala peringkat Likert menjadi 4-poin: STS (sangat tidak setuju) sampai SS (sangat setuju) serta melakukan pengacakan terhadap nomor item (Putra, Rahayu, & Umar, 2019). Dalam studi ini, urutan item mengikuti versi asli GSES-12, yakni item yang mengukur *initiative* adalah item 1, 2, dan 3, item yang mengukur *effort* adalah item 4, 5, 6, 7, dan 8, serta item yang mengukur *persistence* adalah item 9, 10, 11, dan 12. Instrumen ini diterjemahkan ke dalam Bahasa Indonesia oleh komite dari Pusat Bahasa UIN Syarif Hidayatullah Jakarta dengan kualifikasi pendidikan doktor di bidang Pendidikan Bahasa Inggris serta proses *back-translation* oleh dosen Fakultas Psikologi UIN Syarif Hidayatullah Jakarta.

Polytomous Rasch Model (PRM)

Pada awalnya, Rasch model (Rasch, 1960) dikembangkan untuk analisis terhadap data dikotomi. Model ini memungkinkan peneliti untuk melakukan kalibrasi terhadap item dan orang pada garis ataupun skala yang sama. Ketika hal tersebut dilakukan, pemodelan terhadap prediksi respons seseorang terhadap suatu item dapat diketahui (Andrich & Marais, 2019). Parameter orang dan item dalam Rasch model diekspresikan dalam skala *logit*. Mengingat persamaan dasar Rasch model menggunakan fungsi logistik, data kontinu seperti skor total (hasil penjumlahan skor item) dapat dikonversi ke dalam skala interval. Misalnya, terdapat dua orang dengan skor total 14 dan 15 yang memiliki jarak sebesar 1, serta dua orang lainnya dengan skor total 17 dan 18 yang memiliki jarak 1. Meskipun sama-sama memiliki jarak 1, jarak 14 ke 15 dan jarak 17 ke 18 tidaklah sama karena belum berskala interval, sedangkan *person measure* hasil analisis Rasch berskala interval dalam satuan logit. Jarak 2 ke 3 logit sama besarnya dengan jarak 1 ke 2 logit (Wright, 1993). Meskipun secara matematis sederhana, model ini memiliki berbagai karakteristik khusus yang memenuhi standar tinggi dalam pengukuran (Mair, 2018). Seiring berjalannya waktu, telah tersedia model yang dapat digunakan untuk menganalisis data politomi. Model tersebut diklasifikasikan sebagai polytomous Rasch model (lihat, Engelhard & Wind, 2018).

Karena seluruh item pada instrumen GSES-12 yang diadaptasi dalam penelitian ini berbentuk skala peringkat Likert, Polytomous Rasch model (PRM) yang digunakan di antaranya adalah Rating Scale Model (RSM; Andrich, 1978) dan Partial Credit Model (PCM; Masters, 1982), serta beberapa model yang merupakan generalisasi atau modifikasi dari RSM dan PCM (lihat, Embretson & Reise, 2000). Adapun pemilihan model PCM dan RSM dilakukan dengan cara membandingkan *overall model fit* keduanya dengan Andersen Likelihood Ratio (LR) test (Andersen, 1973).

Dalam pengaplikasian PRM (RSM dan PCM), *threshold* (τ) kategori respons diikutsertakan dalam proses estimasi item. *Threshold* adalah titik transisi respons seseorang dari satu kategori ke kategori yang berdekatan pada skala Likert. Banyaknya *threshold* sama dengan jumlah kategori (k) dikurang 1. Dalam penelitian ini, setiap item memiliki empat opsi respons berskala ordinal, maka terdapat tiga *threshold* untuk masing-masing item GSES-12. Jika, data fit terhadap RSM, maka rentang *threshold* kategori respons untuk masing-masing item sama (hanya berlaku satu rentang *threshold* untuk seluruh item), begitupun dengan kurva karakteristik item. Namun jika data fit terhadap PCM, rentang *threshold* kategori respons untuk masing-masing item berbeda. Secara sederhana, PCM merupakan model yang

melonggarkan asumsi *threshold* kategori respons berlaku sama dalam RSM (Embretson & Reise, 2000).

