



Validasi Psikometrik Instrumen Kecemasan Statistik untuk Mahasiswa Psikologi di Indonesia Menggunakan Model Rasch

Siti Munawarah^{1✉}, Bahrul Hayat²

Universitas Islam Negeri Syarif Hidayatullah, Indonesia^{1,2}

e-mail : nana.munawarah@gmail.com¹, bahrulhayat@uinjkt.ac.id²

Abstrak

Kecemasan statistik merupakan secara signifikan mempengaruhi kinerja akademik, terutama di kalangan mahasiswa psikologi yang diharapkan dapat menguasai metode penelitian kuantitatif. Penelitian ini bertujuan untuk mengevaluasi validitas dan keandalan instrumen untuk mengukur kecemasan statistik, berdasarkan skala *Statistical Self-Efficacy for Psychologists* (SES-Psy). Sebanyak 373 mahasiswa psikologi dari enam perguruan tinggi di Jabodetabek yang telah menyelesaikan mata kuliah statistika dasar berpartisipasi dalam penelitian ini. Proses adaptasi mengikuti pedoman Beaton dengan empat orang ahli yang menilai validasi konten. Data dikumpulkan secara online dan dianalisis menggunakan Confirmatory Factor Analysis (CFA) dan model Rasch. Hasil CFA mengkonfirmasi beban faktor yang kuat dan indeks kesesuaian model yang sangat baik. Analisis Rasch memverifikasi bahwa instrumen tersebut memenuhi asumsi unidimensionalitas, independensi lokal, dan kesesuaian statistik. Semua kategori skala ditafsirkan dengan benar oleh responden dan tidak menunjukkan bias gender berdasarkan analisis DIF. Temuan ini menunjukkan bahwa instrumen yang diadaptasi secara psikometris valid untuk mengidentifikasi tingkat kecemasan statistik di kalangan mahasiswa psikologi di Indonesia. Dengan *standard setting* berbasis Rasch, diketahui bahwa mayoritas responden penelitian ini diklasifikasikan mengalami kecemasan statistik sedang hingga tinggi. Instrumen ini dapat digunakan pada lingkungan pendidikan maupun dimanfaatkan untuk pengembangan intervensi kecemasan statistik.

Kata kunci: Kecemasan statistik, Model Rasch, Adaptasi instrumen

Abstract

Statistical anxiety significantly affects academic performance, especially among psychology students expected to master quantitative research methods. This study aims to evaluate the validity and reliability of instruments adapted to measure statistical anxiety, based on the Statistical Self-Efficacy for Psychologists (SES-Psy) scale. A total of 373 psychology students from six universities in Greater Jakarta who completed basic statistics courses participated. The adaptation process adhered to Beaton's guidelines, with four experts assessing content validity. Data were collected online and analyzed using Confirmatory Factor Analysis (CFA) and Rasch models. Results from CFA showed strong factor loadings and excellent model fit indices. Rasch analysis confirmed the instrument meets assumptions of unidimensionality, local independence, and statistical appropriateness. Respondents correctly interpreted all categories and showed no gender bias according to DIF analysis. These findings suggest that the psychometrically adapted instruments are valid for identifying statistical anxiety levels among psychology students in Indonesia. Using a Rasch-based standard setting, most respondents were classified as having moderate to high statistical anxiety. This instrument can be applied in educational settings or for developing interventions targeting statistical anxiety.

Keywords: Statistics anxiety, Rasch Model, Instrument adaptation

PENDAHULUAN

Keterampilan statistik dibutuhkan dalam pendidikan tinggi, terutama dalam program ilmu sosial yang menekankan penguasaan metode penelitian kuantitatif (Garfield, 2002). Pemahaman siswa tentang konsep statistik sangat diperlukan dalam menyusun dan menganalisis data penelitian sosial. Dalam bidang Psikologi, keterampilan statistik diperlukan untuk menafsirkan hasil penelitian sehingga meminimalisir risiko kesalahan dalam melakukan intervensi psikologis (Kaufmann et al., 2022). Kelas statistik pada program sarjana psikologi bermanfaat bagi kemampuan kognitif lulusan psikologi (Dempster & McCorry, 2009). Terlebih lagi, ada prediksi bahwa penggunaan statistik Bayesian, dan pendekatan poin estimasi akan banyak digunakan untuk menggantikan hipotesis nol yang akan dianggap tidak relevan lagi (Howard, 2019).

Penelitian menunjukkan bahwa kinerja akademik mahasiswa dikelas statistik dipengaruhi oleh; sikap, kecemasan, dan efikasi diri mahasiswa terhadap statistik (Ogbonnaya et al., 2019). Kecemasan statistik terbukti menghambat performa siswa di kelas statistika (Hoegler & Nelson, 2018). Penelitian yang dilakukan pada mahasiswa psikologi di Inggris mengungkapkan bahwa kecemasan statistika memiliki pengaruh negatif terhadap performa *self-predicted test*, namun berasosiasi positif dengan performa dari *actual test* (Hunt et al., 2023). Selain itu, kecemasan terhadap statistik dipengaruhi oleh pola pikir negatif, kepuasan terhadap kebutuhan dasar psikologis, dan gender (Mcintee et al., 2022). Didapati mahasiswa perempuan mengalami kecemasan statistik yang lebih besar ketimbang mahasiswa laki-laki (Mcintee et al., 2022). Tidak hanya mahasiswa Psikologi, kecemasan statistik juga mempengaruhi performa akademik pada mahasiswa ilmu sosial (Hoegler & Nelson, 2018; Macher et al., 2015). Lebih tepatnya, dua pertiga dari mahasiswa program ilmu sosial, mengalami kecemasan statistik, terutama disebabkan oleh rasa kurang percaya diri saat menghadapi ujian (Condrón et al., 2018).

