PSYCHE: Jurnal Psikologi Universitas Muhammadiyah Lampung
Vol. 7 No.2, Agustus 2025
ISSN (*electronic*) 2655-6936
ISSN (*printed*) 2686-0430

VALIDASI STRUKTUR INTERNAL DAN LATEN *SUB-CLASS ACADEMIC SELF-REGULATED LEARNING* VERSI INDONESIA DENGAN *RASCH MIXTURE MODEL*

VALIDATION OF INTERNAL STRUCTURE AND LATENT SUB-CLASSES OF THE INDONESIAN VERSION OF ACADEMIC SELF-REGULATED LEARNING SCALE USING THE MIXTURE RASCH MODEL

Sandra Arviyenna, Ni Putu Rahayu Eka Putri, Ananta Yudiarso
Universitas Surabaya
sandraubaya@gmail.com, rahayuekaputri10@gmail.com, ananta@staff.ubaya.ac.id*

ABSTRAK

Penelitian ini bertujuan untuk menguji validitas menggunakan model *Rasch* dan *Rasch Mixture* pada skala *Academic Self-Regulated Learning Scale* (A-SRL-S) (Magno, 2010). Metode penelitian menggunakan survei dengan *non-random sampling* melibatkan 401 responden. Hasil penelitian menunjukkan unidimensionalitas pada skala A-SRL terpenuhi pada semua sub-skala setelah menggugurkan MS4, MS5, SA32, SA37, O53, dan O54, sehingga menjadi 48 item. Reliabilitas item pada semua sub-skala menunjukkan hasil yang sangat baik, sedangkan *person reliability* dan *person separated index* tidak reliabel dengan rentang 0.46-0.79 dan PSI 0.92-1.94. Item *misfit* pada LR46, tidak sesuai dengan model Rasch. Responden mampu membedakan skala dari STS sampai SS. Pada analisis DIF, item LR44 dan LR46 menunjukkan bias *gender* dengan probabilitas Welch 0.0065 (LR44) dan 0.0037 (LR46). *Wright map* menunjukkan tingkat kesulitan item yang kurang mampu menjangkau responden dengan kemampuan tinggi. Pada analisis *Rasch Mixture Model*, sub-skala *learning responsibility* mendapati laten *sub-class* terdiri dari dua kelas. Implikasi temuan menunjukkan perlunya item *misfit* dan item multidimensi, *renorming* berdasarkan *latent class,* serta replikasi penelitian dengan partisipan yang lebih heterogen guna meningkatkan sensitivitas skala.

Kata Kunci*:* A-SRL-S, analisis *rasch*, analisis *rasch mixture*

ABSTRACT

This study aims to examine the validity using the Rasch model and Rasch Mixture Model on the Academic Self-Regulated Learning Scale (A-SRL-S) by Magno (2010). The research method employed a survey with non-random sampling with 401 respondents. The results indicate that the assumption of unidimensionality for the A-SRL scale is met for all subscales after the removal of items MS4, MS5, SA32, SA37, O53, and O54, leaving 48 items. Item reliability for all subscales showed excellent results, while person reliability and person separation index were not reliable, ranging from 0.46 to 0.79, with a PSI of 0.92 to 1.94. Misfit item LR46 *do not fit the Rasch Model. Respondents were able to distinguish the scale from strongly disagreed to strongly agreed. The DIF analysis revealed that items LR44 and LR46 exhibited gender bias, with Welch probabilities of 0.0065 (LR44) and 0.0037 (LR46). The Wright map showed that the item difficulty levels did not adequately reach respondents with high ability. In the Rasch Mixture Model analysis, the learning responsibility subscale exhibited a latent subclass consisting of two classes. The findings imply the need to revise misfit and multidimensional items, conduct renorming based on latent class, and replicate the study with a more heterogeneous sample to enhance the scale's sensitivity.*

Keywords*: A-SRL-S, rasch analysis, rasch mixture analysis*

https://doi.org/10.36269/psyche.v7i2.3049

PENDAHULUAN

Self-regulated learning (SRL) telah muncul sebagai konsep penting dalam psikologi pendidikan, mencerminkan proses di mana peserta didik mengambil kendali atas pengalaman belajar mereka sendiri. Konsep A-*SRL* mencakup berbagai aspek, yaitu penetapan tujuan *(goal setting),* tanggung jawab belajar *(learning responsibility)*, evaluasi diri *(self evaluation)*, mencari bantuan *(seeking assistance)*, strategi memori *(memory strategy),* pengaturan lingkungan *(environmental structuring)*, dan pengorganisasian *(orgnanizing)* (Zimmerman & Martinez-Pons, 1986). Dimensi-dimensi ini sangat penting untuk mendorong pembelajaran mandiri dan keberhasilan akademik, karena memungkinkan peserta didik untuk mengelola strategi belajar mereka secara efektif. Validasi skala pengukuran yang menilai *SRL* sangatlah penting untuk penelitian maupun aplikasi praktis dalam lingkungan pendidikan.

Magno (2010) mengembangkan skala untuk mengukur *SRL*, yang didasarkan pada karya dasar Zimmerman dan Martinez-Pons (1986). Skala ini bertujuan untuk menyediakan alat yang andal bagi pendidik dan peneliti dalam menilai keterampilan pengaturan diri siswa. Penelitian yang dilakukan oleh Magno (2010) mengenai pengukuran *self-regulated learning* (SRL) menggunakan skala A-SRL telah diterapkan pada berbagai populasi dan konteks pendidikan. Magno (2010) menganalisis struktur faktor skala ini menggunakan *explanatory factor analysis* (EFA) dan dikonfirmasi dengan *confirmatory factor analysis* (CFA). Selain itu model *Polychotomous Rasch (Partial Credit Model)* digunakan untuk menguji apakah kategori dalam skala sudah sesuai dan apakah setiap item dalam skala berfungsi dengan baik.

Analisis mengunakan *Item Response Theory* (IRT) menunjukkan bahwa skala ini memiliki urutan kategori yang logis dan sesuai dengan harapan, dengan kata lain *step function* meningkat secara pasti. Semakin tinggi skor responden, semakin besar peluang mereka memilih kategori yang lebih tinggi dalam skala. Namun, ditemukan 4 dari 55 item yang tidak konsisten dengan item lainnya, sehingga perlu dievaluasi lebih lanjut. Hasil penelitian menunjukkan *item reliability* yang tinggi untuk seluruh sub-skala (>0.8).

Pada penelitian selanjutnya (Magno, 2011) menguji validitas konstruk dari A-SRL-S yang terdiri dari 7 sub-skala dan 54 item, dengan membandingkannya terhadap *Motivated Strategies for Learning Quistionnaire* (MLSQ) dan *Learning and Study Strategies Inventory* (LASSI). Studi ini dilakukan pada 755 mahasiswa di Filipina menggunakan

https://doi.org/10.36269/psyche.v7i2.3049

Confirmatory Factor Analysis (CFA). Hasil menunjukkan bahwa model tiga faktor, di mana A-SRL-S, MSLQ, dan LASSI dipisahkan sebagai tiga faktor yang berkorelasi, menunjukkan kecocokan terbaik dibandingkan model lainnya (X^2=473.47, df = 87, AIC = .71).

Meskipun demikian, untuk meningkatkan relevansi dan akurasi pengukuran dalam konteks budaya Indonesia, diperlukan penelitian lanjutan yang lebih mendalam. Van de Vijver dan Leung (1997) menunjukkan pentingnya menyesuaikan instrumen pengukuran untuk mengakomodasi perbedaan budaya, untuk memastikan bahwa instrumen yang digunakan dapat mengukur dengan tepat sesuai dengan makna dalam konteks budaya yang spesifik. Oleh karena itu, validasi lebih lanjut di Indonesia sangat diperlukan. Dengan demikian, penelitian lanjutan dapat memberikan bukti empirik yang lebih kuat untuk mendukung penggunaan skala ini dalam populasi Indonesia, yang pada akhirnya dapat meningkatkan akurasi dan keandalan pengukuran dalam penelitian psikologi di Indonesia. Satu dari beberapa pendekatan yang dapat digunakan dalam proses validasi ini ialah *Rash Mixture Model* (RMM) unidimensional untuk identifikasi perbedaan dalam respons partisipan berdasarkan kelompok laten yang memiliki pola pemahaman atau interpretasi berbeda terhadap item dalam skala (Rost, 1990).

Penelitian ini bertujuan untuk mengukur validitas dan reliabilitas A-*SRL* versi Indonesia yang telah diterjemahkan oleh Andiani (2017). Validasi instrumen pengukuran dalam penelitian psikologi dan pendidikan memperhatikan *American Educational Research Association* (AERA), *American Psychological Association* (APA), dan *National Council on Measurement* (NCME). Standar ini menekankan pentingnya menilai kualitas praktik pengujian, dengan validitas sebagai bukti dan teori yang mendukung interpretasi hasil tes (Sireci & Faulkner-Bond, 2014). Validitas dapat dikaji dari berbagai aspek, termasuk isi item, proses respons, struktur internal, hubungan dengan variabel lain, serta dampak pengukuran.

Penelitian ini menggunakan Rasch unidimensional model, yang mengasumsikan bahwa hanya satu faktor utama yang memengaruhi respon individu terhadap suatu instrumen. Keunggulannya terletak pada kemampuannya menghasilkan skala pengukuran yang konsisten dan memungkinkan perbandingan langsung antarindividu meskipun menggunakan soal yang berbeda (Wright & Linacre, 1989). Selain itu, *Rasch Mixture Model* (RMM) dapat digunakan untuk mengidentifikasi subkelompok laten dalam populasi yang memiliki pola respons berbeda meskipun menggunakan instrumen yang

https://doi.org/10.36269/psyche.v7i2.3049

sama. Oleh karena itu, kombinasi Model Rasch dan RMM diharapkan dapat meningkatkan ketepatan pengukuran dan memastikan instrumen dapat digunakan secara lebih akurat dalam berbagai konteks budaya.

