

**ANALISIS BUTIR *SITUATIONAL JUDGEMENT TEST* KOMPETENSI INTI KARYAWAN
PT ANGKASA PURA INDONESIA DENGAN *RASCH MODEL******ANALYSIS OF SITUATIONAL JUDGMENT TEST OF CORE COMPETENCIES USING THE
RASCH MODEL ON EMPLOYEES OF BUMN PT ANGKASA PURA INDONESIA*****Chasya Nia Khoerunnisa & Moh. Abdul Hakim**

Fakultas Psikologi, Universitas Sebelas Maret, Surakarta

chasya.nia1106@gmail.com*; m.a.hakim@staff.uns.ac.id

ABSTRAK

Asesmen kompetensi dalam manajemen sumber daya manusia sangat penting untuk memetakan kebutuhan pengembangan pegawai. PT Angkasa Pura Indonesia, salah satu perusahaan sektor transportasi di Indonesia, mengembangkan asesmen kompetensi dalam bentuk tes menggunakan metode *Situational Judgment Test* (SJT). Penelitian ini bertujuan untuk melakukan analisis kualitas butir SJT kompetensi inti dengan model Rasch berdasarkan sampel karyawan PT Angkasa Pura Indonesia. Teknik *purposive sampling* digunakan dengan kriteria pemilihan sampel yaitu karyawan PT Angkasa Pura Indonesia dengan kelompok jabatan level 1 (*officer*) sebanyak 3,989 karyawan. Analisis kualitas butir dilakukan dengan model Rasch pada 27 butir soal yang mengukur tiga kompetensi inti, yaitu *Customer Focus* (CFO), *Driving Innovation* (DIN), dan *Building Strategic Partnership* (BSP). Hasil analisis Rasch menunjukkan keseluruhan butir fit dengan nilai *infit* dan *outfit* dalam rentang (0,5 - 1,5), sedangkan pada *measure* seluruh butir berada dalam rentang *logit* sedang (-2,0 - 2,0). *Wright Map* menunjukkan bahwa tiga kompetensi inti memotret kemampuan pada tingkatan rerata. Namun, butir soal pada kompetensi CFO dan DIN tidak memotret abilitas rendah maupun tinggi, sedangkan pada BSP butir soal tidak memotret abilitas rendah. Penelitian ini menyimpulkan bahwa instrumen SJT ini memiliki kualitas dan keandalan yang baik sebagai alat asesmen kompetensi inti di PT Angkasa Pura Indonesia.

Kata Kunci: analisis butir, kompetensi inti, rasch model, *situational judgment test*, validasi alat tes

ABSTRACT

PT Angkasa Pura Indonesia, a transportation company in Indonesia, has developed a competency assessment in the form of a test using the Situational Judgment Test (SJT) method. The purpose of this study is to analyze the quality of the core competency items of the SJT using the Rasch model based on a sample of PT Angkasa Pura Indonesia employees. The sampling technique employed is purposive sampling, with the criteria for selecting samples being PT Angkasa Pura Indonesia employees holding position level 1 (officer) amounting to 3,989 employees. The quality analysis of the items was conducted using the Rasch model on 27 items that measure 3 core competencies: Customer Focus (CFO), Driving Innovation (DIN), and Building Strategic Partnerships (BSP) for position level 1 (officer). The Rasch analysis results showed that all items fit within the infit and outfit values ranging from 0.5 to 1.5, and all items were within the moderate logit range (-2.0 to 2.0). The Wright Map demonstrated that the 3 core competencies captured ability at the average level. However, the items in the CFO and DIN competencies did not capture low or high abilities, while the BSP items did not capture low abilities. This study concludes that the SJT

instrument has good item quality and reliability and can be used as a core competency assessment tool at PT Angkasa Pura Indonesia.

Keywords: *item analysis, core competencies, Rasch model, situational judgment test, validation of test equipment*

PENDAHULUAN

Asesmen kompetensi merupakan upaya penting untuk memberikan gambaran mengenai kemampuan kinerja karyawan, termasuk kekuatan, kelemahan, keterampilan, dan potensi karyawan, serta untuk merancang program pengembangan yang dapat meningkatkan kualitas dan kuantitas kerja (Mintawati et al., 2023). Asesmen kompetensi biasanya digunakan untuk mengukur kemampuan individu di berbagai sektor, mulai dari sektor kesehatan, pendidikan hingga perusahaan (Junaedi et al., 2023; Nurlina, 2019; Setiawati & Hindriana, 2020). Dalam konteks perusahaan, salah satu penerapannya asesmen kompetensi terdapat di Perusahaan Badan Usaha Milik Negara (BUMN). Penerapan ini sejalan dengan Peraturan Menteri mengenai kewajiban karyawan untuk melaksanakan asesmen kompetensi sesuai dengan (1) Permen BUMN No 11/MBU/07/2021 tentang syarat, tata cara pengangkatan, dan pemberhentian calon anggota Direksi BUMN, (2) Permen BUMN No. Per 04/MBU/10/2019 tentang Kamus Kompetensi ASN di Lingkungan Kementerian BUMN (Kementrian BUMN, 2019, 2021).

Sejalan dengan peraturan menteri tersebut, salah satu perusahaan BUMN yang bergerak di bidang transportasi, PT Angkasa Pura Indonesia (PT API), telah menerapkan kamus kompetensi dalam pengembangan asesmen kompetensi menggunakan metode *assessment center* yang biasanya digunakan untuk keperluan seleksi, promosi dan pengembangan karyawan berbasis kompetensi (Kleinmann & Ingold, 2019). Namun, dalam penerapan *assessment center* di PT API, terdapat beberapa implikasi yang mengindikasikan perlunya peningkatan efektivitas pelaksanaan asesmen. Implikasi ini meliputi biaya operasional yang besar dan kebutuhan jumlah asesor yang tinggi, sementara jumlah karyawan yang perlu dinilai cukup banyak. Oleh karena itu, PT API mengembangkan asesmen kompetensi menggunakan metode *Situational Judgment Test* (SJT) yang meminimalkan keterlibatan asesor.