Karena kecemasan statistik mempengaruhi performa akademik mahasiswa, maka diperlukan instrumen untuk mengukur hal tersebut. Saat ini, ada beberapa instrumen untuk mengidentifikasi dan mengukur tingkat kecemasan terhadap statistik, antara lain: *Statistical Anxiety Scale* (SAS), yang merupakan versi alternatif yang lebih singkat dari *Statistics Anxiety Rating Scale* (STARS) (Vigil-Colet et al., 2008). *Statistics Comprehensive Anxiety Response Evaluation* (SCARE) (Griffith et al., 2014), dan *Statistics Anxiety Scale in the Online or Hybrid Setting* (SASOH) yang dikembangkan untuk mengetahui tingkat kecemasan statistik pada kondisi pembelajaran online (Liu, 2021). Selain itu, ada instrumen yang spesifik dikembangkan untuk mahasiswa psikologi, yaitu *Statistical Self-Efficacy for Psychologists* (Kaufmann et al., 2022). Dari seluruh instrumen yang mengukur kecemasan statistik tersebut, belum ada yang divalidasi secara psikometrik dengan metode Rasch model pada populasi mahasiswa psikologi di Indonesia.

Penelitian ini akan mengadaptasi instrumen SES-Psy, khususnya dimensi kecemasan statistik, ke dalam budaya dan bahasa Indonesia. Penelitian ini perlu untuk dilakukan karena kecemasan statistik mempengaruhi performa akademik secara signifikan dan memerlukan alat ukur yang sensitif dan akurat sesuai konteks budaya. Adapun validitas dan reliabilitasnya akan dianalisis menggunakan analisis faktor konfirmatori (CFA) dan Model Rasch. Analisis faktor konfirmasi adalah teknik statistik yang digunakan untuk menguji kesesuaian antara model pengukuran teoretis dengan data yang diamati. CFA sering digunakan dalam penelitian psikologi dan ilmu sosial. CFA adalah metode yang digunakan untuk mengetahui validitas konstruk dalam bidang psikologi, menggantikan pendekatan tradisional seperti korelasi Pearson dan Cronbach Alpha (Kyriazos, 2018). Indeks kesesuaian yang umumnya digunakan untuk mengevaluasi kesesuaian model adalah Chi-square, RMSEA, CFI, dan TLI (DoǺŶan, 2015). Masing-masing indeks tersebut memberikan informasi yang berbeda tentang kesesuaian model model.

Model pengukuran Rasch telah banyak digunakan di berbagai bidang, terutama untuk menilai instrumen psikologis dan asesmen pendidikan. Prinsip utama dari model Rasch adalah bahwa data harus sesuai dengan model probabilistik yang diasumsikan (Yu, 2020). Banyak penelitian terbaru yang menunjukkan peningkatan penggunaan Model Rasch untuk mengevaluasi sifat psikometrik dari instrumen psikologi (Dabaghi et al., 2020).

Rasch model memberikan kemudahan dalam pengembangan instrumen dengan memberikan informasi mengenai unidimensionalitas, kesesuaian item, dan reliabilitas (Tunç, 2023). Rasch model bekerja dengan cara mengubah respons yang berupa data kategorik, menjadi data berbentuk skala yang dapat dibandingkan (Aziz et al., 2013).

METODE

Penelitian ini menggunakan teknik *purposive sampling* untuk memilih enam perguruan tinggi di Jabodetabek berdasarkan pertimbangan ketersediaan program studi psikologi aktif serta kemudahan akses pengumpulan data dan efisiensi proses penelitian. *Non-probability sampling* digunakan untuk mendapatkan responden berdasarkan kriteria mahasiswa jurusan psikologi yang telah mengambil mata kuliah Statistika Dasar. Persetujuan dari komite etik Fakultas Psikologi UIN Syarif Hidayatullah Jakarta didapatkan sebelum proses pengambilan data. Pada akhirnya, diperoleh 337 responden yang telah menyetujui *informed consent* (Laki-laki 23%, Perempuan = 77%; Statistika Dasar = 38.3%, Statistika Lanjutan = 61.7%)

Instrumen yang diadaptasi dalam penelitian ini berupa skala Likert yang terdiri empat item yang berasal dari dimensi Kecemasan Statistik pada instrumen SES-Psy (Kaufmann et al., 2022). Kategori tanggapan yang tersedia adalah "sangat tidak sesuai dengan saya (1)" hingga "sangat sesuai dengan saya (4)".

Tabel 1. Instrumen Efikasi Diri Statistik untuk Mahasiswa Psikologi (SES-Psy); Dimensi Kecemasan Statistik

Items	Pernyataan
Anx_1	Sebelum ujian statistika, saya merasa lebih cemas dibandingkan dengan sebagian besar teman-teman saya
Anx_2	Saya merasa lebih cemas saat belajar statistika dibandingkan saat belajar mata kuliah lainnya
Anx_3	Perhitungan aritmatika pada statistika merupakan sumber kesulitan bagi saya
Anx_4	Saya lebih cemas saat sebelum ujian statistika dibandingkan sebelum ujian mata kuliah lain