METODE PENELITIAN

Desain Penelitian

Validasi psikometri ini dilakukan dengan pendekatan kuantitatif melalui *Rasch Model* dan *Rasch Mixture Model* pada skala *Academic Self-Regulated Learning* versi Indonesia. Penelitian ini disusun untuk mengevaluasi konsistensi dan kemampuan skala dalam mengukur kemampuan *self-regulated learning* secara akurat dan adil, memastikan bahwa skala ini mengukur satu aspek utama, serta mengidentifikasi apakah terdapat kelompok-kelompok peserta dengan pola jawaban yang berbeda.

Partisipan

Ukuran sampel dalam analisis Rasch dipengaruhi oleh prinsip kalibrasi instrumen. Ketika suatu instrumen dikalibrasi pada sampel yang berbeda dari peserta yang sama, perbedaan hasil kecil bisa saja terjadi (Linacre, 1994). Jika ukuran sampel terlalu kecil, hasil kalibrasi menjadi tidak stabil dan kurang mampu mencerminkan kondisi yang sebenarnya (Wright & Stone, 1979). Sebaliknya, ukuran sampel yang besar dapat mengurangi perbedaan dalam hasil kalibrasi, tetapi memerlukan waktu dan biaya yang lebih besar (Linacre, 1994). Menurut Linacre (1994), untuk mendapatkan hasil analisis Rasch yang reliabel dengan tingkat kepercayaan 99% diperlukan ukuran sampel antara 108 hingga 243 responden. *Sample size kalkulator* digunakan pada pengambilan sampel pada penelitian ini, yang terdapat dalam Raosoft.com. Populasi diasumsikan sebanyak 20.000 dengan taraf kepercayaan 95% dan margin of error 5%, didapatkan sampel subjek minimal 377 orang.

Penelitian ini melibatkan 401 responden yang terdiri dari 111 laki-laki dan 290 perempuan. Seluruh responden adalah mahasiswa aktif semester 5 dan 7 dipilih dengan teknik *non-random sampling* dengan subjek dipilih berdasarkan karakteristik tertentu tanpa peluang yang sama untuk seluruh anggota populasi (Fraenkel et al., 2011). Teknik ini menggunakan *convenience sampling*, yaitu pemilihan responden yang mudah dijangkau dan bersedia berpartisipasi dalam penelitian (Etikan, Musa, & Alkassim, 2016).

https://doi.org/10.36269/psyche.v7i2.3049

Instrumen

Peneliti menggunakan instrumen *Academic Self-Regulated of Learning Scale (A-SRL-S)* versi Indonesia yang berjumlah 54 item. Aspek yang diukur mencakup penetapan tujuan *(goal setting),* tanggung jawab belajar *(learning responsibility)*, evaluasi diri *(self evaluation)*, mencari bantuan *(seeking assistance)*, strategi memori *(memory strategy),* pengaturan lingkungan *(environmental structuring)*, dan pengorganisasian *(orgnanizing).* Setiap pernyataan dalam skala ini diukur menggunakan skala Likert dengan empat opsi jawaban, yaitu 1 untuk sangat tidak setuju, 2 untuk tidak setuju, 2 untuk setuju, dan 4 untuk sangat setuju.

Tabel 1. *Blueprint Academic Self-Regulated of Learning Scale (A-SRL-S)*

Aspek	Item Coding	Item
Memory Strategy	MS1	Saya menggunakan catatan kecil untuk menuliskan informasi yang perlu saya ingat
	MS2	Saya membuat daftar informasi berdasarkan kategori
	MS3	Saya menulis ulang catatan kuliah saya dengan kata-kata saya sendiri
	MS4	Saya menggunakan gambar, skema, dan bagan untuk memahami informasi yang tidak jelas (abstrak)
	MS5	Saya menggunakan simbol (gambar, skema, bagan, dll) agar saya mudah mengingatnya
	MS6	Saya membuat rangkuman bacaan saya
	MS7	Saya membuat rangkuman sebagai panduan saya belajar
	MS8	Saya membuat rangkuman semua topik yang akan dijelaskan di kelas
	MS9	Saya membayangkan suatu kata untuk mengingat sesuatu
	MS10	Saya membaca jawaban pertanyaan terkait topik tertentu
	MS11	Saya mencatat materi perkuliahan yang saya ikuti
	MS12	Saya menjawab contoh pertanyaan yang saya buat dari topik tertentu
	MS13	Saya membaca catatan saya saat belajar untuk ujian
	MS14	Saya menulis catatan untuk mengingatkan saya untuk mengerjakan tugas
Goal Setting	GS15	Saya membuat jadwal aktivitas saya dengan detail
	GS16	Saya membuat daftar aktivitas yang harus diselesaikan
	GS17	Saya merencanakan apa yang harus saya lakukan dalam 1 minggu
	GS18	Saya menggunakan buku catatan untuk mengetahui apa yang harus saya kerjakan
	GS19	Saya menggunakan kalender untuk mengetahui apa yang harus saya kerjakan
Self-Evaluation	SE20	Jika saya mengalami kesulitan dalam belajar, saya meminta bantuan dari orang yang lebih pintar
	SE21	Saya menerima masukan dari teman terhadap hasil kerja saya
	SE22	Saya mengevaluasi kesuksesan saya setiap akhir belajar
	SE23	Saya meminta teman memberi komentar mengenai hasil tugas saya sebelum saya menyerahkannya ke dosen

https://doi.org/10.36269/psyche.v7i2.3049

	SE24	Saya mencatat perkembangan kemajuan yang saya alami
	SE25	Saya memeriksa kemajuan saya dalam mengerjakan sesuatu
	SE26	Saya menanyakan pendapat orang lain yang lebih pintar mengenai hasil kerja saya
	SE27	Saya mendengarkan orang yang mengomentari tugas saya
	SE28	Saya terbuka dengan masukan orang lain untuk mengembangkan tugas saya menjadi lebih baik
	SE29	Saya melihat kembali nilai-nilai tugas, ujian dsb sebelumnya untuk melihat perkembangan saya
	SE30	Saya bertanya pada orang lain apa yang harus saya perbaiki dalam tugas saya
	SE31	Saya mau berubah berdasarkan saran atau masukan dari orang lain
Seeking Assistance	SA32	Saya menggunakan beberapa sumber yang berbeda (buku, jurnal, penelitian orang lain, dsb) dalam membuat laporan atau makalah saya
	SA33	Saya menggunakan perpustakaan untuk mencari informasi yang saya butuhkan
	SA34	Saya menulis catatan perkuliahan saya di kelas
	SA35	Saya suka bekerja sama dengan teman karena kami saling membantu
	SA36	Ketika saya tidak masuk kuliah, saya bertanya pada teman mengenai tugas yang diberikan dosen pada hari itu
	SA37	Saya mencari teman yang bisa saling berdiskusi
	SA38	Saya belajar dengan teman untuk membandingkan catatan kuliah
	SA39	Saya menjelaskan pada teman apa yang telah saya pelajari dari topik tertentu
Environmental Structuring	ES40	Saya tidak menonton TV (termasuk video, *youtube*, dsb) saat ada tugas kuliah
	ES41	Saya menghindari tempat yang mengganggu saya belajar
	ES42	Saya tidak bisa belajar atau mengerjakan tugas bila ruangannya tidak terang
	ES43	Saya mematikan TV (termasuk video, *youtube*, dsb) agar bisa berkonsentrasi belajar
Learning Responsibility	LR44	Saya memeriksa ulang tugas saya untuk memastikan semuanya benar sebelum mengumpulkannya
	LR45	Saya langsung mengerjakan tugas yang diberikan dosen
	LR46	Saya cemas dengan batas waktu pengumpulan tugas yang ditetapkan dosen
	LR47	Saya mendahulukan tugas perkuliahan saya daripada aktivitas lainnya
	LR48	Saya menyelesaikan tugas saya sebelum mengerjakan yang lain
Organizing	O49	Saya memberikan stabilo atau menggarisbawahi kata atau informasi penting dalam bacaan saya
	O50	Saya membayangkan bentuk ujian yang akan datang berdasarkan ujian sebelumnya
	O51	Saya menyimpan catatan dan catatan kuliah lama saya di tempat khusus
	O52	Saya belajar sesuai kemampuan maksimal yang saya mampu
	O53	Saya merapikan barang-barang di sekitar tempat belajar sebelum mulai belajar
	O54	Saya memastikan tempat belajar saya bersih sebelum mulai belajar

https://doi.org/10.36269/psyche.v7i2.3049

Analisis Data

Data yang telah dikumpulkan dianalisis menggunakan software Winsteps® Rasch Measurement versi 5.0 untuk menganalisis *Rasch Model* dan Jamovi untuk menganalisis *Rasch Mixture Model* (RMM)**.** Analisis data dilakukan dengan menggunakan *Rasch Model* untuk menghasilkan informasi holistik yang mencakup uji unidimensionalitas, *rating scale*, *reliability, item fit, wright map,* dan *differential item functioning*. Statistik *fit* diterapkan untuk menilai sejauh mana data mendukung model Rasch, berdasarkan nilai infit MNSQ dan outfit MNSQ di mana idealnya berada pada rentang 0,5 hingga 1,5. Model Rasch memungkinkan kemampuan individu dan tingkat kesulitan item untuk dinyatakan dalam suatu skala logaritmik yang sama, yaitu skala *logit*, sehingga mendukung pengukuran yang objektif (Bond & Fox, 2015). Model Rasch mampu memberikan estimasi yang lebih stabil pada data tes, mampu mengidentifikasi item-item yang tidak berfungsi sebagaimana mestinya *(misfitting items),* serta memastikan *unidimensionality,* yaitu bahwa tes hanya mengukur satu konstruk utama.

Jamovi versi 2.3.28 digunakan untuk menganalisis *Rasch Mixture Model* (RMM), yang membantu mengidentifikasi subkelompok tersembunyi dalam data dan memberikan pemahaman lebih mendalam tentang variasi respons antar kelompok (The Jamovi Project, 2024). Penggunaan kedua *software* ini memastikan analisis yang lebih komprehensif dan akurat dalam memvalidasi skala A-SRL versi Indonesia.