Agar menghasilkan hasil estimasi yang akurat, ada beberapa asumsi dari Rasch model yang perlu dipenuhi, yaitu: (1) konstruk bersifat unidimensional, yaitu hanya satu trait yang diukur, (2) *local independence*, yaitu respons yang diberikan oleh penempeh tes terhadap satu item harus bersifat independen secara statistik dengan respons terhadap item lainnya dalam suatu tes, (3) paralel ICC, yaitu meskipun daya pembeda tiap item ditetapkan sebesar 1, data fit terhadap Rasch model (Mair, 2018). Ketiga asumsi tersebut diuji pada penelitian ini.

Selain itu, untuk menguji apakah item cocok (fit) dengan Model Rasch, terdapat dua statistik yang digunakan, yaitu: *chi-square* serta Infit Mean Square (MNSQ) dan Outfit Mean Square (MNSQ) yang dihitung dari pembagian *chi-square* dengan *degrees of freedom* (Linacre, 2018). Ketika *chi-square* tidak signifikan ($p > 0.05$), item fit terhadap model serta Infit dan Outfit berada dalam rentang nilai 0.5 hingga 1.5 yang menunjukkan bahwa item fit (Linacre, 2002). Ketika Infit dan Outfit lebih rendah dari 0.5, kondisi tersebut disebut sebagai '*overfit*' yaitu kondisi data terlalu ideal dan mudah diprediksi. Sedangkan ketika nilai lebih besar dari 1.5 disebut sebagai '*underfit*', yaitu kondisi data tidak konsisten dan sulit diprediksi (Linacre, 2018). Kedua kondisi tersebut dihasilkan oleh *chi-square* yang signifikan ($p < 0.05$).

Adapun tahap analisis Rasch yang akan dilakukan adalah: (1) pengujian *overall model fit* untuk menentukan model yang tepat, (2) pengujian asumsi *local independence*, (3) pemaparan statistik tingkat item (*measure* dan *threshold*) dan item fit, (4) representasi grafis keterkaitan orang-item dengan *Person-item map* dan informasi reliabilitas, dan (5) fungsi informasi tes dari instrumen GSES-12 versi Bahasa Indonesia. Dalam penelitian ini, parameter model diestimasi menggunakan metode estimasi *conditional maximum likelihood* (CML; Andersen, 1970) yang menghasilkan parameter item untuk selanjutnya digunakan dalam mengestimasi kemampuan responden dalam bentuk *plausible values* (Wu, 2005) dengan *package* 'eRm' (Mair & Hatzinger, 2007) yang diimplementasikan pada program R.

Berbeda dengan program analisis Rasch lain berbasis *unconditional (joint) maximum likelihood* (UCON atau JML seperti Winsteps), CML memungkinkan pengujian *goodness-of-fit* terhadap Rasch model baik *overall model fit*, item fit maupun *model comparison* (Andersen, 1973). Mengingat pengujian *goodness-of-fit* menggunakan prosedur umum, statistik inferensial perlu dilakukan dan tidak hanya bergantung pada *rule-of-thumb* (Perline, Wright, & Wainer, 1979). Namun, program berbasis UCON seperti Winsteps juga melaporkan

overall fit dalam bentuk Log-likelihood χ^2 yang hingga kini belum mendapat perhatian peneliti di Indonesia (Putra, Suryadi, & Hayat, 2018). Terdapat studi terkini yang melakukan perbandingan hasil Rasch analysis dengan CML melalui eRm dan JML berbantuan Winsteps sehingga menemukan perbedaan secara 'prosedural' dalam hal estimasi parameter antardua program tersebut (lihat, Linacre, 2020).