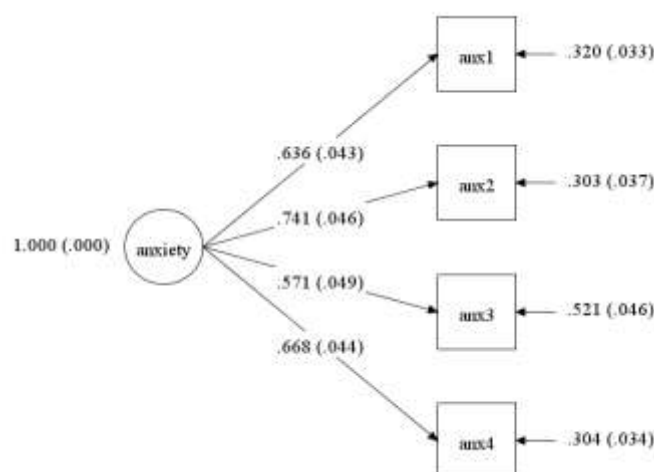
Proses adaptasi dilakukan dengan menggunakan prosedur Beaton (Beaton et al., n.d.). Diawali dengan melakukan menerjemahkan instrumen ke bahasa Indonesia dan *backward* terjemahan ke bahasa Inggris untuk memastikan tidak ada pergeseran makna. Pada penelitian ini, peneliti belum secara khusus melakukan tahap uji coba awal (*pre-test*) maupun wawancara kognitif (*cognitive interview*) untuk mengevaluasi pemahaman responden terhadap butir instrumen. Meskipun demikian, telah dilakukan sintesis terhadap *forward* dan *backward translation*, dan penilaian validitas konten oleh empat orang ahli yang merupakan dosen Statistika di Fakultas Psikologi UIN Syarif Hidayatullah, Jakarta.

Setelah proses adaptasi, kuesioner disebar menggunakan Google form. Data yang terkumpul kemudian dianalisis menggunakan *software* MPlus untuk analisis faktor konfirmatori (CFA) yang memastikan validitas konstruk. Selain itu juga digunakan *software* Winstep untuk analisis model Rasch yang memastikan validitas dan reliabilitas, serta deteksi bias dan penetapan *standard setting*.

HASIL DAN PEMBAHASAN

Validitas Konstruk dengan CFA

Validitas konstruk sangat penting dalam pengembangan instrumen, untuk memastikan bahwa item mengukur konstruk yang dimaksud. Analisis faktor konfirmatori (CFA) dianggap sebagai metode yang dapat diandalkan untuk menilai validitas konstruk dalam bidang psikologi, pendidikan, dan ilmu sosial (Umar & Nisa, 2020). Adapun indeks kesesuaian model CFA yang digunakan pada penelitian ini adalah; RMSEA dan SRMR < 0,08 serta nilai CFI/ FLI > 0,9 (Hu & Bentler, 1999).



Gambar 1. Diagram Jalur

Proses data cleaning perlu dilakukan untuk mengetahui responden yang tidak konsisten dalam memberikan jawaban (23 orang). Dari hasil uji validitas konstruksi dengan CFA didapatkan nilai *loading factor* yang tinggi untuk seluruh item (Gambar 1). Selain itu, nilai RMSEA, TLI/CFI, dan SRMS menunjukkan bahwa data sangat cocok dengan model. (Tabel 2). Sehingga dapat disimpulkan bahwa empat item pada instrumen ini mengukur konstruk yang sama, karena data teoritis sesuai dengan data empiris.

Tabel 2. Indeks Kesesuaian Model

Indeks Kesesuaian Model	Nilai
RMSEA (<i>Root Mean Square Error of Approximation</i>)	0.044
SRMR (<i>Standardised Root Mean Residual</i>)	0.012
TLI (<i>Tucker-Lewis Index</i>)	0.992
CFI (<i>Comparative Fit Index</i>)	0.997

Unidimensionalitas

Unidimensionalitas merupakan syarat utama dalam pemodelan Rasch, yang memastikan bahwa item-item pada instrumen hanya mengukur satu konstruk (Brentani & Golia, 2007). Unidimensionalitas dapat dideteksi melalui analisis komponen utama (PCA) dari residual dan statistik kesesuaian item (Brentani & Golia, 2007). Pada penelitian ini digunakan kriteria unidimensionalitas; RVE > 40% (Holster & Lake, 2016), dan *eigenvalue* dari residuals < 2 (E. V. Smith, 2002).

Tabel 3. Unidimensionalitas

Parameter	Value
<i>Raw Variance Explained by Measure (RVE)</i>	61,5 %
<i>Eigenvalue PCA Residual</i>	1,5

Dari hasil analisis Rach dengan Winstep, didapatkan nilai *Raw Variance Explained by Measure (RVE)* 61,5% dan nilai *eigenvalue* dari residual sebesar 1,5. Berdasarkan kedua indikator tersebut, dimensi kecemasan terhadap Statistika memenuhi asumsi unidimensionalitas, seperti yang ditunjukkan pada Tabel 3.

Lokal Independensi

Independensi lokal adalah asumsi penting dalam Teori Respons Item (IRT), yang menyatakan bahwa respon seseorang terhadap sebuah item tidak dipengaruhi oleh respon terhadap item lainnya, apabila sudah dikendalikan oleh variabel laten (Edwards & Houts, 2018). Pelanggaran terhadap independensi lokal ini perlu dipahami dan diatasi untuk menghasilkan pengukuran yang akurat dalam bidang pendidikan, kesehatan, dan psikologi (Edwards & Houts, 2018). Nilai korelasi residual menggambarkan korelasi antara residual item setelah efek model Rasch dihilangkan.

Tabel 4. Korelasi Residual antar Item

No.	Item	Item	Korelasi Residual
1	Anx_2	Anx_3	-0,43
2	Anx_1	Anx_3	-0,37
3	Anx_3	Anx_4	-0,32
4	Anx_2	Anx_4	-0,31
5	Anx_1	Anx_4	-0,29
6	Anx_1	Anx_2	-0,26

Pasangan item dengan nilai korelasi residual $\geq +0,5$ menunjukkan pelanggaran terhadap asumsi independensi lokal, sehingga item perlu dihapus, digabung/ testlet, atau direvisi. (Yen, 1984). Pada Tabel 4, nilai korelasi residual dari setiap pasangan item adalah $< +0,5$, sehingga dapat disimpulkan bahwa item dalam dimensi kecemasan, memenuhi asumsi independensi lokal.