HASIL PENELITIAN

Unidimensionalitas

Analisis dalam penelitian ini dilakukan secara terpisah untuk setiap aspek dalam A-SRL-S. Setiap sub skala diuji unidimensionalitasnya dengan mengacu pada varians mentah yang dijelaskan oleh pengukuran, menggunakan kriteria yang ditetapkan oleh *Rasch Principal Component Analysis of Residuals* (PCAR). Linacre (2002) menyatakan bahwa varians yang dijelaskan sebesar 40% dapat diterima sebagai indikator unidimensionalitas, dengan syarat nilai eigenvalue dari *first contrast* tidak melebihi 2.0. Nilai *eigenvalue* yang melebihi 2.0 mengindikasikan adanya kemungkinan dimensi tambahan yang dapat memengaruhi hasil pengukuran.

Pada *raw variance explained by measure* dari 54 item A-SRL-S, diperoleh nilai sub-skala GS, SE, ES, dan O yang berada >40%. Namun, sub-skala MS (35.9%), SA (39.7%) dan O (38.9%) masih kurang memenuhi ambang batas minimal, sehingga kontribusinya

https://doi.org/10.36269/psyche.v7i2.3049

terhadap dimensi utama perlu diperhatikan lebih lanjut. Peneliti melakukan penghapusan item MS4, MS5, SA32, SA37, O53, dan O54 dengan melihat *residual loadings for item* yang bernilai >0.5. Setelah dilakukan pengguguran item, pada Tabel 2 menunjukkan bahwa semua sub-skala unidimensional dengan 48 item. Selanjutnya, akan dilakukan analisis lanjutan menggunakan 48 item.

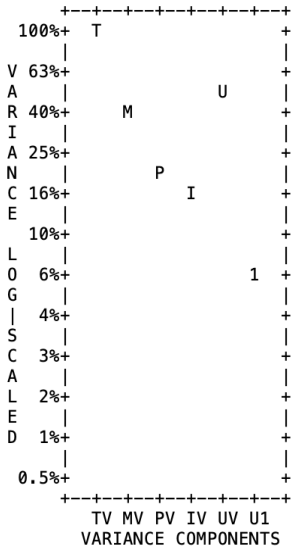
Tabel 2. *Unidimensionalitas*

Sub-skala	***Raw Variance Explained by Measure* (>40%)**	***First Contrast* (<2.0)**	***Item Deleted***	**Kesimpulan**
Memory Strategy (MS)	40.9	1.6	MS4, MS5	Unidimensional
Goal Setting (GS)	59	1.3	-	Unidimensional
Self Evaluation (SE)	42.4	2.1	-	Unidimensional
Seeking Assistance (SA)	44	1.4	SA32, SA37	Unidimensional
Environmental Structuring (ES)	52.8	1.5	-	Unidimensional
Learning Responsibility (LS)	43.2	1.6	-	Unidimensional
Organizing (O)	41.9	1.4	O53, O54	Unidimensional

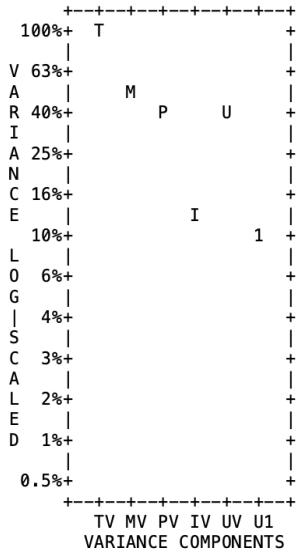
Berdasarkan *scree plo*t dari dimensi A-SRL, *Memory Strategy* dan *Goal Setting* menunjukkan pola unidimensional, dengan satu komponen utama yang dominan dan penurunan tajam setelahnya. Hal ini mengindikasikan bahwa item-item dalam kedua dimensi tersebut cenderung mengukur satu konstruk inti. Sementara itu, dimensi *Self Evaluation*, *Seeking Assistance*, *Environmental Structuring*, dan *Learning Responsibility* menunjukkan kecenderungan multidimensional, karena memiliki lebih dari satu komponen dengan *eigenvalue* yang masih signifikan. Pola ini mengisyaratkan bahwa setiap dimensi tersebut memuat beberapa aspek berbeda yang turut berkontribusi dalam pembentukan konstruk utamanya.

https://doi.org/10.36269/psyche.v7i2.3049

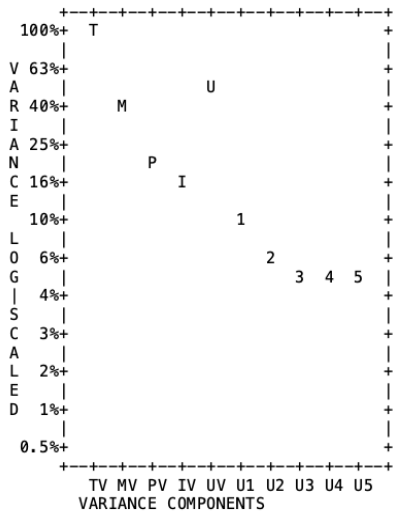
PSYCHE: Jurnal Psikologi Universitas Muhammadiyah Lampung
Vol. 7 No.2, Agustus 2025
ISSN (*electronic*) 2655-6936
ISSN (*printed*) 2686-0430



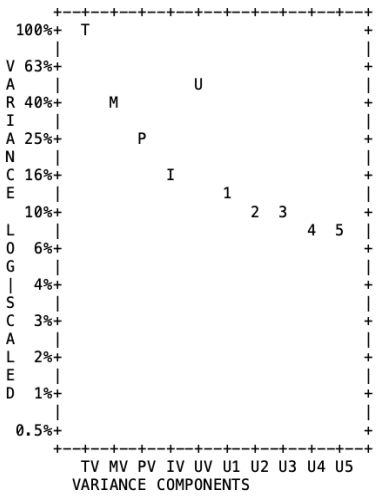
Memory Strategy



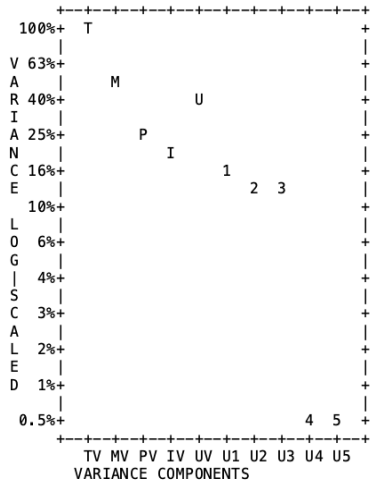
Goal Setting



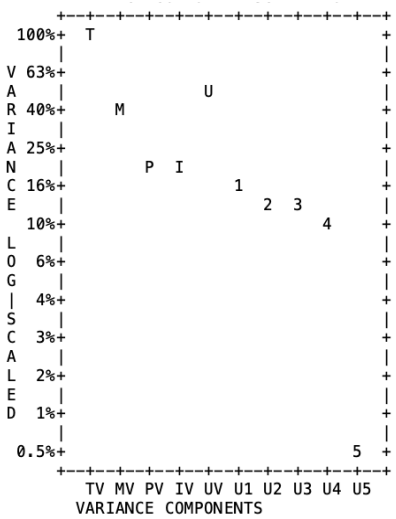
Self Evaluation



Seeking Assistance

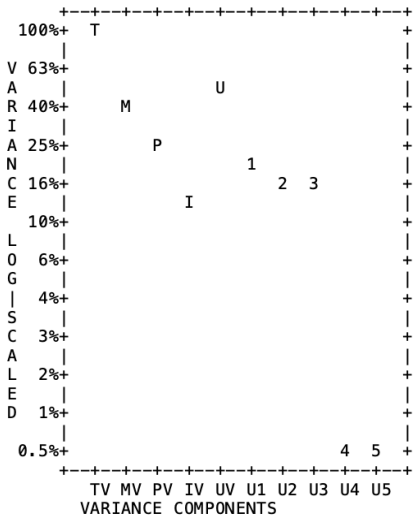


Environmental Structuring



Learning Responsibility

https://doi.org/10.36269/psyche.v7i2.3049



Gambar 1. *Scree Plot* A-SRL

Rating Scale Diagnostic

Skala penilaian yang digunakan dalam A-SRL-S adalah skala *likert* dengan empat poin yang terdiri dari "sangat tidak setuju" (1), "tidak setuju" (2), "setuju" (3), dan "sangat setuju" (4). Diagnostik skala penilaian dilakukan untuk menilai apakah responden mampu membedakan pilihan jawaban yang disediakan. Hasil analisis ini memberikan informasi yang lebih akurat dan mudah dipahami terkait konstruk yang diukur, karena peneliti dapat mengidentifikasi jarak aktual yang digunakan oleh responden saat menentukan pilihan. *Andrich Threshold* adalah nilai ambang batas dalam model Rasch yang menunjukkan titik transisi antara satu kategori respons ke kategori berikutnya dalam skala pengukuran (Andrich, 1978).

Berdasarkan Tabel 3, nilai *Andrich threshold* semua sub-skala menunjukkan ambang batas yang meningkat secara bertahap dari negatif ke positif dari empat opsi respons (Linacre, 2012). Hal ini menunjukkan bahwa responden mampu membedakan satu jawaban dengan jawaban lainnya.

https://doi.org/10.36269/psyche.v7i2.3049

Tabel 3. *Rating Scale Diagnostics* A-SRL-S

Sub-scale	***Rating Scale***	**Average Measure**	***Infit* MNSQ**	***Outfit* MNSQ**	**Andrich Threshold**
Memory Strategy	STS	-.84	1.03	1.10	NONE
	TS	.12	.82	.77	-1.30
	S	1.44	.99	1.02	-.68
	ST	2.28	1.07	1.05	1.97
Goal Setting	STS	-2.24	.96	.99	NONE
	TS	-.55	.79	.78	-2.17
	S	1.57	1.00	1.05	-.45
	ST	2.75	1.13	1.08	2.62
Self Evaluation	STS	-.93	1.07	1.19	NONE
	TS	.04	.79	.77	-1.48
	S	1.53	.96	1.02	-.68
	ST	2.44	1.07	1.05	2.16
Seeking Assistance	STS	-.85	1.01	1.07	NONE
	TS	-.04	.77	.72	-1.19
	S	1.39	.98	1.09	-.80
	ST	2.34	1.07	1.04	1.99
Environmental Structuring	STS	-1.07	.99	1.04	NONE
	TS	-.37	.58	.55	-1.01
	S	1.25	.90	1.18	-.60
	ST	1.95	1.13	1.10	1.61
Learning Responsibility	STS	-.45	1.26	1.35	NONE
	TS	-.12	.70	.66	-.155
	S	1.15	.91	1.04	-.34
	ST	2.17	1.08	1.08	1.89
Organizing	STS	-1.12	1.05	1.09	NONE
	TS	-.01	.83	.80	-1.09
	S	1.56	.98	1.01	-1.00
	ST	2.54	1.05	1.02	2.09