HASIL DAN PEMBAHASAN

Model Comparison dan Overall Model Fit

Pada tahap pertama dilakukan pengujian model fit secara keseluruhan penerapan Rasch model. Metode yang digunakan adalah *Andersen LR (Likelihood ratio) test* (Andersen, 1973). Jika nilainya tidak signifikan ($p > 0.05$), hal tersebut menunjukkan bahwa data fit terhadap model. Adapun data dianalisis menggunakan dua model, yaitu RSM dan PCM. Pada penelitian ini, hasil uji Andersen LR-test pada model RSM menghasilkan LR-value = 33.144, df = 10, dan p-value = 0.000. Sedangkan untuk PCM didapat LR-value = 21.027, df = 23, dan p-value = 0.579. Mengingat *p-value* LR-test PCM menunjukkan hasil yang tidak signifikan dapat disimpulkan bahwa data fit terhadap Rasch model dan mendukung penggunaan PCM. Hasil pengujian ini sekaligus menunjukkan bahwa asumsi unidimensionalitas dan paralel ICC dari PCM terpenuhi (Mair, 2018). Temuan ini sekaligus mengonfirmasi struktur faktor unidimensional dari GSES-12 dari penelitian terdahulu (Putra, Rahayu, & Umar, 2019; Putra & Tresniasari, 2015).

Local Independence

Setelah asumsi unidimensionalitas terpenuhi, selanjutnya dilakukan pengujian terhadap asumsi *local independence*. Kedua asumsi tersebut terkait erat karena ketika unidimensionalitas telah terbukti, respons pada masing-masing item memiliki sifat independen satu sama lainnya selain mengukur satu faktor yang sama (Embretson & Reise, 2000). Ketika asumsi ini diabaikan, seluruh hasil estimasi kemampuan orang, *test information function*, item parameter, dan juga statistik fit akan menghasilkan kesimpulan yang salah (Edwards, Houts, & Cai, 2018).

Pada penelitian ini, metode yang digunakan untuk menguji asumsi *local independence* adalah Q_3 (Yen, 1984). Dengan kriteria indeks Q_3 bahwa korelasi residual antarpasangan item tidak > 0.20 (Yen, 1993), dapat disimpulkan tidak ada item yang mengalami *local dependence*. Item yang memiliki korelasi residual tertinggi adalah pasangan item 5 dan item 8 sebesar 0.16 yang nilainya < 0.20 . Dengan kata lain, asumsi *local independence* pada penelitian ini telah terpenuhi. Temuan ini sejalan dengan terpenuhinya asumsi unidimensionalitas.

Item Measure, Fit Statistics dan Threshold

Tabel 2 berisi gambaran tentang karakteristik psikometris dari instrumen GSES-12 di antaranya adalah item *fit statistics* dan parameter lokasi item untuk seluruh item GSES-12. Seperti yang dapat dilihat pada Tabel 1, semua item menunjukkan *chi-square* yang tidak signifikan sehingga Infit dan Outfit MNSQ seluruh item

berada pada rentang yang dapat diterima (0.5-1.5). Hal ini berarti seluruh item fit terhadap PCM berdasarkan dua indeks kecocokan item yang telah dijelaskan sebelumnya. Temuan ini dapat mengonfirmasi temuan studi terdahulu bahwa kedua belas item signifikan secara statistik dengan metode dan pendekatan berbeda (Putra, Rahayu, & Umar, 2019).

Tabel 1. Item measure, fit statistics, and threshold parameter

Item	Measure	χ^2	p-value	Infit MNSQ	Outfit MNSQ	Threshold d 1	Threshold 2	Threshold d3
SEff 6	1.511	277.758	0.838	0.928	0.917	0.386	1.929	2.220
SEff 2	1.083	300.585	0.512	1.012	0.992	0.020	1.326	1.902
SEff 8	0.755	299.223	0.534	0.914	0.988	-0.583	0.664	2.184
SEff 3	0.650	338.823	0.071	1.016	1.118	-0.340	0.424	1.865
SEff 12	0.540	336.405	0.084	1.156	1.110	-0.984	0.343	2.264
SEff 10	0.513	307.126	0.407	0.963	1.014	-0.461	0.123	1.878
SEff 1	0.062	263.447	0.947	0.948	0.869	-0.511	-0.054	0.750
SEff 9	0.036	336.072	0.086	1.120	1.109	-0.827	0.088	0.847
SEff 5	-0.091	190.965	1.000	0.538	0.630	-1.118	-0.804	1.649
SEff 4	-0.110	241.164	0.996	0.792	0.796	-1.360	-0.534	1.562
SEff 11	-0.145	283.613	0.769	0.929	0.936	-1.371	-0.353	1.289
SEff 7	-0.315	260.888	0.958	0.911	0.861	-1.032	-0.260	0.348