PT API mengembangkan asesmen kompetensi dengan SJT dengan tiga model kompetensi yaitu *core competency* (kompetensi inti), *leadership competency* (kompetensi kepemimpinan), dan *functional competency* (kompetensi fungsional). Dengan

mengembangkan model asesmen SJT ini, diharapkan kompetensi karyawan dapat diukur melalui respon karyawan terhadap situasi pekerjaan yang relevan (Lievens & Motowidlo, 2015). Asesmen kompetensi berbasis SJT yang dikembangkan PT API telah melalui uji validitas isi yang dilakukan oleh ahli psikometri dan asesor SDM (Hakim et al., 2024). Namun, masih diperlukan penelitian lebih lanjut untuk mengevaluasi kualitas psikometris butir-butir tes tersebut untuk mengoptimalkan keandalannya (Courville, 2004). Oleh karena itu, perlu dilakukan analisis kualitas butir dengan cara merevisi atau menghapus soal yang tidak efektif, serta memberikan informasi diagnostik mengenai pemahaman karyawan terhadap kompetensi yang diuji (Aiken, 1994).

Dalam analisis kualitas butir, terdapat beberapa pendekatan yang sering digunakan, yaitu teori tes klasik dan teori respon butir (*item response theory*). Kedua pendekatan ini telah digunakan dalam berbagai penelitian untuk menganalisis kualitas butir soal (Jumini et al., 2023; Pratama, 2020). Namun, untuk analisis butir pada asesmen kompetensi di PT API, pendekatan teori respon butir (*item response theory*) akan memberikan hasil yang lebih optimal karena IRT memiliki hasil analisis yang lebih akurat apabila diterapkan pada sampel yang besar (Siregar & Panjaitan, 2022). Teori respon butir (*item response theory*) menganalisis hubungan antara respon individu dan karakteristik item dengan Rasch Model sebagai salah satu modelnya yang mengklasifikasikan item dan individu dalam peta distribusi untuk menilai kemampuan dan kesukaran item secara akurat (Azizah & Wahyuningsih, 2020). *Rasch Model* menjadi pilihan yang tepat karena bermanfaat untuk mengukur tingkat kesukaran item, kualitas instrumen, dan tingkat kemampuan individu secara bersamaan (Rahman et al., 2022).

Penelitian sebelumnya oleh Kusuma & Hakim (2024) telah melakukan analisis butir menggunakan *Rasch Model* pada asesmen kompetensi berbasis SJT di PT API, khususnya untuk model kepemimpinan pada level jabatan 1 (*officer*). Hasil penelitian tersebut membuktikan bahwa alat tes kompetensi kepemimpinan memiliki kualitas dan keandalan butir yang dapat menjadi instrumen asesmen berkelanjutan dengan rekomendasi perbaikan pada butir tertentu. Meskipun penelitian tersebut telah memberikan kontribusi dalam pengembangan asesmen kompetensi di PT API, masih terdapat model kompetensi lain yang belum di analisis kualitas butirnya seperti *core competency* (kompetensi inti) dan *functional competency* (kompetensi fungsional).

Berdasarkan kamus kompetensi PT API, terdapat tiga kompetensi inti yang menjadi tujuan strategis organisasi yaitu *customer focus (service orientation)*, *building strategic partnerships*, dan *driving innovation (creativity)*. Oleh karena itu, analisis butir terhadap kompetensi inti ini menjadi penting untuk memastikan kontribusi optimal karyawan terhadap kinerja organisasi. Hal ini didukung oleh penelitian Liang et al. (2013) yang menunjukkan korelasi positif antara kompetensi inti dan kinerja organisasi. Mempertimbangkan pentingnya level jabatan 1 (*officer*) sebagai pondasi awal karir, penelitian ini fokus pada analisis butir tes kompetensi inti untuk level tersebut. Hasil analisis ini dapat menjadi dasar yang kuat untuk merancang program pengembangan karyawan yang efektif dan selaras dengan kebutuhan strategis perusahaan.

Berbagai penelitian terdahulu telah menerapkan *Rasch Model* dalam menganalisis kualitas butir soal (Aqida, 2024; Sugitri & Fitri, 2024; Widyaningsih et al., 2023). Lebih lanjut, penggunaan *Rasch Model* dalam menganalisis butir *Situational Judgment Test* (SJT) juga telah dilakukan sebelumnya baik dalam sektor organisasi maupun non-organisasi sebagai instrumen evaluasi kompetensi bagi tenaga pengajar dan tenaga kesehatan (Chao et al., 2019; Heininger et al., 2021). Sementara itu, dalam lingkup BUMN, Kusuma & Hakim (2024) telah menerapkan *Rasch Model* untuk menganalisis kualitas butir SJT pada kompetensi kepemimpinan. Namun, belum ada penelitian yang secara khusus menganalisis kompetensi inti. Oleh karena itu, penelitian ini bertujuan untuk menganalisis kualitas butir SJT instrumen pengukuran kompetensi inti level jabatan 1 (*officer*) dengan *Rasch Model* pada karyawan PT Angkasa Pura Indonesia.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan data sekunder yang diperoleh melalui hasil asesmen kompetensi yang telah dilakukan oleh PT API. Pengumpulan data dalam penelitian ini diperoleh melalui asesmen kompetensi yang dilakukan oleh PT API menggunakan sistem *Computer Based Test* (CBT) yang dikembangkan oleh PT API. Pengumpulan data dilakukan mulai tanggal 26 Februari 2024 hingga 4 Juni 2024, dengan cakupan wilayah meliputi kantor pusat dan 20 kantor cabang yang tersebar di seluruh Indonesia dengan persebaran distribusi responden berdasarkan wilayah meliputi provinsi Aceh (3%), Banten (4%), DKI Jakarta (54,7%), Jambi (2,7%), Jawa Barat (4,4%), Jawa Tengah (0,2%), Jawa Timur (1,1%), Kalimantan Barat (3,4%),

Kalimantan Tengah (1,1%), Kepulauan Bangka Belitung (3,0%), Lampung (1,2%), Riau (5,3%), Sumatera Barat (3,4%), Sumatera Selatan (3,5%), dan Sumatera Utara (9,0%).