Kesesuaian dengan Model Rasch

Statistik kesesuaian item sangat penting dalam pengukuran Rasch untuk memastikan unidimensi dan mendeteksi adanya gangguan dalam pengukuran. Nilai INFIT dan OUTFIT MNSQ dapat mengidentifikasi item yang berpotensi *misfit* (Padgett & Morgan, 2019). Indikator kesesuaian item yang digunakan pada penelitian ini adalah nilai MNSQ yang berada dalam rentang 0,5 hingga 1,5 (Linacre, 2012).

Tabel 5. Kesesuaian Item dengan Model Rasch

Item	Measure	MNSQ Infit	MNSQ Outfit
Anx_3	0,05	1,18	1,18
Anx_2	0,49	1,00	1,01
Anx_4	-0,51	0,9	0,89
Anx_1	-0,03	0,88	0,88

Tabel 5 menunjukkan bahwa seluruh item sesuai secara statistik dengan model Rasch. Hal tersebut menandakan bahwa respon peserta terhadap item-item pada instrumen ini dapat diprediksi oleh model Rasch. Selain itu, tabel 5 juga memberikan informasi bahwa pernyataan dalam Anx_2 (0,49 logit) lebih mudah direspons positif oleh responden, dibandingkan Anx_4 (-0,51 logit).

Indeks Separasi & Reliabilitas

Indeks separasi dan reliabilitas *person* dan *item* menunjukkan sejauh mana instrumen mampu membedakan responden berdasarkan tingkat kecemasannya, serta sejauh mana *item* memiliki variasi tingkat kesulitan yang mencerminkan perbedaan nyata antar individu, dengan konsistensi pengukuran yang baik. *Item* dan *Person* memiliki sebaran yang baik apabila nilai separasi > 2 (Linacre, 2012). Reliabilitas *item* dan *person* dianggap baik apabila nilainya $> 0,8$ (Bond & Fox, 2007). Dari tabel 6 diketahui bahwa nilai *Alpha Cronbach*, separasi *person-item*, and reliabilitas *person-item* pada instrumen penelitian ini dianggap baik.

Tabel 6. Separasi dan Reliabilitas Item-Person

	Indikator	Nilai
Item	Item Separation	2,73
	Item Reliability	0,88
	Alpha Cronbach	0,86
Person	Person Separation	2,21
	Person Reliability	0,83

Analisis Kategori Skala Peringkat

Selanjutnya, kita perlu menguji apakah peringkat skala 1-4 yang kita gunakan dalam dimensi Kecemasan Statistik berfungsi dan dipahami dengan baik oleh responden. Analisis kategori skala peringkat perlu dilakukan untuk memastikan pengukuran yang valid dan optimal dalam segala aspek (Linacre, 2002).

Tabel 7. Analisis Kategori Skala Peringkat

Label kategori	Observed Sample Average	Structure Measure	MNSQ Outfit
1	-3,90	None	0,90
2	-0,75	-4,10	0,96
3	1,95	0,23	1,01
4	3,74	3,87	1,04

Dari nilai *Observed Sample Average* pada Tabel 7, diketahui bahwa tingkat kecemasan responden yang dinyatakan dalam logit meningkat seiring dengan meningkatnya kategori skala. Hal ini menunjukkan bahwa kategori pada skala berfungsi dengan baik. Selain itu, keempat peringkat pada skala Likert memiliki nilai *Outfit* MNSQ mendekati satu, yang artinya, responden menggunakan kategori peringkat secara konsisten sesuai asumsi model Rasch (Linacre, 2002).



Gambar 2. Kurva Kategori Skala Peringkat

Dari kurva kategori skala peringkat pada Gambar 2, diketahui bahwa kategori satu (kurva paling kiri) dipilih oleh banyak responden dengan tingkat kecemasan statistik yang paling rendah, diikuti secara berurutan oleh peringkat dua, tiga dan empat (paling kanan). Keempat kurva tersebut juga saling berpotongan sesuai dengan urutan kategori, sehingga mendukung validitas kategori *rating scale*.

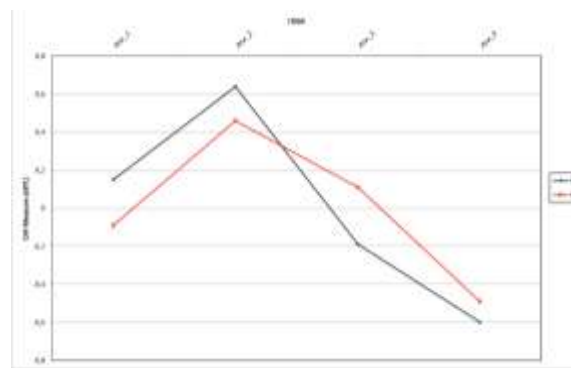
Differential Item Functioning (DIF)

Analisis *Differential Item Functioning* (DIF) sangat penting untuk memastikan keadilan dan validitas pengujian dengan mengidentifikasi item yang berfungsi secara berbeda di seluruh sub kelompok (Lee & Geisinger, 2014). Item terdeteksi DIF apabila nilai kontrasnya $\geq$ logit 0,40 dan nilai *p-value* $\leq$ 0,05 (A. B. Smith et al., 2016). Penelitian ini menguji adanya DIF berdasarkan gender (laki-laki dan perempuan) serta kelas statistik (dasar dan lanjutan).