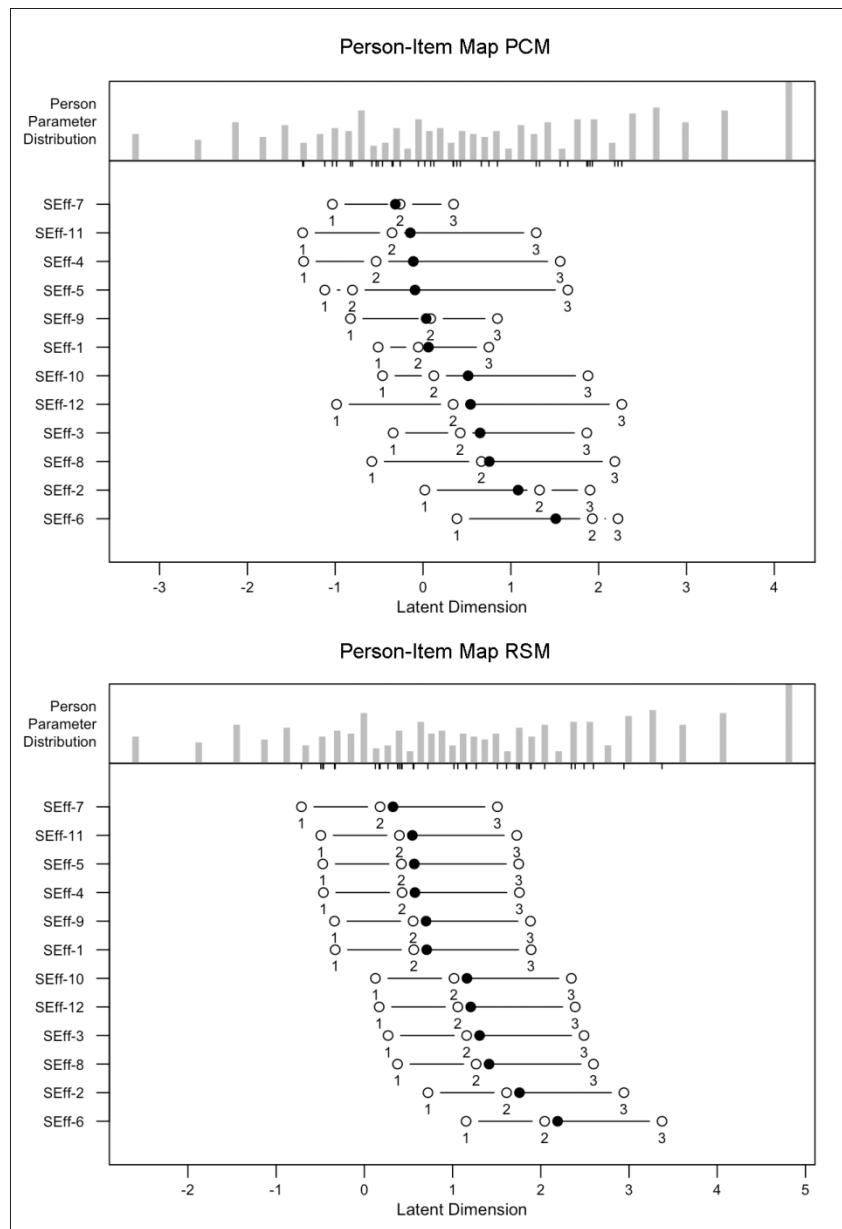
Lokasi item berada dalam rentang nilai -0.315 sampai 1.511. Dapat dilihat bahwa item SEff 7 “Ketika saya memutuskan untuk melakukan suatu pekerjaan, saya akan bertanggung jawab untuk menyelesaikannya dengan baik (*When I decide to do something, I go right to work on it*)” dengan lokasi pada -0.315 logit adalah item yang paling mudah untuk mendapatkan skor 4 (paling mudah disetujui) sedangkan item SEff 6 “Ketika saya harus melakukan suatu pekerjaan yang kurang menyenangkan untuk diri saya, saya akan tetap bertahan untuk mengerjakannya hingga terselesaikan dengan baik (*When I have something unpleasant to do, I stick to it until I finish it*)” yang berlokasi pada 1.511 logit merupakan item yang paling sulit untuk mendapatkan skor 4 (paling sulit untuk disetujui).