Populasi merupakan keseluruhan objek atau subjek yang memiliki karakteristik dan kualitas tertentu yang ditetapkan peneliti untuk diteliti dan dijadikan dasar penarikan kesimpulan (Sugiyono, 2019). Populasi dalam penelitian ini adalah karyawan PT Angkasa Pura Indonesia. Sampel merupakan bagian dari populasi yang memiliki karakteristik tertentu yang mewakili suatu populasi (Sugiyono, 2019). Metode pengambilan sampel dalam penelitian ini diambil berdasarkan *non probability sampling* dengan teknik *purposive sampling*, dimana pemilihan sampel dilakukan berdasarkan kriteria atau pertimbangan khusus yang telah ditentukan oleh peneliti (Sugiyono, 2019). Kriteria pemilihan sampel yaitu karyawan PT API yang merupakan kelompok jabatan level 1 (*officer*). Oleh karena itu, jumlah sampel yang digunakan sebanyak 3.989 orang.

Teknik analisis data dalam penelitian ini dilakukan dengan menganalisis butir masing-masing kompetensi inti menggunakan metode *Rasch Model Unidimensional*, dengan bantuan aplikasi Jamovi 2.5.5. Analisis ini menghasilkan berbagai parameter, yaitu skor *Item Mean* (rata-rata butir), *Measure* (tingkat kesukaran), *Standard Error (SE) Measure*, nilai *Outfit & Infit*, dan *Wright Map*. Selain itu, uji *Martin-Löf* dilakukan guna memastikan bahwa instrumen yang digunakan hanya mengukur satu konstruk utama (unidimensi) (Smith, 2024). Uji ini bertujuan untuk memverifikasi asumsi unidimensionalitas dalam model Rasch.

Penelitian ini menggunakan instrumen SJT yang dikembangkan oleh PT API untuk mengukur 3 kompetensi inti pada karyawan PT API. Instrumen tersebut terdiri dari 27 butir yang mengukur 3 kompetensi inti pada level kompetensi 1. *Blueprint* kompetensi inti dapat dilihat pada tabel 1.

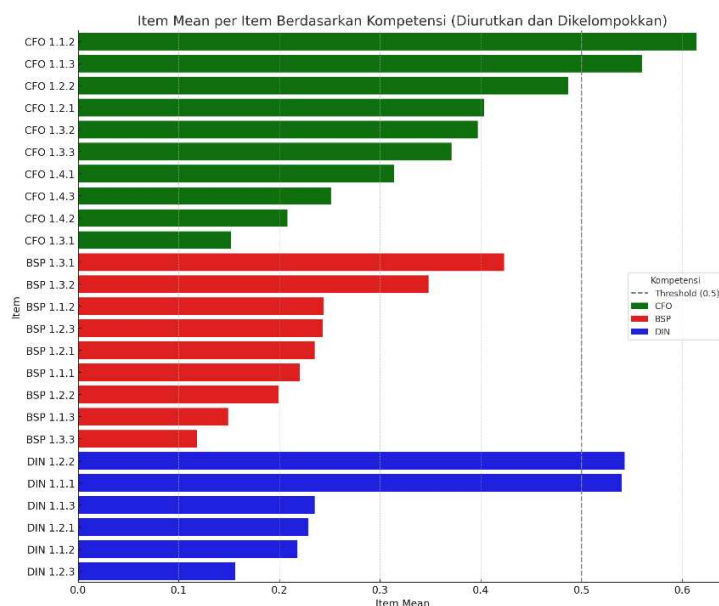
Tabel 1. *Blueprint Kompetensi Inti*

<i>Kode Aitem/Butir</i>	<i>N</i>
<i>Customer Focus (Service Orientation) - CFO</i>	12
CFO 1.1.1, CFO 1.1.2, CFO 1.1.3, CFO 1.2.1, CFO 1.2.2, CFO 1.2.3, CFO 1.3.1, CFO 1.3.2, CFO 1.3.3 CFO 1.4.1, CFO 1.4.2, CFO 1.4.3	
<i>Building Strategic Partnership - BSP</i>	9
BSP 1.1.1, BSP 1.1.2, BSP 1.1.3 BSP 1.2.1, BSP 1.2.2, BSP 1.2.3 BSP 1.3.1, BSP 1.3.2, BSP 1.3.3	
<i>Customer Focus (Service Orientation) - CFO</i>	6
DIN 1.1.1, DIN 1.1.2, DIN 1.1.3 DIN 1.2.1, DIN 1.2.2, DIN 1.2.3	
TOTAL	27