Tabel 8. Nilai DIF kontras berdasarkan Gender dan Kelas Statistika

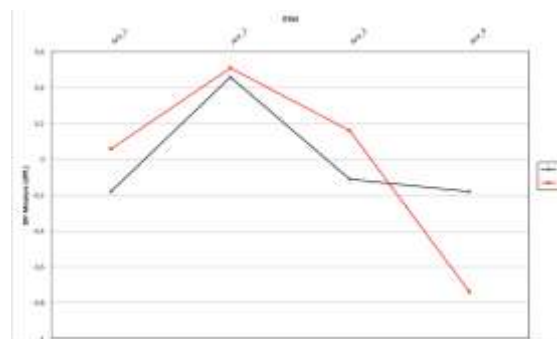
Item	DIF Contrast by Gender	P-value	DIF Contrast by Statistical Class	P-value
Anx_1	$\pm$ 0,24	0,42	$\pm$ 0,24	0,33
Anx_2	$\pm$ 0,18	0,54	$\pm$ 0,05	0,82
Anx_3	$\pm$ 0,30	0,30	$\pm$ 0,26	0,28
Anx_4	$\pm$ 0,11	0,72	$\pm$ 0,56	0,02

Dari gambar 3 kita ketahui ada perbedaan tingkat kesulitan item untuk kedua kelompok gender, namun informasi dari Tabel 8 menunjukkan bahwa perbedaan tersebut tidak signifikan. Hal ini menunjukkan bahwa seluruh item pada instrumen, bebas bias gender. Dengan kata lain, perbedaan skor kecemasan yang mungkin muncul antar kelompok gender mencerminkan perbedaan tingkat kecemasan yang sesungguhnya, bukan karena ketidakadilan item (Arslan et al., 2023).



Gambar 3. DIF Kontras berdasarkan Gender: Laki-laki (L) dan Perempuan (P)

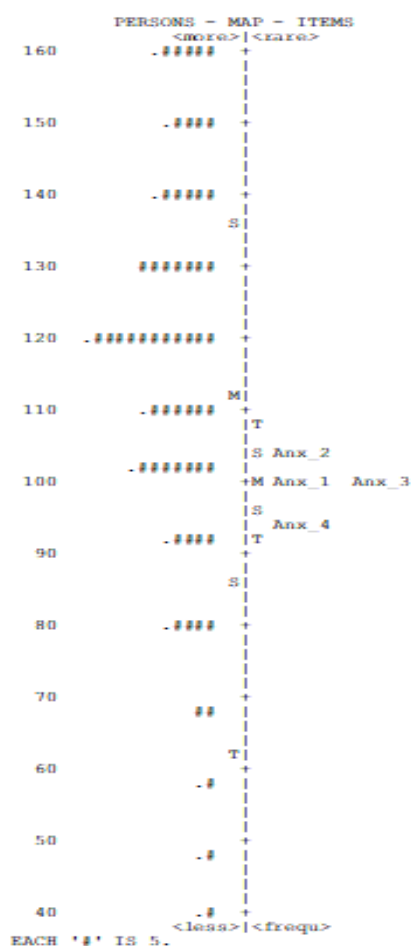
Sementara itu, dari gambar 4 dapat kita lihat DIF kontras terbesar pada item Anx_4 yang didukung dengan informasi dari Tabel 8 bahwa perbedaan tersebut signifikan. Dengan demikian item Anx_4 ini berpotensi mengandung bias pemahaman yang tidak sepenuhnya menggambarkan perbedaan dalam konstruk kecemasan statistik, namun menggambarkan perbedaan latar belakang pendidikan statistik dari responden. Implikasinya, apabila item Anx_4 ini disertakan dalam asesmen komparatif pada kelompok dengan pengalaman belajar statistik yang berbeda, akan berpotensi menghasilkan interpretasi yang keliru. Dengan demikian, item tersebut perlu penyempurnaan redaksi agar dapat dipahami dengan setara oleh semua kelompok dengan perbedaan latar belakang pendidikan statistik.



Gambar 4. DIF Kontras berdasarkan Kelas Statistik: Dasar (D) dan Lanjutan (L)

Wright Map

Wright map, juga dikenal sebagai *person-item map*, adalah alat yang berharga dalam teori respons item untuk memvisualisasikan keselarasan kemampuan orang dan kesulitan item pada skala yang sama (Kabic & Alexandrowicz, 2023). Wright map menampilkan skor orang dan kesulitan item dalam satuan logits pada skala yang sama, sehingga memungkinkan interpretasi simultan dari keduanya (Ismail et al., 2021).



Gambar 5. Wright Map

Setelah dilakukan transformasi skala logit dengan $U_{mean} = 100$ dan $U_{scale} = 10$, nilai tingkat kesulitan item dan kemampuan responden ditampilkan pada Peta Wright pada Gambar 5. Dari Peta Responden (kiri), diketahui bahwa distribusi tingkat kecemasan responden terhadap statistik berkisar antara 40 (Tidak cemas sama sekali) hingga 140 logit (Sangat cemas), dengan posisi netral adalah 100 ($U_{MEAN} = 100$). Dari *Wright map* diketahui bahwa sebagian besar responden berada di atas 100 logits. Hal ini menunjukkan bahwa sebagian besar responden dalam penelitian ini mengalami kecemasan terhadap statistik. Hal ini sejalan dengan penelitian-penelitian sebelumnya bahwa mahasiswa psikologi dan ilmu sosial cenderung mengalami kecemasan terhadap statistik (Condrón et al., 2018; Macher et al., 2015).

Standard Setting and Performance Level Descriptor

Kontribusi praktis dari penelitian ini adalah menetapkan standard setting untuk mengklasifikasikan tingkat kecemasan statistik mahasiswa. Model Rasch dapat memvalidasi level kecakapan dengan mengklasifikasikan responden ke dalam kelompok laten yang berbeda (Jiao et al., 2011). Rasch model menjembatani hasil pengukuran numerik ke dalam kategori interpretatif yang dapat langsung digunakan oleh pihak kampus, konselor ataupun praktisi pendidikan tinggi untuk mendeteksi dan memberikan intervensi psikologis yang sesuai.