Informasi selanjutnya adalah *threshold* (τ) untuk masing-masing item. Seperti dapat dilihat pada Tabel 1, mengingat model yang digunakan adalah PCM, *threshold* untuk masing-masing item memiliki rentang yang bervariasi (misal, τ_1 untuk item 5 berada pada -1.118 logit, sedangkan τ_1 untuk item 4 berada pada -1.360 logit). Selain itu,

ditemukan bahwa *threshold* untuk masing-masing item nilainya naik. Hal ini menunjukkan bahwa seluruh kategori respons dalam GSES-12 versi Bahasa Indonesia berfungsi dengan baik, sesuai dengan apa yang diteorikan. Jika ada *threshold* yang terbalik (*disordered threshold*) (Adams, Wu, & Wilson, 2012) dibutuhkan modifikasi terhadap data. Misalnya, dengan cara menggabungkan kategori yang terbalik dengan kategori sebelumnya (*collapsing categories*). Dalam penelitian ini, *threshold* instrumen GSES-12 terbukti berfungsi dengan baik dan tidak memerlukan *category collapsing*.

Person-Item Map and Separation Reliability

Setelah dipaparkan informasi mengenai hasil estimasi parameter item, kaitan antara tingkatan *self-efficacy* yang dimiliki responden dan juga lokasi masing-masing item dapat dibandingkan secara bersamaan dengan menggunakan *person-item map* yang dikenal juga sebagai Wright Map (Wilson & Draney, 2002). Adapun Wright Map hasil analisis GSES-12 dengan 2 parameterisasi PRM dilihat pada Gambar 3.



Gambar 3. Person-Item Map GSES-12 versi Indonesia

Untuk melengkapi informasi pada Gambar 3, rata-rata tingkatan *trait* responden sebesar 0.816 ($SD = 1.918$) dengan rentangan -3.274 sampai 4.116 lebih tinggi dibandingkan rata-rata tingkat kesukaran item sebesar 0 ($SD = 0.561$). Hal ini terlihat ketika rata-rata tingkat kemampuan berada 1 SD (standar deviasi) di atas rata-rata tingkat kesukaran item. Hal ini menunjukkan bahwa kecenderungan responden memiliki *self-efficacy* yang tinggi diketahui dari banyaknya responden yang berada pada posisi 1 SD di atas rata-rata item. Terdapat sejumlah besar responden yang memiliki skor (total) maksimum. Selain itu, distribusi person cenderung mengikuti distribusi *uniform*. Meskipun distribusi *uniform* dalam Gambar 3 mencakup rentangan *trait* yang lebar, dengan jumlah item sebanyak 12 dengan 4 opsi respons, sebaran item tidak akan mungkin mampu menjangkau distribusi

uniform dari person selebar itu. Test Information Function dapat digunakan untuk mendapatkan gambaran rentang *trait*. Tes akan menghasilkan informasi yang maksimal di antara rentangan *trait* yang sangat lebar.

Dapat dilihat juga pada *person-item map* PCM, *threshold* kategori respons untuk masing-masing item berbeda-beda. Ada yang rentangannya lebar, ada juga yang sempit. Sedangkan, dalam *person-item map* RSM terlihat bahwa *threshold* untuk seluruh item memiliki rentangan yang sama. Hal ini membedakan PCM dan RSM yang berkaitan dengan *overall model fit* yang menunjukkan data fit terhadap PCM. Hal ini terjadi mengingat item 1 ($\tau_1 = -0.511, \tau_2 = -0.054, \tau_3 = 0.750$) misalnya memiliki rentang *threshold* yang sangat sempit apabila dibandingkan dengan item 12