HASIL PENELITIAN

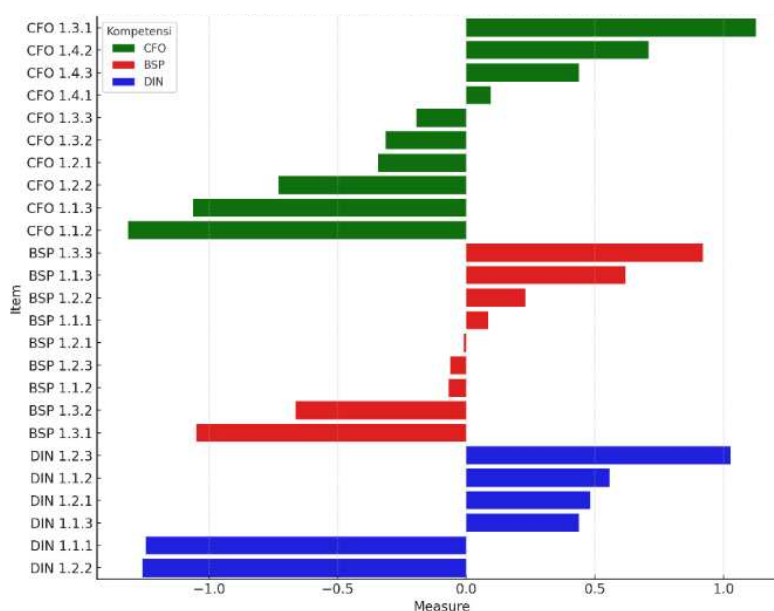
1. *Item Mean, Measure, SE Measure*

Gambar 1 disusun berdasarkan urutan *Measure* (tingkat kesukaran) setiap *item*, mulai dari yang paling sukar hingga yang paling mudah. *Item Mean* memberikan gambaran nilai rata-rata bagi setiap butir (Kusuma & Hakim, 2024). Analisis pada alat tes kompetensi ini dilakukan dengan format dikotomi (dengan jawaban hanya benar atau salah), yang artinya secara umum skor rata-rata butir (*item mean*) sebesar 0.5. Hal ini mengindikasikan bahwa 50% responden memberikan jawaban benar dan 50% responden memberikan jawaban salah. Artinya, apabila *item mean* < 0.5, maka lebih dari 50% responden memberikan jawaban salah pada butir terkait (nilai 0), sedangkan apabila *item mean* > 0.5, maka lebih dari 50% responden memberikan jawaban benar pada butir terkait (nilai 1). Berdasarkan hasil analisis pada tabel 6, terdapat 23 butir memiliki *item mean* < 0,5 dan 4 butir memiliki *item mean* > 0,5.



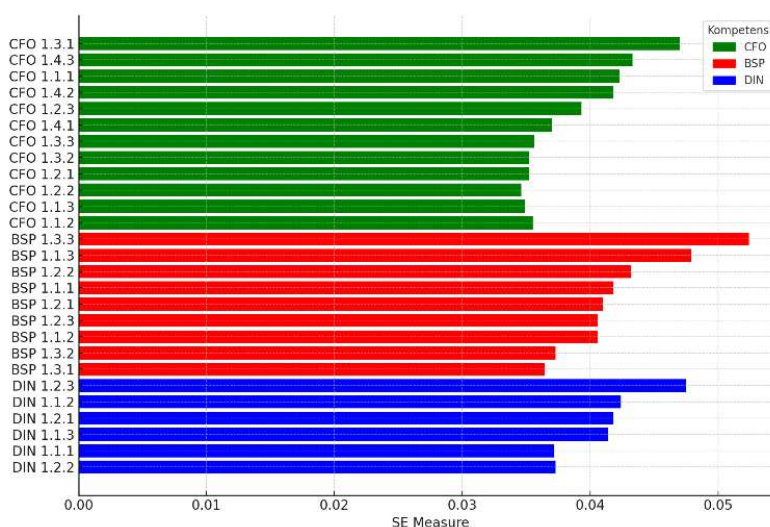
Gambar 1. Sebaran Item Mean per Item SJT

Measure memberikan gambaran tentang kesukaran bagi setiap butir (Sumintono & Widhiarso, 2015). Berdasarkan hasil analisis, seluruh butir soal berada pada tingkat kesukaran butir sedang, dengan rentang $-2,0 \text{ logit}$ hingga $2,0 \text{ logit}$ (Hadiyanti et al., 2024). Pada kompetensi CFO, nilai *logit* butir soal berada di rentang $-1,3146 - 1,1246$. Dari seluruh butir soal, CFO 1.3.1 memiliki tingkat kesukaran tertinggi dengan nilai *logit* sebesar $1,1246$, sedangkan CFO 1.1.2 memiliki tingkat kesukaran paling rendah dengan nilai *logit* sebesar $-1,3146$. Pada kompetensi BSP, nilai *logit* butir soal berada di rentang $-1,04907 - 0,92048$. Dari seluruh butir soal, BSP 1.3.3 memiliki tingkat kesukaran tertinggi dengan nilai *logit* sebesar $0,92048$, sedangkan BSP 1.3.1 memiliki tingkat kesukaran paling rendah dengan nilai *logit* sebesar $-1,04907$. Pada kompetensi DIN, nilai *logit* butir soal berada di rentang $-1,258 - 1,028$. Dari seluruh butir soal, DIN 1.2.3 memiliki tingkat kesukaran tertinggi dengan nilai *logit* sebesar $1,028$, sedangkan DIN 1.2.2 memiliki tingkat kesukaran paling rendah dengan nilai *logit* sebesar $-1,258$.



Gambar 2. Sebaran Measure per Item SJT

Nilai *SE Measure* memberikan gambaran evaluasi mengenai stabilitas estimasi kemampuan individu dalam satuan *logit*, di mana semakin besar nilai *standard error*, maka semakin kurang akurat estimasi kemampuannya (Katz et al., 2021). Berdasarkan hasil analisis, nilai *SE Measure* pada seluruh kompetensi inti berada dalam rentang 0.0346 - 0.0524 dengan nilai *SE Measure* terendah adalah CFO 1.2.2 sebesar 0.0346, yang menunjukkan stabilitas pengukuran paling akurat dibandingkan aitem lainnya. Sedangkan, nilai *SE Measure* tertinggi adalah BSP 1.3.3 sebesar 0.0524, yang mengindikasikan stabilitas pengukurannya kurang akurat dibandingkan aitem lainnya.

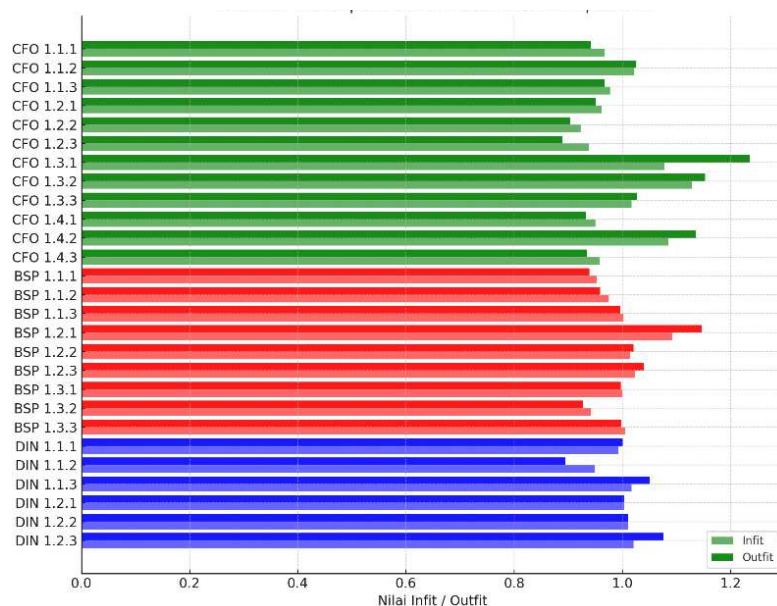


Gambar 3. Sebaran SE Measure per Item SJT

2. Item Fitness

Nilai *infit* menunjukkan seberapa baik butir soal sesuai dengan pola keseluruhan tes, sedangkan nilai *outfit* mengukur hubungan butir soal dengan kemampuan yang diukur dan lebih sensitif terhadap respons yang tidak terduga atau *outlier* (Aqida & Khairunisa, 2019). Berdasarkan hasil analisis, seluruh butir soal pada kompetensi inti menunjukkan nilai *infit* dan *outfit* yang dapat dikatakan *fit* dengan nilai dalam rentang 0.5 - 1.5 (Andrich & Marais, 2019).