Reliabilitas

Model Rasch digunakan untuk menilai reliabilitas baik pada responden maupun pada item. Kemampuan instrumen dalam membedakan responden berdasarkan variabel yang diukur disebut *person reliability*. Nunnally & Bernstein (1994) menyatakan nilai reliabilitas minimal 0.8 disarankan terutama untuk keperluan evaluatif serta diagnostik, karena memberikan dasar yang lebih kuat dalam pengambilan keputusan. Nilai *Cronbach's α* di atas 0.8 menunjukkan konsistensi internal yang tinggi, sehingga mengurangi kesalahan pengukuran dan meningkatkan validitas data (Gliem & Gliem, 2003). Standar ini sangat penting, khususnya dalam penelitian di bidang psikologi klinis atau pendidikan, di mana hasil yang akurat sangat menentukan interpretasi data (Streiner, 2003). Meskipun nilai 0.7 masih dapat diterima dalam penelitian eksploratif

https://doi.org/10.36269/psyche.v7i2.3049

(Hair et al., 2019), namun reliabilitas 0.8 dianggap lebih ideal karena menawarkan tingkat kepercayaan yang lebih tinggi terhadap stabilitas alat ukur. Oleh karena ini, dengan mempertahankan reliabilitas pada angka 0.8 berkontribusi pada peningkatan validitas, kredibilitas temuan, dan juga mengurangi potensi bias dalam analisis.

Tabel 4 menunjukkan nilai *Cronbach's α* untuk semua sub-skala A-SRL berada di atas 0.8, yang menunjukkan konsistensi internal yang tinggi dalam mengukur konstruk yang dimaksud. Menurut Nunnally & Bernstein (1994), nilai reliabilitas sebesar >0.8 dianggap memadai untuk instrumen yang digunakan dalam konteks pengukuran terapan. Namun, reliabilitas person pada seluruh sub-skala tidak reliabel. Reliabilitas person yang rendah menunjukkan bahwa perbedaan kemampuan individu tidak dapat diestimasi secara akurat. Selain itu, *item separation index* menunjukkan bahwa seluruh sub-skala memiliki nilai di atas 2.0, yang berarti instrumen memiliki kemampuan yang memadai untuk membedakan secara statistik antara peserta dengan performa tinggi dan rendah (Linacre, 2012). *Person separation index* seluruh sub-skala tidak memenuhi kriteria ideal (2.0) yang berarti bahwa instrumen belum cukup andal untuk membedakan individu berdasarkan tingkat kemampuannya secara bermakna (Linacre, 2012). Sedangkan strata pada seluruh sub-skala berada di bawah 3.0 yang menunjukkan bahwa instrumen belum mampu membedakan secara jelas tiga kelompok performa (rendah, sedang, tinggi) secara statistik.

Tabel 4. *Reliability* Skala A-SRL

	Sub-skala 1:MS	**Sub-skala 2: GS**	**Sub-skala 3: SE**	**Sub-skala 4: SA**	**Sub-skala 5:ES**	**Sub-skala 6: LR**	**Sub-skala 7: O**
N	12 item	5 item	12 item	6 item	4 item	5 item	4 item
Person Reliability (>.8)	.79	.79	.79	.62	.53	.55	.46
Alpha Cronbach (>.8)	.94	.88	.96	.97	.99	.98	.89
Item Separation Index (>2.0)	3.85	2.71	5.23	5.26	8.18	7.25	2.82
Person Separation Index (>2.0)	1.94	1.92	1.95	1.27	1.06	1.10	.92
Strata (>3.0)	2.92	2.89	2.93	2.03	1.75	1.8	1.56

Item Misfit

Item misfit dalam *Rasch Model* mengacu pada sejauh mana respon responden terhadap item, sesuai dengan model *Rasch* yang diharapkan secara teoretis. Evaluasi *item fit* penting untuk menentukan apakah suatu item dalam instrument pengukuran

https://doi.org/10.36269/psyche.v7i2.3049

berfungsi dengan baik dalam mengukur konstruk yang sama secara konsisten. *Rasch Model* menganalisis efektivitas setiap item dengan memperhatikan nilai *Infit* MNSQ, *Outfit* MNSQ, dan korelasi *point-measure*. Menurut Linacre, nilai MNSQ ideal untuk *infit* dan *outfit* berada di antara 0.5 sampai 1.5. Nilai ZSTD (*standardized z-score*) ideal berada di rentang ±2.0 untuk menunjukkan kecocokan item yang baik (Sumintono & Widhiarso, 2013). Item dengan nilai di luar rentang tersebut menunjukkan anomali yang memerlukan perhatian lebih. Sedangkan korelasi *point-measure* yang mengacu pada hubungan antara kemampuan responden dan tingkat kesulitan item, dikatakan baik apabila hasilnya berada pada rentang 0.4 hingga 0.85 (Sumintono, 2016).

Tabel 5 menunjukkan bahwa item SE 20, SA33, ES42, dan LR 46 tergolong misfit karena nilai infit dan outfit melebihi batas 2.0. Nilai ZSTD yang tinggi menunjukkan kontribusi item yang kurang sesuai terhadap model (Bond & Fox, 2015). Sedangkan item SE22, SE27, SE30, SA38, SA39, ES40, ES43, LR45, LR47, dan LR48 dikategorikan overfit karena nilai ZTD berada di bawah -2.0. Hal ini menunjukkan adanya inkonsistensi dalam pola respons yang dapat mengindikasikan item tidak berfungsi dengan baik dalam mengukur konstruk yang dimaksud (Linacre, 2012). LR46 memiliki Infit dan Outfit MNSQ sebesar 1.63 dengan ZSTD 7.7, yang menunjukkan tingkat misfit yang tinggi dan item yang tidak sesuai dengan model pengukuran atau memiliki interpretasi yang berbeda di antara responden (Linacre, 2012). Sementara itu, LR47 dan LR48 memiliki infit dan outfit ZSTD -4.1 dan -4.2, menunjukkan bahwa item ini terlalu mudah bagi responden atau kurang memberikan informasi diskriminatif yang cukup (Zumbo, 2007).

Dalam analisis Rasch, nilai ZSTD yang berada di luar rentang -2.0 – 2.0 dapat diabaikan jika nilai MNSQ dan Point Measure Correlation berada dalam batas wajar, karena ZSTD sangat sensitif terhadap ukuran sampel (Sumintono & Widhiarso, 2013). Dengan demikian, item LR46 perlu dikaji ulang untuk memastikan kesesuaian dengan model pengukuran, karena selain nilai ZSTD yang berada di luar rentang batas, nilai infit dan outfit MNSQ juga berada di atas 1.5, sedangkan item lainnya dapat dinyatakan sesuai dengan model.

https://doi.org/10.36269/psyche.v7i2.3049

PSYCHE: Jurnal Psikologi Universitas Muhammadiyah Lampung
Vol. 7 No.2, Agustus 2025
ISSN (*electronic*) 2655-6936
ISSN (*printed*) 2686-0430

Tabel 5. *Item Misfit*

Item	*Measure*	*Infit MNSQ*	*Outfit MNSQ*	*Infit ZSTD*	*Outfit ZSTD*	*Point Measure Correlation*
MS9	-.11	1.14	1.14	1.8	1.9	.49
MS1	-.47	1.11	1.06	1.4	.7	.56
MS12	.61	1.07	1.11	1.0	1.5	.64
MS2	.30	1.10	1.10	1.4	1.3	.57
MS13	-.55	1.02	1.04	.3	.5	.50
MS14	.01	1.03	1.03	.4	.5	.58
MS3	.18	1.01	1.02	.2	.2	.63
MS10	-.21	.92	.95	-1.1	-.7	.56
MS11	-.12	.93	.92	-1.0	-1.0	.62
MS8	.58	.92	.91	-1.2	-1.2	.70
MS6	-.10	.81	.79	-2.7	-3.0	.70
MS7	-.12	.78	.79	-3.1	-3.1	.66
GS19	.30	1.08	1.06	1.2	.9	.76
GS17	.05	1.05	1.02	.7	.3	.78
GS16	-.47	.96	1.01	-.5	.1	.74
GS18	-.06	.96	.97	-.6	-.4	.79
GS15	.17	.88	.89	-1.8	-1.7	.81
SE20	-.73	1.39	1.33	4.5	3.7	.46
SE24	.97	1.06	1.14	.8	1.9	.70
SE31	-.11	1.11	1.10	1.5	1.3	.54
SE21	-.50	1.04	1.05	.6	.6	.49
SE29	.06	.99	1.01	.0	.1	.60
SE26	-.01	.92	.99	-1.0	-.1	.64
SE28	-.52	.97	.98	-.4	-.2	.50
SE23	.41	.96	.95	-.5	-.7	.66
SE25	.58	.93	.93	-1.0	-.9	.69
SE22	.26	.85	.88	-2.1	-1.8	.62
SE27	-.48	.91	.83	-2.6	-2.2	.61
SE30	.07	.73	.76	-3.9	-3.7	.68
SA33	.58	1.20	1.23	2.6	2.9	.65
SA36	-.80	1.01	1.07	.1	.8	.49
SA34	-.16	1.06	1.05	.8	.6	.61
SA35	-.29	.96	1.00	-.5	.1	.57
SA38	.43	.84	.86	-2.3	-2.0	.72
SA39	.24	.77	.78	-3.3	-3.2	.69
ES42	-.55	1.19	1.30	2.3	3.3	54
ES41	-.67	.93	1.01	-.9	.2	62
ES43	.21	.85	.84	-2.2	-2.3	73
ES40	1.01	.83	.84	-2.6	-2.3	78
LR46	.33	1.63	1.63	7.7	7.7	.43
LR44	-1.15	1.03	1.09	.4	.9	.48
LR45	.70	.91	.85	-3.0	-2.3	.71
LR47	.05	.73	.75	-4.1	-3.8	.70
LR48	.08	.73	.74	-4.2	-4.0	.72
O49	.18	1.11	1.07	1.4	1.0	.68
O51	.35	.99	1.01	-.1	.1	.69
O50	-.11	.91	.95	-1.1	-.7	.62
O52	-.42	.92	.91	-.9	-1.2	.66