Dalam mengklasifikasi mahasiswa berdasarkan tingkat kecemasan statistiknya, peneliti menggunakan nilai mean dan standard deviasi (logit) dari responden *ability/measure* untuk pengelompokan kriteria (MacCann, 2009). Adapun *standard setting* dari tingkat kecemasan statistik mahasiswa ditampilkan pada Tabel 9.

Tabel 9. Standard Setting and Performance Level Descriptors

Level Kecemasan Statistika	Kategori	Deskripsi
Sangat Tinggi	>Mean + 2SD	Pada kategori ini, mahasiswa memiliki kecemasan yang secara signifikan menghambat pembelajaran mereka. Mereka mengalami ketakutan dan kebingungan dengan konten statistik. Mahasiswa dalam kelompok ini membutuhkan dukungan dan intervensi untuk membantu membangun kepercayaan diri dan meningkatkan pemahaman tentang Statistik.
Tinggi	Mean + 1SD to Mean + 2SD	Mahasiswa merasa sangat cemas dengan materi Statistik. Mereka menghindari keterlibatan dengan statistik dan membutuhkan bantuan yang cukup besar untuk menyelesaikan tugas. Kecemasan mereka menyebabkan kesulitan dalam menerapkan pengetahuan statistik.
Sedang	Mean -1SD to Mean +1SD	Pada kelompok ini, mahasiswa mengalami kecemasan sedang, yang memengaruhi kemampuan mereka untuk fokus pada tugas statistik. Meskipun mereka memahami beberapa konsep statistik, mereka masih membutuhkan dukungan pada materi yang lebih kompleks.
Rendah	Mean - 2SD to Mean - 1SD	Mahasiswa pada kelompok ini mengalami kecemasan ringan tetapi masih dapat mengikuti materi tanpa banyak kesulitan. Mereka terkadang merasa bingung tetapi umumnya dapat mengatasi tantangan dengan bantuan minimal. Mereka memiliki pemahaman dasar tentang konsep statistik.
Sangat Rendah	< Mean - 2SD	Pada level ini, siswa tidak menunjukkan kecemasan terhadap pembelajaran statistik. Mereka percaya diri dan nyaman dengan konsep dan keterampilan analisis statistik. Para siswa memahami materi dengan baik dan berpartisipasi aktif dalam kelas.

Penelitian ini telah memvalidasi instrumen untuk mengukur kecemasan statistik pada mahasiswa psikologi menggunakan analisis model Rasch. Metode ini telah terbukti efektivitas penggunaannya untuk validasi berbagai instrumen pengukuran psikologis dalam konteks Pendidikan tinggi, antara lain *MASQUE/ Multicultural Attitude Scale Questionnaire* (Tanujaya & Yudiarso, 2023), *Self-control instruments for secondary school student* (Prayoga et al., n.d.), dan instrumen aplikasi STEM (Qudratuddarsi et al., 2022).

Validasi instrumen dengan model Rash pada penelitian ini memberi lebih banyak informasi ketimbang menggunakan metode teori tes klasik. Model Rasch banyak digunakan karena memberikan gambaran yang objektif terhadap kemampuan, sikap, dan karakter mahasiswa, serta memungkinkan analisis terperinci tentang karakteristik item dan responden (Ibrahim et al., 2025). Keunggulan analisis model Rasch dibandingkan teori uji klasik dalam validasi instrumen adalah menghasilkan skala linier dengan interval yang sama, memprediksi data yang hilang, memberikan perkiraan yang lebih akurat, mendeteksi ketidakcocokan model, dan menghasilkan pengukuran yang dapat direplikasi (Sumintono & Widhiarso, 2015). Selain itu, analisis model Rasch juga memberikan wawasan yang lebih baik tentang kualitas item dibandingkan dengan teori tes klasik untuk mengevaluasi pertanyaan ujian (Fernanda & Hidayah, 2020).

KESIMPULAN

Penelitian ini memberikan bukti empiris bahwa dimensi kecemasan statistik dalam instrumen SES-Psy valid dan reliabel untuk digunakan dalam mengukur kecemasan statistik pada mahasiswa psikologi di Indonesia. Keberadaan instrumen yang terstandar memungkinkan untuk mengidentifikasi mahasiswa yang mengalami kecemasan statistik secara lebih akurat. Hasil pengukuran ini memberikan dasar empiris dalam merancang intervensi akademik maupun psikologis untuk mengurangi hambatan belajar akibat kecemasan statistik. Pada

konteks pendidikan yang lebih luas, penggunaan instrumen ini berkontribusi meningkatkan kesetaraan akses belajar, karena upaya peningkatan lingkungan belajar yang lebih responsif pada kebutuhan emosional mahasiswa. Selanjutnya penulis merekomendasikan agar dilakukan penelitian lanjutan untuk menguji validitas instrumen ini pada mahasiswa dari disiplin ilmu sosial lain, sehingga cakupan penggunaannya menjadi lebih luas dan lintas bidang. Selain itu, penting untuk menambahkan tahap uji coba awal dan wawancara kognitif guna memastikan keterpahaman setiap item oleh responden dan meningkatkan validitas budaya dari instrumen. Penambahan item dengan cakupan indikator yang lebih variatif juga disarankan untuk meningkatkan sensitivitas dan daya diskriminasi instrumen untuk mendeteksi tingkat kecemasan statistik secara lebih mendalam.