Pada kompetensi *Customer Focus* (CFO), rentang skor *infit* adalah 0.923 - 1.129, sedangkan skor *outfit* berada pada rentang 0.889 - 1.236. Kompetensi *Building Strategic Partnership* (BSP) memiliki rentang skor *infit* sebesar 0.942 - 1.092 dan rentang skor *outfit* sebesar 0.927 - 1.146. Terakhir, pada kompetensi *Driving Innovation* (DIN), rentang skor *infit* sebesar 0.949 - 1.021, sedangkan skor *outfit* berada di rentang 0.894 - 1.076.

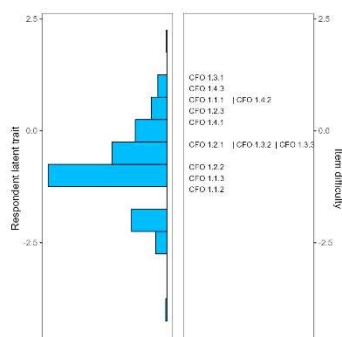


Gambar 4. Sebaran Item Fitness per Item SJT

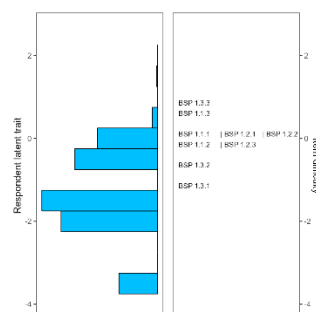
3. Persebaran *Wright Map*

Wright Map adalah representasi grafis dua dimensi yang digunakan untuk menunjukkan hubungan antara kemampuan responden dan tingkat kesulitan butir (Green & Frantom, 2002). Area kanan pada *Wright Map* menunjukkan kemampuan individu, dengan kemampuan lebih rendah berada di bawah dan yang lebih tinggi di atas, sedangkan area kiri menunjukkan tingkat kesukaran soal, dengan soal yang lebih sulit berada di atas dan yang lebih mudah di bawah (Sukarelawan et al., 2024). Dalam analisis butir kompetensi inti, butir-butir dengan tingkat kesukaran tinggi terlihat pada area kanan atas, sedangkan butir dengan tingkat kesukaran rendah berada di bagian kanan bawah. Nilai pada area kanan menggambarkan visualisasi distribusi dari *item measure* yang menyajikan informasi secara kuantitatif. Sementara itu, kelompok responden (karyawan) dengan abilitas atau kompetensi tinggi terlihat pada area kiri atas, sedangkan pada bagian kiri bawah merupakan kelompok responden (karyawan) dengan abilitas atau kompetensi rendah.

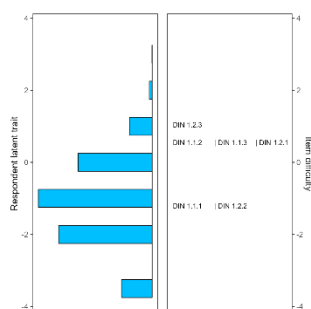
Berdasarkan distribusi *Wright Map*, area kanan pada kompetensi *Customer Focus* (CFO), *Building Strategic Partnership* (BSP) dan *Driving Innovation* (DIN) menunjukkan bahwa tidak terdapat butir soal yang lebih tinggi dari nilai logit (2) atau lebih rendah dari nilai logit (-2), sehingga seluruh butir soal pada kompetensi tersebut memiliki tingkat kesukaran yang sedang. *Wright map* ini menunjukkan bahwa persebaran butir soal pada kompetensi seluruh kompetensi CFO dapat memotret abilitas rata-rata. Namun, tidak ada butir soal yang dapat memotret abilitas rendah maupun tinggi. Sedangkan seluruh kompetensi BSP dapat memotret abilitas rata-rata. Namun, tidak ada butir soal yang dapat memotret abilitas rendah. *Wright map* ini menunjukkan bahwa persebaran butir soal pada seluruh kompetensi DIN dapat memotret abilitas rata-rata. Namun, tidak ada butir soal yang dapat memotret abilitas rendah maupun tinggi.



Gambar 5. Wright Map Kompetensi CFO



Gambar 6. Wright Map Kompetensi BSP



Gambar 7. Wright Map Kompetensi DIN

4. Uji Unidimensi (*Martin-Löf*)

Uji unidimensionalitas (*Martin-Löf*) dilakukan dengan melihat nilai signifikansi (*p-value*), di mana $p < 0,05$ mengindikasikan adanya multidimensionalitas (Horton et al., 2012). Berdasarkan hasil uji unidimensionalitas yang disajikan pada tabel 6, diperoleh bahwa kompetensi *Customer Focus* (CFO) dan *Driving Innovation* (DIN) menunjukkan unidimensionalitas yang baik dengan nilai p untuk CFO sebesar 0.085 ($p > 0.05$) dan DIN sebesar 0.079 ($p > 0.05$). Sementara kompetensi *Building Strategic Partnership* (BSP) menunjukkan kecenderungan multidimensionalitas dengan nilai p sebesar 0.004 ($p < 0.05$). Tabel hasil uji unidimensionalitas per kompetensi dapat dilihat pada tabel 2.