https://doi.org/10.36269/psyche.v7i2.3049

Wright Map

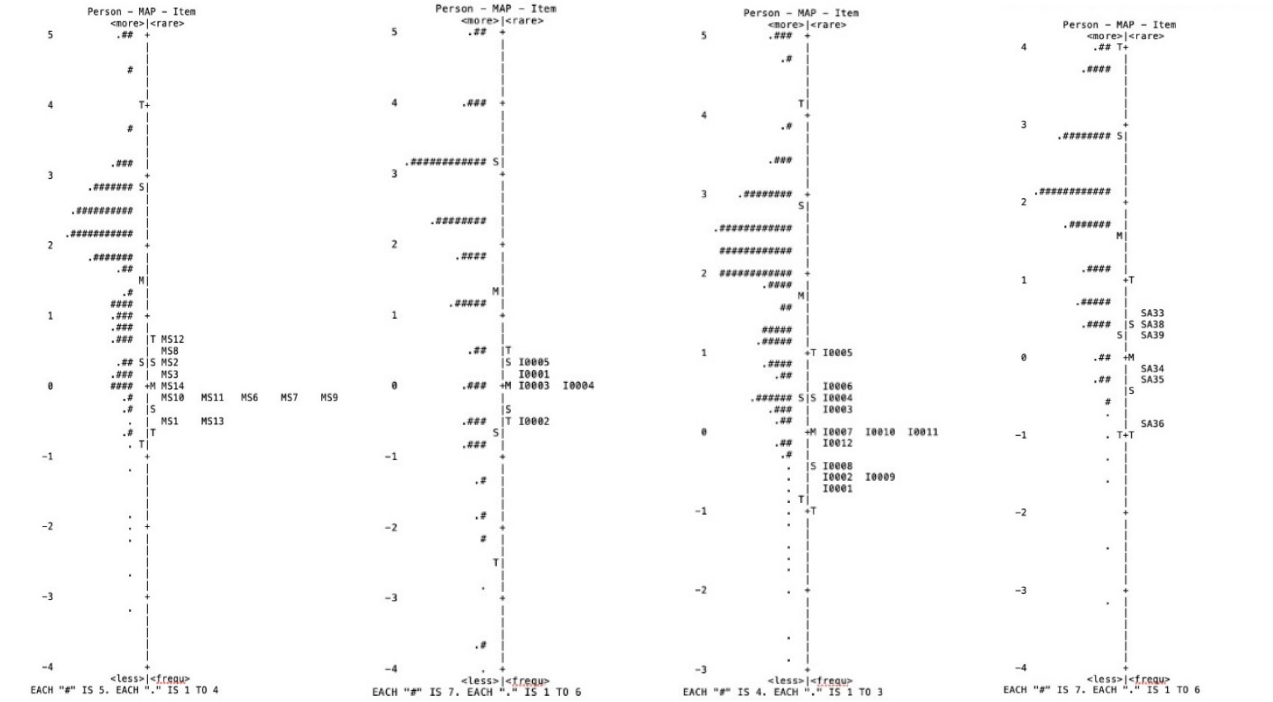
Validitas konstruk dapat dianalisis dan divisualisasikan menggunakan *Wright Map* yang diterapkan dalam model Rasch. *Wright Map,* atau yang juga dikenal sebagai *Item-Person Map*, memberikan representasi grafis yang memadukan tingkat kemampuan responden dengan tingkat kesulitan item dalam satu skala logit yang sama (Boone, Staver, & Yale, 2014). Dalam *Wright Map,* kemampuan responden ditempatkan pada sisi kiri, sedangkan tingkat kesulitan item terletak di sisi kanan.

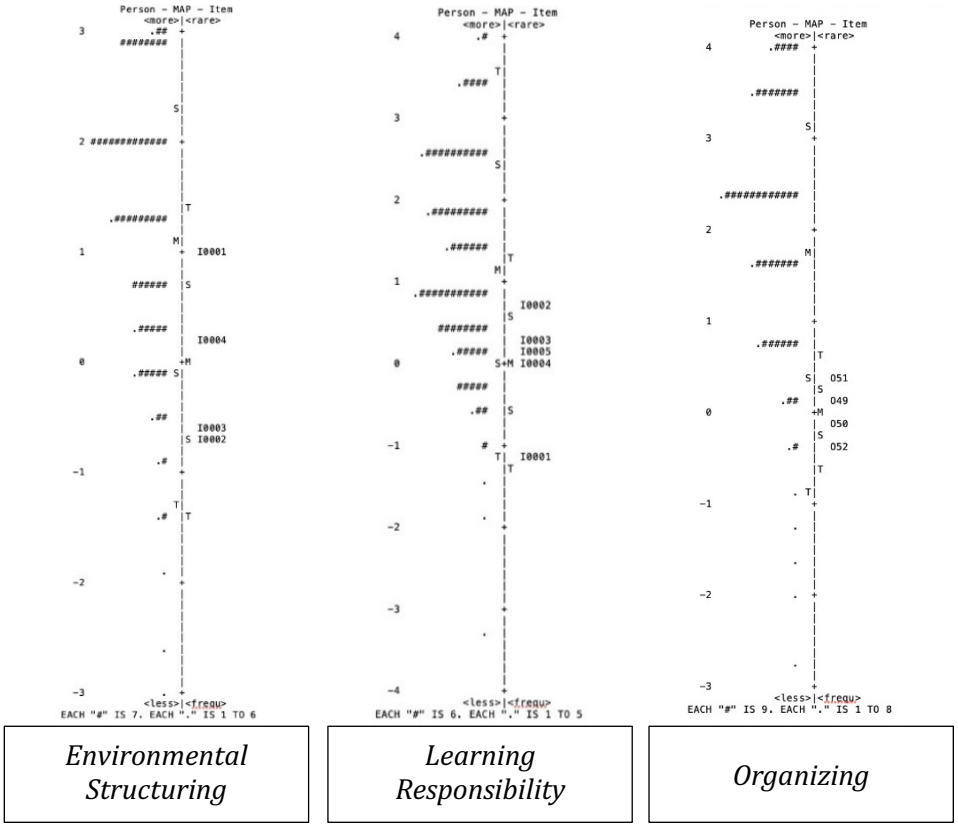
Tingkat kesulitan item ditampilkan dalam urutan dari yang paling mudah di bagian bawah hingga yang paling sulit di bagian atas. Hal ini memungkinkan peneliti untuk memeriksa apakah tingkat kemampuan responden tersebar secara merata di sepanjang skala, serta apakah terdapat kecocokan antara kemampuan responden dengan kesulitan item yang diukur (Linacre, 2007). Dengan kata lain, *Wright Map* tidak hanya membantu mengevaluasi sejauh mana item mencerminkan konstruk yang diukur, tetapi juga memberikan gambaran tentang sejauh mana item tersebut mampu membedakan individu dengan kemampuan yang berbeda. Pemetaan ini juga bermanfaat untuk mengidentifikasi item yang mungkin tidak sesuai dengan kemampuan mayoritas responden, seperti item yang terlalu mudah atau terlalu sulit. Selain itu, *Wright Map* dapat digunakan untuk mendeteksi item yang mungkin memerlukan revisi, terutama jika item tersebut tidak berfungsi sebagaimana mestinya atau tidak memberikan informasi yang cukup pada tingkat kemampuan tertentu (Bond & Fox, 2015).

Gambar 2 menunjukkan bahwa skala A-SRL memiliki tingkat kesulitan item yang kurang merata. Item lebih banyak mengukur responden dengan kemampuan rendah hingga sedang, serta kurang mampu mengukur responden dengan kemampuan yang tinggi. Ketidakmerataan tingkat kesulitan item menunjukkan bahwa skala lebih terfokus pada kelompok responden dengan kemampuan rendah hingga sedang. Instrumen dapat memberikan informasi yang baik untuk responden pada tingkat kemampuan ini, tetapi kurang efektif untuk menilai responden dengan kemampuan tinggi. Selain itu, kurangnya item yang menargetkan kemampuan tinggi menciptakan kesenjangan dalam pengukuran, sehingga skala tidak mampu membedakan antara responden dengan tingkat kemampuan tinggi. Dalam konteks validitas konstruk, skala belum sepenuhnya representatif untuk seluruh dimensi atau variasi dari konstruk yang diukur.

https://doi.org/10.36269/psyche.v7i2.3049

PSYCHE: Jurnal Psikologi Universitas Muhammadiyah Lampung
Vol. 7 No.2, Agustus 2025
ISSN (*electronic*) 2655-6936
ISSN (*printed*) 2686-0430





Gambar 2. *Wright Map* Skala A-SRL Versi Indonesia

https://doi.org/10.36269/psyche.v7i2.3049

Differential Item Functioning

Analisis Differential Item Functioning (DIF) bertujuan untuk mengidentifikasi dan mengevaluasi apakah terdapat perbedaan pola respons antar sub-kelompok dalam sampel yang memiliki tingkat karakteristik serupa terhadap konstruk yang diukur (Zumbo, 1999). DIF mengukur apakah item tertentu dalam instrumen cenderung memberikan keuntungan atau kerugian kepada kelompok tertentu, sehingga dapat digunakan untuk memastikan keadilan dan validitas instrumen dalam populasi yang heterogen (Penfield & Camilli, 2007).

Dalam penelitian ini, sub-kelompok dibedakan berdasarkan jenis kelamin, yaitu laki-laki dan perempuan. DIF dianalisis menggunakan metode item *trait chi-square* sebagaimana dijelaskan oleh Linacre (2007). Metode ini menggunakan pendekatan berbasis Rasch, yang secara khusus mengevaluasi apakah parameter item, seperti tingkat kesulitan, berbeda antara sub-kelompok. Item yang memiliki nilai probabilitas kurang dari 0,05 dianggap menunjukkan bias signifikan, yang berarti item tersebut memperlakukan kelompok secara tidak adil (Linacre, 2007). Selain itu, teknik DIF dapat memberikan wawasan tentang sejauh mana item berfungsi secara seragam di seluruh populasi. Bias pada item dapat merusak interpretasi hasil pengukuran, sehingga mendeteksi dan mengeliminasi item bias merupakan langkah penting dalam pengembangan instrumen yang valid dan adil (Holland & Wainer, 1993). Teknik ini juga relevan dalam konteks pengujian berbasis kelompok yang beragam, seperti pendidikan, psikologi, atau survei populasi umum, untuk memastikan bahwa hasil tidak terpengaruh oleh perbedaan sub-kelompok non-konstruk (Clauser & Mazor, 1998).