REFERENSI

- Arslan, D., & Sak, U. (2023). Factorial Invariance of The Anadolu Sak Intelligence Scale Across Gender. *Journal of Psychoeducational Assessment*, 41(5), 542–555. <https://doi.org/10.1177/07342829231158834>
- Aziz, A. A., Zaharim, A., Fuaad, N. F. A., & Nopiah, Z. M. (2013). Students' Performance on Engineering Mathematics: Applying Rasch Measurement Model. *2013 12th International Conference on Information Technology Based Higher Education and Training (ITHET)*. <https://doi.org/10.1109/ITHET.2013.6671040>
- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for The Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine*, 25(24), 3186–3191. https://journals.lww.com/spinejournal/fulltext/2000/12150/Guidelines_for_the_Process_of_Cros_Cultural.14.aspx
- Bond, T. G., & Fox, C. M. (2007). *Applying The Rasch Model: Fundamental Measurement in The Human Sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9781410614575>
- Brentani, E., & Golia, S. (2007). Unidimensionality in The Rasch Model: How to Detect and Interpret. *Statistica*, 67(3), 253–261. <https://doi.org/10.6092/ISSN.1973-2201/3508>
- Condrón, D. J., Becker, J. H., & Bzhetaj, L. (2018). Sources of Students' Anxiety in A Multidisciplinary Social Statistics Course. *Teaching Sociology*, 46(4), 315–326. <https://doi.org/10.1177/0092055X18780501>
- Dabaghi, S., Esmailzadeh, F., & Rohani, C. (2020). Application of Rasch Analysis for Development and Psychometric Properties of Adolescents' Quality of Life Instruments: A Systematic Review. *Adolescent Health, Medicine and Therapeutics*, 11, 91–104. <https://doi.org/10.2147/ahmt.s265413>
- Dempster, M., & McCorry, N. K. (2009). The Role of Previous Experience and Attitudes Toward Statistics in Statistics Assessment Outcomes Among Undergraduate Psychology Students. *Journal of Statistics Education*, 17(2). <https://doi.org/10.1080/10691898.2009.11889515>
- Doğan, M. (2015). Influence of Sample Size, Estimation Method and Normality on Fit Indices in Confirmatory Factor Analysis. *Journal of Social Science Research*, 9(2), 75–91.
- Edwards, M. C., & Houts, C. R. (2018). A Diagnostic Procedure to Detect Departures From Local Independence in Item Response Theory Models. *Psychological Methods*, 23(1), 138–149. <https://doi.org/10.1037/met0000121>
- Fernanda, J. W., & Hidayah, N. (2020). Analisis Kualitas Soal Ujian Statistika Menggunakan Classical Test Theory dan Rasch Model. *Square: Journal of Mathematics and Mathematics Education*, 2(1), 35–42. <https://doi.org/10.21580/square.2020.2.1.5363>
- Garfield, J. (2002). The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education*, 10(3). <https://doi.org/10.1080/10691898.2002.11910676>
- Griffith, J. D., Mathna, B., Sappington, M., Turner, R., Evans, J., Gu, L., Adams, L. T., & Morin, S. (2014). The Development and Validation of The Statistics Comprehensive Anxiety Response Evaluation. *International Journal of Advances in Psychology*, 3(2), 21–29.

- 883 *Validasi Psikometrik Instrumen Kecemasan Statistik untuk Mahasiswa Psikologi di Indonesia Menggunakan Model Rasch - Siti Munawarah, Bahrul Hayat*
DOI: <https://doi.org/10.31004/edukatif.v7i4.8404>
- Hoegler, S., & Nelson, M. (2018). The Influence of Anxiety and Self-Efficacy on Statistics Performance: A Path Analysis. *Psi Chi Journal of Psychological Research*, 23(5), 364–371.
<https://doi.org/10.24839/2325-7342.jn23.5.364>
- Holster, T. A., & Lake, J. (2016). Guessing and The Rasch Model. *Language Assessment Quarterly*, 13(2), 124–141. <https://doi.org/10.1080/15434303.2016.1160096>
- Howard, G. S. (2019). The Present and Future of Methodology and Statistics in Psychology. *The Humanistic Psychologist*, 47(1), 1–12. <https://doi.org/10.1037/hum0000111>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling*, 6(1), 1–55.
<https://doi.org/10.1080/10705519909540118>
- Hunt, B. W., Mari, T., Knibb, G., Christiansen, P., & Jones, A. (2023). Statistics Anxiety and Predictions of Exam Performance in UK Psychology Students. *PLoS ONE*, 18(8), e0290467.
<https://doi.org/10.1371/journal.pone.0290467>
- Ismail, M. S., Din, M. H., & Jusoh, M. S. (2021, July). Predictive Modelling Using Rasch's Person-Item Map: Interpreting and Assessing in Malaysia Manufacturing Firms. In *AIP Conference Proceedings* (Vol. 2347, No. 1, p. 020288). AIP Publishing LLC. <https://pubs.aip.org/aip/acp/article-abstract/2347/1/020288/614840>
- Jiao, H., Lissitz, R. W., Macready, G., Wang, S., & Liang, S. (2011). Exploring Levels of Performance Using The Mixture Rasch Model for Standard Setting. *Psychological Test and Assessment Modeling*, 53(4), 499–522.
https://www.psychologie-aktuell.com/fileadmin/download/ptam/4-2011_20111217/06_Jiao.pdf
- Kabic, M., & Alexandrowicz, R. W. (2023). RMX/PIccc: An Extended Person–Item Map and A Unified IRT Output for eRm, Psychotools, ltm, mirt, and TAM. *Psych*, 5(3), 265–280.
<https://doi.org/10.3390/psych5030062>
- Kaufmann, L., Ninaus, M., Weiss, E. M., Gruber, W., & Wood, G. (2022). Self-Efficacy Matters: Influence of Students' Perceived Self-Efficacy on Statistics Anxiety. *Annals of the New York Academy of Sciences*, 1514(1), 41–52. <https://doi.org/10.1111/nyas.14797>
- Kyriazos, T. A. (2018). Applied psychometrics: The Application of CFA to Multitrait-Multimethod Matrices (CFA-MTMM). *Psychology*, 9(12), 2933–2970. <https://doi.org/10.4236/psych.2018.912150>
- Lee, H. S., & Geisinger, K. F. (2014). The Effect of Propensity Scores on DIF Analysis: Inference on The Potential Cause of DIF. *International Journal of Testing*, 14(4), 345–368.
<https://doi.org/10.1080/15305058.2014.922567>
- Linacre, J. M. (2002). Optimizing Rating Scale Category Effectiveness. *Journal of Applied Measurement*, 3(1), 85–106.
- Linacre, J. M. (2012). *A user's guide to Winsteps/Ministep: Rasch-Model Computer Programs*. Winsteps.
- Ibrahim, N. L., Ekonomi da Bisnis, F., & Gorontalo, P. (2025). Analisis Literatur Penggunaan Model Rasch untuk Validasi Instrumen Asesmen Pendidikan: Tren, Temuan, dan Implikasi. *Lencana: Jurnal Inovasi Ilmu Pendidikan*, 3(2), 46–60. <https://doi.org/10.55606/lencana.v3i2.5012>
- Liu, L. (2021). Assessing Statistics Anxiety in An Online or Hybrid Setting: The Adaption and Development of a New Instrument SASOH. *Statistics Education Research Journal*, 20(1).
<https://doi.org/10.52041/SERJ.V20I1.96>
- MacCann, R. (2009). Standard Setting with Dichotomous and Constructed Response Items: Some Rasch Model Approaches. *Journal of Applied Measurement*, 10(4), 438–454.
<https://www.kingarthursvoyagetoamerica.com/wp-content/uploads/2016/08/01-MacCann-Oxford-University-Journal-of-Applied-Measurement-10-4.pdf>