Tabel 2. Hasil Uji Unidimensionalitas Per Kompetensi

Kompetensi	Value	df	p
<i>Customer Focus</i> (CFO)	25.7	35	0.875
<i>Building Strategic Partnership</i> (BSP)	39.6	19	0.004
<i>Driving Innovation</i> (DIN)	1.58	7	0.979

DISKUSI

Rata-rata butir (*item mean*) secara umum adalah 0.5 yang artinya setengah responden menjawab benar dan setengahnya menjawab salah. Hasil analisis menunjukkan 23 butir dijawab salah oleh sebagian besar responden, sementara 4 butir lainnya dijawab benar oleh sebagian besar responden, mengindikasikan adanya butir yang terlalu mudah atau terlalu sulit. Oleh karena itu, analisis selanjutnya berfokus pada pendalaman tingkat kesukaran setiap butir soal (*measure*) menggunakan nilai *logit*.

Hasil analisis menunjukkan bahwa butir soal secara keseluruhan berada pada tingkat kesulitan yang baik, yakni dalam kategori sedang (Hadiyanti et al., 2024). Butir soal yang diterima adalah yang memiliki tingkat kesulitan sesuai dengan rentang yang ditentukan (Nuryanti et al., 2018). Selanjutnya, stabilitas estimasi kemampuan individu dievaluasi melalui nilai *Standard Error (SE) Measure*. Semakin besar nilai SE, semakin kurang akurat estimasi kemampuan responden (Katz et al., 2021). Meskipun sebagian besar butir soal mengukur kompetensi dengan presisi, butir soal dengan *SE Measure* tertinggi (BSP 1.3.3) menunjukkan estimasi kemampuan yang kurang akurat jika dibandingkan dengan butir soal lainnya, sehingga perlu evaluasi lebih lanjut.

Item fitness digunakan untuk mengevaluasi kecocokan setiap butir soal. Hasil analisis menunjukkan bahwa seluruh butir soal berada dalam rentang *infit* dan *outfit* yang baik (0.5 - 1.5), serta tidak ada butir soal yang *misfit* (Andrich & Marais, 2019). Oleh karena itu, meskipun terdapat indikasi awal bahwa terdapat butir soal yang mungkin kurang andal berdasarkan hasil pengukuran *SE Measure*, hasil evaluasi *item fitness* menunjukkan bahwa instrumen ini valid dan andal untuk mengukur kompetensi (Chasanah et al., 2024). Tidak ditemukannya *item misfit* dalam instrumen ini mengonfirmasi bahwa setiap butir soal sesuai untuk mengukur kompetensi inti karyawan PT API (Napitupulu, 2018). Hal ini menunjukkan bahwa semua soal berfungsi

sesuai tujuan pengukuran. Soojung et al., (2018) berpendapat bahwa apabila terdapat butir soal yang tidak sesuai (*misfit*), maka aitem tersebut mengukur konstruk yang berada di luar tujuan pengukuran instrumen sehingga perlu dilakukan revisi atau eliminasi. Selain itu, butir soal *misfit* juga dapat menurunkan keandalan tes (Zubairi & Kassim, 2006). Dengan tidak adanya item misfit, instrumen ini memenuhi standar kualitas yang diperlukan dan tidak memerlukan perubahan lebih lanjut.

Berdasarkan hasil analisis *wright map*, yang menunjukkan sejauh mana kemampuan responden (karyawan) dengan tingkat kesulitan soal pada skala *logit* yang sama, diperoleh seluruh butir soal pada 3 kompetensi inti yaitu *Customer Focus* (CFO), *Building Strategic Partnership* (BSP), dan *Driving Innovation* (DIN) hanya mampu mengukur kemampuan responden pada tingkat kesukaran yang sedang (rata-rata). Sementara itu, pada kompetensi BSP, tidak ada butir soal yang dapat mengukur kemampuan responden dengan abilitas rendah, sedangkan pada kompetensi CFO dan DIN, tidak ada butir soal yang dapat mengukur kemampuan peserta dengan abilitas rendah maupun tinggi. Hal ini menunjukkan bahwa instrumen tes ini kurang mampu menangkap kemampuan peserta yang ekstrem, baik yang sangat rendah maupun sangat tinggi. Temuan ini sejalan dengan penelitian Novriyanti & Arthur (2024), yang menyarankan untuk menambahkan butir soal pada tingkat kesulitan yang sangat mudah dan sangat sulit, sehingga pengukuran instrumen tes dapat lebih efektif dalam mengukur semua tingkat kemampuan peserta.

Namun, dalam konteks pengukuran kompetensi inti pada alat ukur PT API, penambahan butir soal yang efektif sebaiknya difokuskan pada butir soal yang memiliki tingkat kesulitan tinggi. Hal ini relevan untuk mengukur kemampuan peserta dengan abilitas tinggi, sehingga dapat memenuhi standar kompetensi yang dibutuhkan PT API. Temuan ini sejalan dengan penelitian oleh Karst et al., (2022), yang menyoroti pentingnya keberadaan soal-soal sulit untuk membedakan responden dengan kemampuan tinggi yang mempermudah identifikasi individu dengan potensi terbaik.

Penambahan butir soal yang sulit dapat dilakukan dengan melakukan analisis mendalam terhadap kompetensi yang ingin diukur, serta mengintegrasikannya dengan situasi nyata yang relevan dalam penerapannya di PT API, sehingga membutuhkan keterlibatan para ahli/praktisi dalam perusahaan untuk merancang soal yang menantang dan nyata. Selain itu, alternatif lain yang dapat dilakukan adalah melakukan revisi dengan

memasukkan soal isomorfik, di mana soal dirancang untuk memiliki struktur dan format yang sama tetapi dengan konteks atau detail yang berbeda (Lievens & Sackett, 2007).

Sebagai contoh, revisi pada butir soal CFO 1.1.1 (Butir soal terlampir pada lampiran), dapat dilakukan dengan menambahkan kalimat “...*Namun, Anda juga menyadari bahwa jika Anda tidak memenuhi permintaan Bapak Andi, dia mungkin akan mencari vendor lain yang menawarkan harga lebih baik*”.... Dengan penambahan situasi ini, dilema muncul, sehingga responden harus mempertimbangkan konsekuensi baik untuk menjaga hubungan berkelanjutan dengan pelanggan ataupun untuk menjaga kestabilan finansial perusahaan. Penambahan ini membuat butir soal menjadi lebih kompleks sehingga diharapkan mampu meningkatkan kesulitan soal.