($\tau_1 = -0.984, \tau_2 = 0.343, \tau_3 = 2.264$) yang memiliki rentang *threshold* yang lebar. Jika menggunakan RSM, *threshold* item 1 dan item 12 dianggap memiliki rentang yang sama padahal cukup jauh berbeda dan berpotensi menyebabkan kesalahan dalam penarikan kesimpulan. Selain itu, terlihat angka 1, 2, dan 3 yang masing-masing melambangkan *threshold* 1, 2 dan 3 dan berurut dari kiri ke kanan. Jika ada yang terbalik (misal, 2, 1, dan 3), terjadi *disordered threshold* yang harus ditelaah lebih lanjut oleh peneliti. Prosedur yang harus dilakukan ketika menghadapi *disordered threshold* dapat mengacu pada prosedur yang digunakan oleh Houghton et al. (2017).

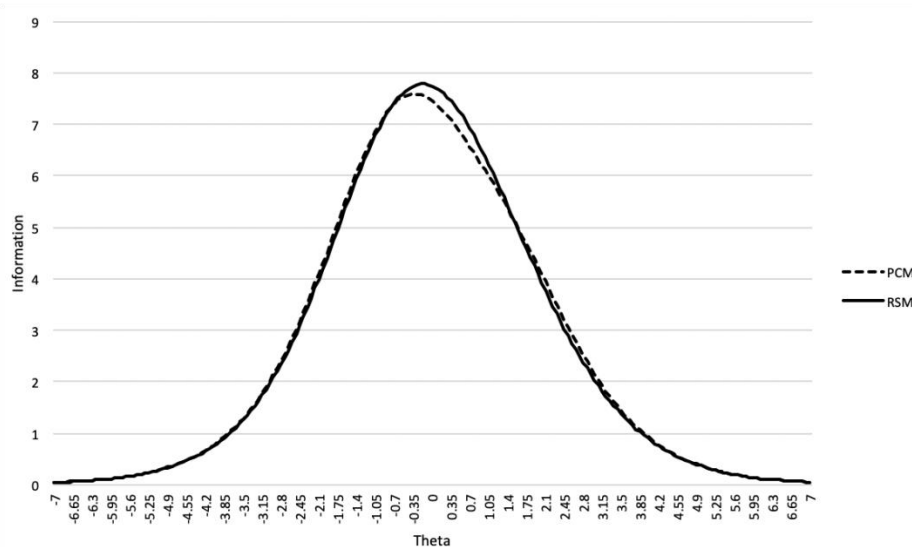
Adapun informasi reliabilitas dalam Rasch analysis didapat melalui *person separation reliability* (PSR; Wright & Stone, 1979) yang memiliki rentang nilai 0-1. Dalam penelitian ini didapat PSR sebesar 0.910 yang artinya konsistensi internal GSES-12 versi Bahasa Indonesia sangat baik karena nilainya lebih besar daripada 0.700 (Geldenhuis & Bosch, 2019), yang artinya sangatlah tinggi. Tingginya reliabilitas sebesar 0.910 melewati temuan pengembang GSES-12 yang menghasilkan reliabilitas sebesar 0.690 (Bosscher & Smit, 1998). Faktor penyebab tingginya reliabilitas dalam penelitian ini di

antaranya adalah lebarnya rentangan trait, *targeting item-person* (Linacre, 2018).

Berbeda dengan program lain berbasis UCON seperti Winsteps, eRm menghasilkan *koefisien* reliabilitas PSR dengan rentang nilai 0-1, sedangkan Winsteps melampirkan dua bentuk yaitu *koefisien* (rentang 0-1) serta *indeks* reliabilitas (misal, *person separation index*) yang memiliki rentang nilai 0- ∞ (Wright & Stone, 1999). Sedangkan software analisis Rasch lain berbasis MML (*marginal maximum likelihood*), seperti ConQuest, melaporkan *koefisien* reliabilitas dalam bentuk PSR, sehingga informasi reliabilitas dalam penilaian skala besar (misal, PISA) dilaporkan dalam bentuk koefisien dengan rentang nilai 0-1 (Adams, 2005). Komputasi manual serta rangkuman konsep dasar dari PSR dapat dilihat dalam Wright dan Stone (1999).