- 884 *Validasi Psikometrik Instrumen Kecemasan Statistik untuk Mahasiswa Psikologi di Indonesia Menggunakan Model Rasch - Siti Munawarah, Bahrul Hayat*
DOI: <https://doi.org/10.31004/edukatif.v7i4.8404>
- Macher, D., Papousek, I., Ruggeri, K., & Paechter, M. (2015). Statistics Anxiety and Performance: Blessings in Disguise. *Frontiers in Psychology*, 6, 1116. <https://doi.org/10.3389/fpsyg.2015.01116>
- Mcintee, S. E., Goulet-Pelletier, J. C., Williot, A., Deck-L  ger, E., Lalande, D., Cantinotti, M., & Cousineau, D. (2022). (Mal)adaptive Cognitions as Predictors of Statistics Anxiety. *Statistics Education Research Journal*, 21(1). <https://doi.org/10.52041/serj.v21i1.364>
- Ogbonnaya, K. E., Okechi, B. C., & Nwankwo, B. C. (2019). Impact of Statistics Anxiety and Self-Efficacy on Statistical Performance of Psychology Students. *Nigerian Journal of Social Psychology*, 2(2), 15–23. <https://www.nigerianjps.com/index.php/NJSP/article/view/33>
- Padgett, R. N., & Morgan, G. B. (2019). An Examination of Sensitivity to Measurement Error in Rasch Residual-Based Fit Statistics. *Journal of Applied Measurement*, 20(4), 358–371.
- Prayoga, K., Suryana, D., & Suhendar, M. S. (2024). Penggunaan Rasch model untuk Menganalisis Konstruk Instrumen Kontrol Diri pada Siswa Sekolah Menengah. *Guidance: Jurnal Bimbingan dan Konseling*, 7(1), 35–44. <https://journal.upy.ac.id/index.php/bk/article/view/4459>
- Qudratuddarsi, H., Hidayat, R., Shah, R. L. Z. R. M., Nasir, N., Imami, M. K. W., & Mat Nor, R. (2022). Rasch Validation of Instrument Measuring Gen-Z Science, Technology, Engineering, and Mathematics (STEM) Application in Teaching During The Pandemic. *International Journal of Learning, Teaching and Educational Research*, 21(6), 104–121
- Smith, A. B., Cocks, K., Parry, D., & Taylor, M. (2016). A Differential Item Functioning Analysis of the EQ-5D in cancer. *Value in Health*, 19(8), 1063–1067. <https://doi.org/10.1016/j.jval.2016.06.005>
- Smith, E. V. (2002). Detecting and Evaluating The Impact of Multidimensionality Using Item Fit Statistics and Principal Component Analysis of Residuals. *Journal of Applied Measurement*, 3(2), 205–231. <https://europepmc.org/article/med/12011501>
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Trim Komunikata.
- Tanujaya, L., & Yudiarso, A. (2023). Uji Validitas dan Reliabilitas Multicultural Attitude Scale Questionnaire (MASQUE) Versi Indonesia dengan Model Rasch. *Humanitas (Jurnal Psikologi)*, 7(2), 161–174. <https://doi.org/10.28932/humanitas.v7i2.7420>
- Tun  , E. B. (2023). A Review of Measurement Tools Developed and Adapted Based on the Rasch Model. *Journal of Human and Social Sciences*, 6(2), 249–275.
- Umar, J., & Nisa, Y. F. (2020). Uji Validitas Konstruk dengan CFA dan Pelaporannya. *Jurnal Pengukuran Psikologi dan Pendidikan Indonesia*, 9(2), 1–11. <https://doi.org/10.15408/jp3i.v9i2.16964>
- Vigil-Colet, A., Lorenzo-Seva, U., & Condon, L. (2008). Development and Validation of The Statistical Anxiety Scale. *Psicothema*, 20(1), 174–180.
- 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. <https://doi.org/10.1177/014662168400800201>
- Yu, C. H. (2020). Objective Measurement: How Rasch Modeling Can Simplify and Enhance Your Assessment. *Rasch measurement: Applications in quantitative educational research*, 47–7