Selain itu penerapan Rasch Model dalam analisis ini memungkinkan evaluasi terhadap asumsi unidimensionalitas dan independensi lokal yang merupakan aspek penting dalam validitas instrumen (Sumintono & Widhiarso, 2014). Analisis terhadap asumsi unidimensionalitas menunjukkan bahwa kompetensi CFO dan DIN menunjukkan unidimensionalitas yang baik, sedangkan kompetensi BSP cenderung multidimensional, dengan nilai $p < 0.05$ (Horton et al., 2012). Hal ini menunjukkan bahwa kompetensi BSP melibatkan berbagai aspek perilaku yang relevan meskipun tidak terfokus pada satu faktor laten tunggal (Kusuma & Hakim, 2024). Penelitian ini menggarisbawahi bahwa unidimensionalitas bukanlah keharusan dalam pengukuran kompetensi, karena fokus utama adalah pada keberadaan perilaku kunci yang relevan dalam tugas pekerjaan (Kusuma & Hakim, 2024; Taylor, 2007).

Hasil menunjukkan bahwa kompetensi dengan unidimensionalitas baik, seperti CFO dan DIN, memiliki kecenderungan untuk memenuhi kriteria independensi lokal, sesuai dengan pendapat Retnawati (2014) yang menyatakan bahwa jika asumsi unidimensionalitas terpenuhi, maka asumsi independensi lokal juga cenderung terpenuhi. Sebaliknya, pada kompetensi BSP yang tidak unidimensional, independensi lokal tidak dapat dipastikan. Temuan ini dapat menjadi bahan evaluasi bagi PT API untuk memastikan apakah perilaku-perilaku kunci (*keys behavior*) yang disusun dalam kamus kompetensi bersifat unidimensional (faktor laten tunggal) atau multidimensional.

Secara keseluruhan, hasil penelitian menunjukkan bahwa instrumen SJT dapat diterapkan secara berkelanjutan sebagai alat ukur kompetensi inti di PT API. Instrumen ini menunjukkan validitas dan reliabilitas yang baik, serta dapat mengevaluasi

kompetensi dengan akurat dan efisien (Ida & Musyarofah, 2021; Rost & Davier, 1994). Meskipun demikian, hasil analisis *Wright map* menunjukkan perlunya penambahan butir soal dengan tingkat kesulitan yang sangat sulit, untuk mengukur spektrum kemampuan responden secara lebih luas. Penelitian ini memberikan beberapa rekomendasi untuk penelitian lebih lanjut, seperti penambahan butir soal yang lebih beragam tingkat kesulitannya dan penerapan pendekatan *multidimensional* untuk alat tes yang bersifat multidimensi. Penelitian ini juga mencatat keterbatasan dalam hanya menganalisis kompetensi pada level 1, dan tidak mencakup evaluasi parameter psikometris lainnya yang juga dapat mempengaruhi validitas instrumen. Dengan keterbatasan-keterbatasan tersebut, diharapkan penelitian ini dapat menjadi landasan bagi penelitian lanjutan yang lebih mendalam serta memperkuat kontribusi penelitian terhadap pengembangan asesmen kompetensi karyawan di PT API dan bidang psikometri secara umum.

KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa butir-butir soal dalam instrumen *Situational Judgment Test* (SJT) kompetensi inti karyawan PT Angkasa Pura Indonesia memiliki kualitas dan keandalan butir yang baik dilihat dari tingkat kesukaran butir (*item measure*) dan kesesuaian butir (*item fitness*). Tidak ditemukan butir yang gugur berdasarkan kriteria analisis *Rasch Model*, yang menunjukkan bahwa instrumen ini secara psikometris layak digunakan untuk asesmen kompetensi inti di PT API. Meskipun demikian, penambahan atau revisi butir dengan tingkat kesukaran yang lebih tinggi perlu dipertimbangkan, sehingga cakupan pengukuran kompetensi dapat mengukur kompetensi karyawan dengan kemampuan/abilitas yang tinggi.

Dari sisi penerapan, penelitian ini memberikan kontribusi praktis bagi pengembangan asesmen di lingkungan BUMN. Selain itu, secara teoritis, penelitian ini juga memperkaya literatur mengenai penerapan *Rasch Model* dalam pengembangan alat tes berbasis metode SJT di ranah industri, khususnya dalam lingkup BUMN yang belum banyak dieksplorasi. Temuan ini menunjukkan bahwa *Rasch Model* efektif dalam mengevaluasi kualitas butir alat tes SJT.

Penelitian selanjutnya disarankan untuk melakukan analisis pada instrumen level jabatan 2 (kompetensi manajerial) agar dapat memastikan butir-butir soal pada level jabatan yang lebih tinggi memiliki kualitas yang baik serta dapat memberikan wawasan yang lebih komprehensif mengenai kemampuan kompetensi karyawan pada level

tersebut. Selain itu, Penelitian selanjutnya juga dapat menggunakan pendekatan *multidimensional* seperti *Multidimensional Item Response Theory* (MIRT), yang lebih tepat diterapkan pada instrumen dengan karakteristik multidimensi (Embretson & Reise, 2000). Pengembangan analisis juga dapat diperluas dengan mempertimbangkan penggunaan parameter psikometrik tambahan, seperti model 2PL (*Two-Parameter Logistic Model*) yang memungkinkan pengukuran terhadap tingkat kesukaran dan daya diskriminasi item serta model 3 PL (*Three-Parameter Logistic Model*) dengan pengukuran terhadap tingkat kesukaran, daya diskriminasi, dan tebakan semu (*pseudo-guessing*) yang dapat memengaruhi hasil respons peserta terhadap item tertentu (Amelia & Kriswantoro, 2017).

PENDANAAN PENELITIAN

Penelitian ini didanai oleh RKAT Universitas Sebelas Maret Tahun Anggaran 2025 melalui skema Penelitian PENGUATAN KAPASITAS GRUP RISET (PKGR-UNS) C dengan Nomor Perjanjian Penugasan Penelitian: 371/UN27.22/PT.01.03/2025.