Selain metode *item trait chi-square* sebagaimana dijelaskan oleh Linacre (2007), terdapat pendekatan lain yang sering digunakan dalam analisis DIF berbasis Rasch, seperti Welch's t-test dan Mantel-Haenszel test. Welch's t-test membandingkan perbedaan estimasi DIF *contrast* antara dua kelompok dengan mempertimbangkan varians yang tidak sama. Jika nilai t menunjukkan signifikansi ($p < 0.05$), maka item tersebut memiliki perbedaan yang signifikan antar kelompok dan berpotensi bias (Zumbo, 1999). Sementara itu, Mantel-Haenszel test mengevaluasi rasio peluang dari kemungkinan menjawab benar antara kelompok dengan tingkat kemampuan setara. Jika nilai *chi-square* menunjukkan signifikansi ($p < 0.05$), maka item dianggap memiliki DIF yang signifikan (Mantel & Haenszel, 1959). Dengan menggunakan kedua metode ini, peneliti dapat mengidentifikasi item yang berfungsi secara tidak adil dan

https://doi.org/10.36269/psyche.v7i2.3049

mempertimbangkan revisi atau eliminasi item untuk meningkatkan validitas instrumen (Penfield & Camilli, 2007).

Berdasarkan temuan dari hasil analisis DIF yang dilakukan untuk menguji bias *gender*, diperoleh probabilitas (Welch) 0.0065 pada item LR44 ("saya memeriksa ulang tugas saya untuk memastikan semuanya benar sebelum mengumpulkannya"), dan 0.0038 pada item LR46 ("Saya cemas dengan batas waktu pengumpulan tugas yang ditetapkan dosen"). Hal ini mengindikasikan adanya perbedaan interpretasi antara laki-laki dan perempuan dalam memahami item LR44 dan LR46. Selain itu, pada probability Mantel-Haenzel, hanya item LR44 saja yang mengalami bias gender ($p<0.05$).

Laki-laki cenderung memilih skala penilaian yang lebih tinggi dibandingkan perempuan, sedangkan pada item LR46, perempuan cenderung memilih skala penilaian yang lebih tinggi dibandingkan laki-laki. Oleh karena itu, perlu kehati-hatian dalam menggunakan item LR44 dan LR46 karena terdapat kecenderungan untuk bias *gender.*

Tabel 6. *Differential Item Functioning*

Item	Probability (Welch)	Probability (Mantel-Haenzel)
LS44	0.0065	0.059
LS46	0.0037	.1485

Rasch Mixture Model

Rasch Mixture Model (RMM) dalam penelitian ini digunakan untuk mengidentifikasi heterogenitas data dengan mengelompokkan individu ke dalam kelas-kelas laten. Model ini sangat efektif dalam menganalisis data yang memiliki populasi responden yang tidak homogen, memungkinkan peneliti untuk mengungkap pola-pola tersembunyi berdasarkan karakteristik responden. Dengan demikian, RMM memberikan pemahaman yang lebih dalam mengenai hubungan antara item dan pola respons, serta memperlihatkan bagaimana kelompok-kelompok responden berbeda dapat memengaruhi hasil analisis (Rost, 1990).

Fit indices digunakan untuk mengevaluasi kecocokan model ini. Salah satunya adalah *Akaike Information Criterion* (AIC), yang mengukur keseimbangan antara kecocokan model dan kompleksitasnya. AIC memberikan penalti terhadap jumlah parameter yang digunakan, sehingga nilai AIC yang lebih rendah menunjukkan model yang lebih optimal dalam menjelaskan data tanpa *overfitting* (Burnham & Anderson, 2004). Selain itu, *Bayesian Information Criterion* (BIC) bekerja dengan cara serupa, tetapi

https://doi.org/10.36269/psyche.v7i2.3049

memberikan penalti yang lebih besar terhadap kompleksitas model, khususnya pada ukuran sampel besar, sehingga lebih konservatif dalam menentukan model terbaik. Nilai BIC yang lebih rendah menunjukkan bahwa model memiliki keseimbangan yang baik antara kecocokan data dan kesederhanaan struktur model (Schwarz, 1978).

Selanjutnya, *Consistent Akaike Information Criterion* (CAIC) menawarkan pendekatan lebih ketat dibandingkan AIC dan BIC, dengan memberikan penalti tambahan pada kompleksitas model untuk menjaga konsistensi, terutama pada sampel besar. Hal ini menjadikan CAIC sebagai indeks yang sangat selektif dalam memilih model yang optimal (Bozdogan, 1987). *Log-likelihood* (LL), yang juga digunakan, mengukur seberapa baik model memprediksi data yang diamati. Nilai LL yang lebih tinggi atau mendekati nol menunjukkan kemampuan prediksi model yang lebih baik. Namun, LL sering digunakan bersama indeks lainnya, seperti AIC, BIC, dan CAIC, untuk memberikan penilaian komprehensif terhadap model (Christensen et al., 2013).

Berdasarkan Tabel 7, diketahui bahwa aspek *learning responsibility* lebih optimal dengan model dua kelas karena menunjukkan penurunan nilai AIC, BIC, dan CAIC, serta peningkatan *log-likelihood,* yang mencerminkan kemampuan model untuk menangkap heterogenitas responden secara lebih baik (Burnham & Anderson, 2004; Schwarz, 1978). Sementara itu, pada aspek *seeking assistance*, *memory strategy*, *self-evaluation* dan *environmental structuring*, meskipun nilai AIC menurun pada model dua kelas, terjadi kenaikan nilai BIC dan CAIC yang berarti bahwa penalti terhadap kompleksitas model menyebabkan model dua kelas menjadi kurang optimal dibandingkan model satu kelas (Bozdogan, 1987). Oleh karena itu, model satu kelas dipilih untuk aspek-aspek ini, karena lebih sederhana dan efisien dalam menggambarkan data.

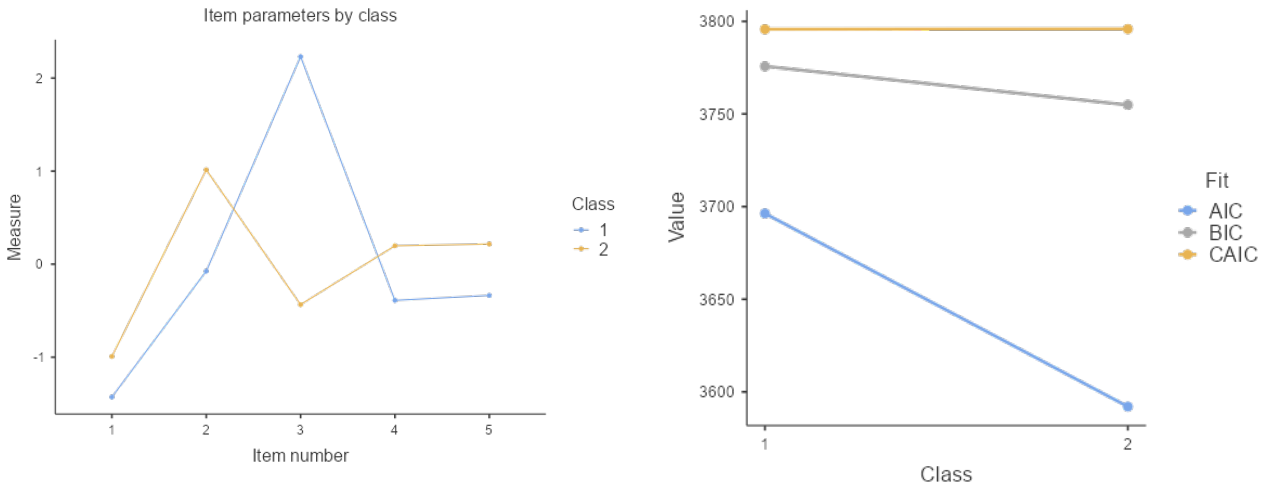
Lebih lanjut, pada aspek *goal setting* dan *organizing*, terjadi peningkatan nilai AIC, BIC, dan CAIC pada model dua kelas, yang mengindikasikan bahwa model satu kelas lebih sesuai. Peningkatan nilai tersebut menunjukkan bahwa model dua kelas tidak memberikan keunggulan signifikan dalam memprediksi data, tetapi menambah kompleksitas model sehingga kurang efisien dibandingkan model satu kelas (Schwarz, 1978). Dalam penerapan analisis model laten *sub-class*, penurunan AIC dan BIC sering digunakan sebagai indikator untuk menambahkan laten *sub-class*, terutama jika BIC menunjukkan penurunan karena sifatnya lebih konservatif. Jika kedua kriteria tersebut mengalami penurunan, hal ini semakin menguatkan bukti bahwa penambahan laten *sub-*

https://doi.org/10.36269/psyche.v7i2.3049

class meningkatkan kualitas model. BIC direkomendasikan sebagai kriteria utama karena kemampuannya mencegah *overfitting* dengan memilih model yang lebih sederhana (Nylund et al., 2007). Oleh karena itu, penilaian penambahan laten *sub-class* sebaiknya berfokus pada penurunan AIC dan BIC, dengan perhatian khusus pada BIC. Dengan demikian, berdasarkan analisis *Rasch Mixture Model*, diperoleh kesimpulan bahwa aspek *Learning Responsibility* terdapat laten *sub-class* yang terdiri dari dua kelas, sementara aspek lainnya tidak terdapat laten *sub-class.*