Test Information Function

Penerapan Rasch model (maupun model IRT lainnya) juga menghasilkan informasi berupa fungsi informasi tes (*test information function* atau TIF) yang menggambarkan nilai informasi bagi setiap tingkatan trait yang diukur beserta *standard error* masing-masing (Embretson & Reise, 2000). Adapun Gambar 4 berisi TIF dari GSES-12 versi Bahasa Indonesia.



Gambar 4. Kurva Test Information Function GSES-12 versi Indonesia

Seperti informasi pada Gambar 4, sepanjang rentangan trait perbedaan kurva TIF antara PCM dan RSM berbeda pada bagian puncak. Pada PCM, puncak kurva berlokasi pada -0.350 logit, sedangkan pada RSM puncak kurva berlokasi pada -0.210 logit. Mengingat PCM digunakan dalam studi ini, temuan ini menunjukkan bahwa GSES-12 merupakan alat ukur yang sangat baik dan optimal untuk mengukur responden yang berada dalam rentangan *self-efficacy* menengah (*moderate*) dengan rentang cakupan trait yang cukup lebar.

Saat posisinya sangat tinggi (+3) dan sangat rendah (-3) barulah keakuratannya menurun.

SIMPULAN

Artikel ini merupakan kontribusi pertama Rasch model dalam mengevaluasi karakteristik psikometris instrumen GSES-12 yang merupakan instrumen pengukuran *general self-efficacy*. Berdasarkan hasil analisis Rasch, instrumen GSES-12 versi Bahasa Indonesia terbukti memiliki karakteristik psikometris yang baik. Data yang

digunakan fit terhadap PCM, begitupun dengan terpenuhinya seluruh asumsi dari penerapan Rasch model. Pada tingkat item ditemukan bahwa seluruh item fit terhadap model. Temuan ini melengkapi bukti validitas konstruk dari penelitian terdahulu. Ke depannya, instrumen ini dapat digunakan pada karakteristik sampel maupun konteks lainnya di luar bidang pendidikan dan psikologi.

REFERENSI

- Adams, R. J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation*, 31(2-3), 162-172
- Adams, R. J., Wu, M. L., & Wilson, M. (2012). The Rasch rating model and the disordered threshold controversy. *Educational and Psychological Measurement*, 72(4), 547-573.
- Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society: Series B*, 32(2), 283-301
- Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38(1), 123-140.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561-573.
- Andrich, D., & Marais, I. (2019). *A course in rasch measurement theory: Measuring in the educational, social and health sciences*. Singapore: Springer Nature Singapore Pte Ltd.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191-215
- Bandura, A. (1990). *Multidimensional scales of perceived self-efficacy*. Stanford, CA: Stanford University Press.
- Bonsaksen, T., Kottorp, A., Gay, C., Fagermoen, M. S., & Lerdal, A. (2013). Rasch analysis of the general self-efficacy scale in a sample of persons with morbid obesity. *Health and Quality of Life Outcomes*, 11:202
- Bosscher, R. J., & Smit, J. H. (1998). Confirmatory factor analysis of the General Self-Efficacy Scale. *Behaviour Research and Therapy*, 36(3), 339-343.
- Celik, E. (2015). Mediating and moderating role of academic self-efficacy in the relationship between student academic support and personal growth initiative. *Australian Journal of Career Development*, 24(2), 105-113
- Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new General Self-Efficacy Scale. *Organizational Research Methods*, 4(1), 62-83.
- Choi, N. (2004). Sex role group differences in specific, academic, and general self-efficacy. *The Journal of Psychology*, 138(2), 149-159
- Edwards, M. C., Houts, C. R., & Cai, L. (2018). A diagnostic procedure to detect departures from local independence in item response theory models. *Psychological Methods*, 23(1), 138-149
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associate, Inc.
- Engelhard, G., & Wind, S. A. (2018). *Invariant measurement with raters and rating scales: rasch models for rater-mediated assessment*. New York, NY: Routledge
- Fischer, G. H. (1987). Applying the principles of specific objectivity and of generalizability to the measurement of change. *Psychometrika*, 52(4), 565-587
- Geldenhuis, M., & Bosch, A. (2019). A rasch adapted version of the 30-item bem sex role inventory (bsri). *Journal of Personality Assessment*. Advance online publication. <https://doi.org/10.1080/00223891.2018.1527343>
- Houghton, S., Wood, L., Marais, I., Rosenberg, M., Ferguson, R., & Pettigrew, S. (2017). Positive mental well-being: a validation of a rasch-derived version of the Warwick-Edinburgh Mental Well-Being Scale. *Assessment*, 24(3), 371-386
- Linacre, J. M. (2002). What do Infit and Outfit, Mean-square and Standardized mean? *Rasch Measurement Transactions*, 16(2), 878
- Linacre, J. M. (2018). *Winsteps® Rasch measurement computer program user's guide*. Beaverton, OR: Winsteps.com
- Linacre, J. M. (2020). CMLE - a problem, its solution and a useful approximation. *Rasch Measurement Transactions*, 33(1), 1749-1753
- Mair, P. (2018). *Modern psychometrics with R*. Cham, Switzerland: Springer International Publishing AG
- Mair, P., & Hatzinger, R. (2007). Extended Rasch Modeling: The eRm Package for the Application of IRT Models in R. *Journal of Statistical Software*, 20(9), 1-20
- Makransky, G., Rogers, M. E., & Creed, P. A. (2014). Analysis of the construct validity and measurement invariance of the career decision self-efficacy scale: a rasch model approach. *Journal of Career Assessment*, 23(4), 645-660
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174
- Perline, R., Wright, B. D., & Wainer, H. (1979). The Rasch model as additive conjoint measurement. *Applied Psychological Measurement*, 3(2), 237-255