Tabel 7. *Model fit Information*

Aspek	Class	AIC	BIC	CAIC	Log-likelihood
Memory Strategy	1	9759	9945	9992	-4832
	2	9635	10011	10106	-4722
Goal Setting	1	3230	3313	3334	-1594
	2	3323	3403	3446	-1547
Self Evaluation	1	7877	8051	8095	-3895
	2	7728	8080	8169	-3775
Seeking Assistance	1	5263	5378	5407	-2602
	2	5189	5423	5482	-2536
Environmental Structuring	1	2812	2879	2898	-1389
	2	2760	2898	2933	-1234
Learning Responsibility	1	3969	3776	3796	-1828
	2	3592	3755	3796	-1755
Organizing	1	3670	3760	3783	-1812
	2	3677	3861	3908	-1792



Gambar 3. *Grafik Distribusi Kelas dan Elbow Plot sub-skala Learning Responsibility*

Hasil pada Tabel 8 menunjukkan bahwa distribusi gender tidak secara langsung menjadi faktor utama pembentukan laten *sub-class.* Namun pada Tabel 9 menunjukkan adanya perbedaan signifikan dalam distribusi kelas berdasarkan angkatan, di mana

https://doi.org/10.36269/psyche.v7i2.3049

mahasiswa angkatan 2020 cenderung mendominasi kelas satu, sementara kelas dua lebih banyak diisi oleh angkatan 2021. *Learning responsibility* dapat dipengaruhi oleh tugas perkembangan individu, seperti penyesuaian terhadap tanggung jawab akademik, membangun hubungan sosial yang matang, dan pengelolaan diri secara mandiri (Havighurst, 1972). Dalam konteks ini, mahasiswa angkatan 2020 lebih mungkin telah menyelesaikan fase perkembangan tersebut dibandingkan dengan angkatan 2021 yang masih dalam tahap awal adaptasi. Selain itu, pada fase *identity vs role confusion* menurut Erikson (1968), individu mulai membangun identitas yang lebih bertanggung jawab, yang dapat tercermin dalam perilaku belajar mereka.

Tabel 8. *Frekuensi Kelas berdasarkan Gender pada Aspek Learning Responsibility*

Gender	Kelas	Jumlah	%	Kumulatif (%)
Perempuan	1	60	4.85	4.85
	2	222	23.21	28.06
Laki-laki	1	19	15.31	43.37
	2	91	56.63	100

Tabel 9. *Frekuensi Kelas berdasarkan Angkatan pada Aspek Learning Responsibility*

Angkatan	Kelas	Jumlah	%	Kumulatif (%)
2020	1	74	18.45	4.85
	2	37	9.23	28.06
2021	1	180	44.89	43.37
	2	110	27.43	100

DISKUSI

Hasil analisis unidimensionalitas menunjukkan bahwa setelah penghapusan enam item (MS4, MS5, SA32, SA37, O53, dan O54), setiap sub-skala dalam A-SRL-S mencapai unidimensionalitas dengan 48 item. Penghapusan ini dilakukan karena item-item tersebut memiliki *residual loadings* >0.5, yang dapat mengindikasikan adanya varians yang tidak dijelaskan oleh dimensi utama (Linacre, 2009). Sebelum penghapusan, sub-skala MS, SA, dan O memiliki *raw variance explained by measure* yang lebih rendah (<40%), yang menunjukkan kemungkinan multidimensionalitas atau kurangnya kontribusi terhadap faktor utama (Smith, 2002).

Penghapusan item ini meningkatkan kualitas pengukuran dengan memastikan bahwa setiap sub-skala tetap unidimensional dan sesuai dengan prinsip dasar model Rasch (Bond & Fox, 2015). Faktor seperti kesulitan pemahaman item atau perbedaan

https://doi.org/10.36269/psyche.v7i2.3049

interpretasi dalam konteks budaya dapat berkontribusi terhadap rendahnya varian yang dijelaskan sebelum reduksi (Van de Vijver & Poortinga, 1997). Dengan demikian, analisis lanjutan dengan 48 item dapat memberikan hasil yang lebih valid dan reliabel dalam mengukur aspek *self-regulated learning.*

Analisis *rating scale* menunjukkan bahwa semua sub-skala dalam A-SRL-S memiliki threshold yang meningkat secara bertahap dari negatif ke positif, yang mengindikasikan bahwa responden mampu membedakan setiap opsi respons dengan baik (Linacre, 2012). Sementara itu, seluruh sub-skala memiliki nilai infit dan outfit MNSQ sesuai dengan model Rasch, dengan rentang 0.5-1.5 (Bond & Fox 2015). Struktur skala respons yang berfungsi dengan baik harus menunjukkan perbedaan yang jelas di antara kategori pilihan jawaban (Wright & Masters, 1982). Oleh karena itu, hasil ini mengonfirmasi bahwa skala A-SRL-S mampu mengukur setiap aspek *self-regulated learning* dengan konsistensi yang baik.

Hasil analisis reliabilitas menunjukkan bahwa semua sub-skala A-SRL memiliki nilai Cronbach's Alpha di atas 0.8. Hal ini mengindikasikan bahwa item-item dalam masing-masing sub-skala memiliki konsistensi internal yang tinggi dalam mengukur konstruk yang dimaksud. Temuan ini menunjukkan bahwa alat ukur A-SRL yang digunakan dalam penelitian ini cukup andal. Berdasarkan kriteria kualitas instrumen yang dikembangkan oleh William P. Fisher, Jr., nilai reliabilitas pengukuran (baik person maupun item) yang berada pada rentang ≥ 0.81 hingga > 0.94 dikategorikan sebagai *good* hingga *excellent.* Sejalan dengan hal ini, nilai reliabilitas sebesar >0.8 dianggap memadai, terutama ketika instrumen digunakan dalam konteks pengukuran terapan (Nunnally & Bernstein, 1994). *Person separation index* seluruhnya berada di bawah 2.0 yang dapat mengindikasikan bahwa partisipan dalam penelitian ini memiliki karakteristik yang relatif homogen, sehingga skala kurang mampu membedakan individu dengan tingkat kemampuan yang berbeda (Boone et al., 2014). Homogenitas partisipan dapat terjadi ketika sampel berasal dari kelompok dengan pengalaman atau kemampuan serupa, yang menyebabkan distribusi respons menjadi terbatas (Linacre, 2012). Dalam kondisi seperti ini, diperlukan sampel yang lebih beragam untuk meningkatkan sensitivitas skala dalam mengukur perbedaan individu secara lebih akurat (Wright & Masters, 1982).

Selain itu, hasil strata pada semua sub-skala berada di bawah ambang batas minimal (3.0) yang menunjukkan bahwa ketajaman instrumen dalam mengelompokkan

https://doi.org/10.36269/psyche.v7i2.3049

individu ke dalam level kemampuan yang berbeda masih terbatas. Dengan kata lain, strata <3.0 mencerminkan bahwa *separation* yang dihasilkan oleh data belum cukup untuk mengidentifikasi setidaknya tiga level kemampuan yang berbeda dalam populasi yang diuji (Linacre, 2012).

Hasil analisis item misfit menunjukkan bahwa item LR46 memiliki nilai ZSTD yang berada di luar rentang ideal (-2.0 – 2.0) dan nilai infit-outfit MNSQ yang melebihi batas ideal (0.5 - 1.5). Hal ini mengindikasikan bahwa item tersebut memiliki pola respons yang tidak sesuai dengan model dan perlu dikaji ulang. Sebaliknya, meskipun sejumlah item memiliki nilai ZSTD di luar rentang –2.0 hingga +2.0, namun nilai infit-outfit MNSQ dan *point-measure correlation* menunjukkan hasil yang masih berada dalam batas yang dapat diterima (Sumintono & Widhiarso, 2013).

Tingkat kesulitan item dalam skala A-SRL yang tidak merata menunjukkan bahwa instrumen lebih efektif dalam mengukur responden dengan kemampuan rendah hingga sedang, tetapi kurang sensitif dalam menilai responden dengan kemampuan tinggi. Dari analisis reliabilitas person yang rendah pada seluruh sub-skala menunjukkan bahwa skala kurang mampu membedakan individu berdasarkan tingkat kemampuannya secara konsisten (Linacre, 2012). Kurangnya item yang menargetkan kemampuan tinggi juga dapat menyebabkan bias dalam pengukuran dan menurunkan validitas konstruk (Bond & Fox, 2015). Diperlukan pengembangan item tambahan yang lebih menargetkan responden dengan kemampuan tinggi agar skala dapat memberikan gambaran yang lebih komprehensif mengenai seluruh spektrum kemampuan yang diukur.

Hasil analisis DIF menunjukkan bahwa item LR44 dan LR46 mengalami bias gender, di mana laki-laki dan perempuan memiliki kecenderungan respon yang berbeda terhadap item tersebut. Pada analisis item mistfit, LR46 memiliki nilai Infit dan Outfit MNSQ tertinggi, yang mengindikasikan adanya variasi respons yang tidak konsisten. Ketidaksesuaian ini dapat disebabkan oleh perbedaan interpretasi terhadap makna item berdasarkan pengalaman atau persepsi yang berbeda antara laki-laki dan perempuan (Zumbo, 2007). Selain itu, bias gender dapat terjadi ketika formulasi item tidak mewakili pengalaman semua *gender,* sehingga menimbulkan perbedaan pola respons (Camilli & Shepard, 1994). Evaluasi lebih lanjut diperlukan untuk memastikan bahwa LR44 dan LR46 tidak menghambat keandalan serta interpretasi hasil pengukuran secara adil bagi seluruh responden.

https://doi.org/10.36269/psyche.v7i2.3049

Analisis *Rasch Mixture* Model menunjukkan bahwa aspek LR lebih optimal dengan model dua kelas, sedangkan sub-skala lainnya lebih sesuai dengan model satu kelas. Temuan ini mengindikasikan adanya heterogenitas responden dalam LR, yang mungkin berkaitan dengan perbedaan strategi belajar atau keterlibatan akademik (Nylund et al., 2007). Sementara itu, kesesuaian model satu kelas pada MS, GS, SE, SA, MS, dan ES menunjukkan bahwa variabilitas respons di aspek-aspek ini tidak cukup kuat untuk membentuk sub-kelas terpisah (Bozdogan, 1987). Pada item misfit, ketidakkonsistenan respons pada LR46 menunjukkan bahwa faktor tambahan seperti bias gender atau interpretasi berbeda dapat mempengaruhi model yang dipilih. Oleh karena itu, dalam pengembangan instrumen lebih lanjut, penting untuk mempertimbangkan faktor-faktor ini agar model dapat menangkap struktur laten responden secara lebih akurat.

Penggunaan teknik *convenience sampling* dapat menimbulkan bias seleksi karena tidak semua anggota populasi memiliki peluang yang sama untuk terpilih sebagai partisipan. Hal ini dapat membuat hasil penelitian kurang merepresentasikan keseluruhan populasi mahasiswa dan membatasi generalisasi temuan. Penelitian selanjutnya, disarankan menggunakan metode sampling yang lebih representatif untuk meningkatkan validitas eksternal.

KESIMPULAN DAN SARAN

Setelah dilakukan penghapusan enam item (MS4, MS5, SA32, SA37, O53, dan O54), setiap sub-skala pada A-SRL-S berhasil memenuhi unidimensionalitas dengan 48 item. Reliabilitas item di seluruh sub-skala menunjukkan hasil yang sangat baik (> 0.8), namun *person reliability* dan *person separation index* tidak menunjukkan reliabilitas yang memadai, dengan nilai *person reliability* berada dalam rentang 0.46 – 0.79 dan *person separation index* berada dalam rentang 0.92 – 1.94, yang mengindikasikan bahwa instrumen belum cukup sensitif untuk membedakan individu berdasarkan tingkat kemampuannya secara bermakna (Linacre, 2023). Strata sub-skala juga berada di bawah ambang batas (3.0) yang mengindikasikan perlunya peningkatan jumlah atau kualitas item agar instrumen lebih sensitif dan akurat dalam mengukur perbedaan kemampuan.

Hasil analisis item *misfit* mengidentifikasi bahwa item LR46 memiliki *infit* dan *outfit* MNSQ sebesar 1.63 dengan ZSTD 7.7, yang menunjukkan *misfit* yang signifikan. Selain itu, bias gender ditemukan pada item LR44 dan LR46 dengan probabilitas Welch

https://doi.org/10.36269/psyche.v7i2.3049

0.0065 dan 0.0037, yang menunjukkan perbedaan respon yang signifikan antara laki-laki dan perempuan (Camilli & Shepard, 1994). *Rasch Mixture Model* menunjukkan bahwa sub-skala *learning responsibility* lebih optimal dengan model dua kelas, sementara sub-skala lainnya lebih sesuai dengan model satu kelas. Dengan adanya pembagian ini, instrumen dapat disesuaikan dengan menetapkan *cut-off score* yang lebih spesifik untuk masing-masing kelas, memungkinkan perbedaan tanggung jawab belajar yang lebih tepat sesuai dengan karakteristik individu dalam setiap kelompok. Penyesuaian ini memberikan instrumen kemampuan yang lebih baik untuk menangkap perbedaan tingkat tanggung jawab belajar antara individu dengan kemampuan yang berbeda.

Diperlukan pengembangan item yang lebih menargetkan responden dengan kemampuan tinggi untuk meningkatkan sensitivitas skala dalam mengukur variabilitas individu. Penambahan item yang lebih mengakomodasi responden dengan kemampuan tinggi dapat memperbaiki reliabilitas person dan PSI, yang saat ini menunjukkan keterbatasan dalam membedakan individu dengan kemampuan yang berbeda (Boone et al., 2014). Selain itu, penggunaan sampel yang lebih beragam di masa mendatang akan membantu meningkatkan distribusi respons, sehingga skala dapat lebih akurat dalam menggambarkan perbedaan individu secara lebih konsisten (Linacre, 2012).

Secara keseluruhan, implikasi dari temuan penelitian adalah perlunya revisi item-item misfit yaitu SE20, SA33, ES42, dan LR46. Meskipun infit dan outfit MNSQ serta *point biserial correlation* memenuhi batas item fit. Selain itu, perlu adanya revisi item, terutama pada item yang dipersepsikan multidimensi, yaitu item MS4, MS5, SA32, SA37, O53, dan O54. Lebih lanjut, berdasarkan temuan pada *latent class analysis,* meskipun gender tidak secara langsung memengaruhi pembentukan *latent sub-class*, perbedaan angkatan memberikan kontribusi yang signifikan terhadap distribusi kelas. Hal ini mengindikasikan bahwa karakteristik perkembangan individu berdasarkan tahap pendidikan dapat memengaruhi tanggung jawab belajar (*learning responsibility*). Oleh karena itu perlu dilakukan *renorming* berdasarkan kelompok *laten class,* khususnya dengan mempertimbangkan perbedaan tahap perkembangan psikososial antar angkatan. Reliabilitas *person* yang rendah juga menunjukkan perlunya replikasi penelitian dengan partisipan yang lebih heterogen untuk meningkatkan sensitivitas skala dalam mengukur perbedaan individu secara lebih akurat (Wright & Masters, 1982). Dalam proses adaptasi skala ke dalam konteks Indonesia, penting untuk mempertimbangkan

https://doi.org/10.36269/psyche.v7i2.3049

aspek-aspek budaya lokal, seperti orientasi kolektivisme dan norma-norma akademik yang dapat berbeda dengan konteks asal pengembangan skala di Filipina. Penyesuaian ini diperlukan untuk memastikan bahwa instrumen tidak hanya valid secara linguistik, tetapi juga relevan secara kultural.

DAFTAR PUSTAKA

Andiani, S. (2017). Hubungan prestasi akademik dan strategi regulasi diri dalam belajar pada mahasiswa tunarungu. *SRLCalyptra: Jurnal Ilmiah Mahasiswa Universitas Surabaya*, *6*(2), 1–10.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573. https://doi.org/10.1007/BF02293814

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer. https://doi.org/10.1007/978-94-007-6857-4

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd edition). Routledge.

Bozdogan, H. (1987). Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions. *Psychometrika*, *52*(3), 345–370. https://doi.org/10.1007/BF02294361

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Sage Publications.

Christensen, K. B., Kreiner, S., & Mesbah, M. (Eds.). (2012). Front Matter. In *Rasch Models in Health* (1st ed.). Wiley. https://doi.org/10.1002/9781118574454.fmatter

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, *17*(1), 31–44.

Duncan, T. G., & McKeachie, W. J. (2005). The Making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist*, *40*(2), 117–128. https://doi.org/10.1207/s15326985ep4002_6

Erikson, E. H. (1968). *Identity: Youth and crisis*. W. W. Norton & Company.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1. https://doi.org/10.11648/j.ajtas.20160501.11

Fisher, W. P., Jr. (n.d.). *Rating scale instrument quality criteria*. Retrieved April 11, 2025, from https://www.winsteps.com/facetman/reliability.htm

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed). McGraw-Hill Humanities/Social Sciences/Languages.

Gliem, J. A., & Gliem, R. R. (n.d.). Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. *2003*.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (Eighth edition). Cengage.

Havighurst, R. J. (1972). *Developmental tasks and education* (3rd ed.). Longman.

Holland, P. W., & Wainer, H. (1993). *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, *7*(4), 328.

https://doi.org/10.36269/psyche.v7i2.3049

Linacre, J. M. (2002). What do Infit and Outfit, Mean-square and Standardized mean? *Rasch Measurement Transactions*, *16*(2), 878.

Linacre, J. M. (2007). *A User's Guide to WINSTEPS: Rasch Model Computer Program*. Winsteps.com.

Linacre, J. M. (2007). Sample size and item calibration stability. *Rasch Measurement Transactions*, *21*(1), 1095.

Linacre, J. M. (2009). *A User's Guide to WINSTEPS*. Winsteps.com.

Linacre, J. M. (2012). *Reliability and separation of measures*. https://www.winsteps.com/winman/reliability.htm

Linacre, J. M. (2023). *Reliability and separation of measures*. https://www.winsteps.com/winman/reliability.htm

Magno, C. (2011). The predictive validity of the academic self-regulated learning scale. *The International Journal of Educational and Psychological Assessment, 9*(1), 45-58. Time Taylor Academic Journals.

Magno, C. (2010). *Assessing Academic Self-Regulated Learning among Filipino College Students: The Factor Structure and Item Fit*. *5*.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometri theory* (3rd edition). McGraw-Hill.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396

Padilla, J.-L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, *1*(26), 136–144. https://doi.org/10.7334/psicothema2013.259

Penfield, R. D., & Camilli, G. (2007). Differential item functioning and item bias. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics: Vol. 26. Psychometrics* (pp. 125–167). Elsevier.

Raosoft. (2024). Sample size calculator. http://www.raosoft.com/samplesize.html.

Rost, J. (1990). Rasch Models in Latent Classes: An Integration of Two Approaches to Item Analysis. *Applied Psychological Measurement*, *14*(3), 271–282. https://doi.org/10.1177/014662169001400305

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2). https://doi.org/10.1214/aos/1176344136

Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, *1*(26), 100–107. https://doi.org/10.7334/psicothema2013.256

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.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18

Sumintono, B. (2016). *Aplikasi Model Rasch dalam Penelitian Pendidikan dan Psikologi*. Penerbit Unika Soegijapranata.

Sumintono, B., & Widhiarso, W. (2013). *Aplikasi model rasch: Untuk penelitian ilmu-ilmu sosial*. Trim Komunikata Publishing House.

The Jamovi Project. (2024). *Jamovi* (Version 2.6) [Computer software]. https://www.jamovi.org

Van De Vijver, F. J. R., & Leung, K. (2021). *Methods and Data Analysis for Cross-Cultural Research* (V. H. Fetvadjiev, J. R. J. Fontaine, & J. He, Eds.; 2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781107415188

https://doi.org/10.36269/psyche.v7i2.3049

Van De Vijver, F. J. R., & Poortinga, Y. H. (1997). Towards an Integrated Analysis of Bias in Cross-Cultural Assessment. *European Journal of Psychological Assessment*, *13*(1), 29–37. https://doi.org/10.1027/1015-5759.13.1.29

Wright, B. D., & Linacre, J. M. (1989). *Rasch measurement: Transactions of the Rasch Measurement SIG*. MESA Press.

Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis*. MESA Press.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. MESA Press.

Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Zimmerman, B. J., & Martinez-Pons, M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal, 23*(4), 614-628.

Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-type (Ordinal) Item Scores*. Ottawa, Canada: Directorate of Human Resources Research and Evaluation.

Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, *4*(2), 223–233.

https://doi.org/10.36269/psyche.v7i2.3049