- Peterson, S. L. (1993). Career decision-making self-efficacy and institutional integration of underprepared college students. *Research in Higher Education*, 34(6), 659-685
- Putra, M. D. K., & Tresniasari, N. (2015). Pengaruh dukungan sosial dan self-efficacy terhadap orientasi masa depan pada remaja. *TAZKIYA Journal of Psychology*, 3(1), 71-82.
- Putra, M. D. K., Rahayu, W., & Umar, J. (2019). Indonesian-language version of general self-efficacy scale-12 using Bayesian confirmatory factor analysis: a construct validity testing. *Jurnal Penelitian dan Evaluasi Pendidikan*, 23(1), 12-25
- Putra, M. D. K., Suryadi, B., & Hayat, B. (2018). *What we can learn from Indonesian scholars about the Rasch model*. The paper is presented at The 2nd International Conference on Educational Research and Evaluation (ICERE), Jakarta, Indonesia, November 23-24
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research. (Expanded edition, 1980. Chicago, IL: University of Chicago Press)
- Rasch, G. (1966). An item analysis which takes individual differences into account. *The British Journal of Mathematical and Statistical Psychology*, 19(1), 49-57
- Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio* (pp. 35-37). Windsor, UK: NFER-NELSON.
- Shelton, S. H. (1990). Developing the construct of general self-efficacy. *Psychological Reports*, 66, 987-994
- Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The self-efficacy scale: Construction and validation. *Psychological Reports*, 51(2), 663-671.
- Trzesniewski, K. H., Donnellan, M. B., & Lucas, R. E. (2011). *Secondary data analysis: an introduction for psychologists*. Washington, DC: American Psychological Association
- Wilson, M., & Draney, K. (2002). A technique for setting standards and maintaining them over time. In S. Nishisato, Y. Baba, H. Bozdogan, & K. Kanefugi (Eds.), *Measurement and multivariate analysis* (pp. 325-332). Tokyo, Japan: Springer-Verlag
- Woodruff, S. L., & Cashman, J. F. (1993). Task, domain, and general efficacy: A reexamination of the self-efficacy scale. *Psychological Reports*, 72(2), 423-432.
- Wright, B. D. (1968). Sample-free test calibration and person measurement. In *Proceedings of the 1967 Invitational Conference of Testing Problems*. Princeton, NJ: Educational Testing Service.
- Wright, B. D. (1993). Logits? *Rasch Measurement Transactions*, 7(2), 288
- Wright, B. D., & Stone, M. (1999). *Measurement essentials* (2nd ed.). Wilmington, DE: Wide Range, Inc.
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: MESA Press
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31, 114-128.
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8(2), 125-145.
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30(3), 187-213.