DAFTAR PUSTAKA

- Aiken, L. R. (1994). *Psychological Testing and Assessment (8th ed.)*. Allyn And Bacon.
- Amelia, R. N., & Kriswantoro, K. (2017). Implementation of Item Response Theory for Analysis of Test Items Quality and Students' Ability in Chemistry. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, 2(1), 1. <https://doi.org/10.20961/jkpk.v2i1.8512>
- Andrich, D., & Marais, I. (2019). *A Course in Rasch Measurement Theory*. Springer Nature Singapore. <https://doi.org/10.1007/978-981-13-7496-8>
- Aqida, D. S. (2024). Penggunaan Rasch Model dalam Menganalisis Butir Soal Matematika Kelas V Sekolah Dasar. *Indonesian Journal of Islamic Elementary Education*, 4(1), 91–101. <https://doi.org/10.28918/ijjee.v4i1.6947>
- Aqida, D. S., & Khairunisa, D. (2019). The Use Of Rasch Model In Analyzing Grade V Elementary School Mathematics Items. *Sunan Kalijaga International Journal on Islamic Educational Research*, 7(1), 58–68. <https://ejournal.uin-suka.ac.id/tarbiyah/SKIJIER/article/download/7589/3008/24555>
- Azizah, A., & Wahyuningsih, S. (2020). Penggunaan Model Rasch Untuk Analisis Instrumen Tes Pada Mata Kuliah Matematika Aktuaria. *JUPITEK: Jurnal Pendidikan Matematika*, 3(1), 45–50. <https://doi.org/10.30598/jupitekvol3iss1pp45-50>
- Chao, T.-Y., Sung, Y.-T., & Huang, J.-L. (2019). Construction of the Situational Judgment Tests for Teachers. *Asia-Pacific Journal of Teacher Education*, 48(4), 355–374. <https://doi.org/10.1080/1359866x.2019.1633621>
- Chasanah, N., Susongko, P., & Suriswo, S. (2024). Pengembangan Penilaian untuk Mengukur Kemampuan Literasi Membaca Peserta Didik SMP dengan Standar PISA 2018. *Journal of Education Research*, 5(3), 3816–3826.

- <https://doi.org/10.37985/jer.v5i3.1472>
- Courville, T. (2004). *An Empirical Comparison of Item Response Theory and Classical Test Theory*. *Item/Person Statistics a Dissertation*.
<https://oaktrust.library.tamu.edu/bitstream/handle/1969.1/1064/etd-tamu-2004B-EPSY-Courville-2.pdf>
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. L. Erlbaum Associates.
- Green, K. E., & Frantom, C. G. (2002). *Survey Development and Validation with The Rasch Model*. International Conference on Questionnaire Development, Evaluation, and Testing, Charleston, SC.
- Hadiyanti, H., Susongko, P., & Munnadi, M. (2024). Pengembangan Instrumen Higher Order Thinking Skill Mata Pelajaran Matematika dengan Rasch Model. *Journal of Education Research*, 5(1), 399–407. <https://doi.org/10.37985/jer.v5i1.765>
- Hakim, M. A., Kusuma, N. K. L., Prasetyani, R. N. N., Pradana, R. G., Anggrani, F. K., & Saniatuzzulda, R. (2024). *Laporan final pekerjaan: Pengembangan alat ukur dan sistem pendukung asesmen kompetensi PT API*.
- Heininger, S. K., Baumgartner, M., Zehner, F., Burgkart, R., Söllner, N., Berberat, P. O., & Gartmeier, M. (2021). Measuring Hygiene Competence: the Picture-Based Situational Judgement Test HygiKo. *BMC Medical Education*, 21(410). <https://doi.org/10.1186/s12909-021-02829-y>
- Horton, M., Marais, I., & Christensen, K. B. (2012). Dimensionality. In *Rasch models in health* (pp. 137–158). <https://doi.org/10.1002/9781118574454.ch9>
- Ida, F. F., & Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'arrib: Journal of Arabic Education*, 1(1), 34–44. <https://doi.org/10.32923/al-muarrib.v1i1.2100>
- Jumini, S., Madnasri, S., Cahyono, E., & Parmin, P. (2023). Analisis Kualitas Butir Soal Pengukuran Literasi Sains Melalui Teori Tes Klasik Dan Rasch Model. In *Prosiding Seminar Nasional Pascasarjana*, 6(1), 758–765.
- Junaedi, A. T., Sudarno, S., Suhardjo, S., Wijaya, E., & Ng, M. (2023). Asesmen Kompetensi Karyawan Bank Perkreditan Rakyat Kabupaten Rokan Hulu, Provinsi Riau. *Jurnal Pengabdian Masyarakat Multidisiplin*, 7(1), 31–39. <https://doi.org/https://doi.org/10.36341/jpm.v7i1.3442>
- Karst, K., Bonefeld, M., Dotzel, S., Fehringer, B. C. O. F., & Steinwascher, M. (2022). Data-Based Differentiated Instruction: The Impact of Standardized Assessment and Aligned Teaching Material on Students' Reading Comprehension. *Learning and Instruction*, 79. <https://doi.org/10.1016/j.learninstruc.2022.101597>
- Katz, D., Clairmont, A., & Wilton, M. (2021). Chapter 3 the Rasch Model | Measuring what Matters: Introduction to Rasch Analysis in R. In *Bookdown*. https://bookdown.org/chua/new_rasch_demo2/viewdifficulty.html
- Kementrian BUMN. (2019). *Kamus Kompetensi ASN di Lingkungan Kementerian BUMN*. https://jdih.bumn.go.id/storage/peraturan/PER_04_MBU_10_2019.pdf
- Kementrian BUMN. (2021). *Permen BUMN no. PER-11/MBU/07/2021 Tahun 2021*. <https://peraturan.bpk.go.id/Details/181405/permen-bumn-no-per-11mbu072021-tahun-2021>
- Kleinmann, M., & Ingold, P. V. (2019). Toward a Better Understanding of Assessment Centers: a Conceptual Review. *Annual Review of Organizational Psychology and Organizational Behavior*, 6(1), 349–372. <https://doi.org/10.1146/annurev-orgpsych-012218-014955>

- Kusuma, N. K. L., & Hakim, M. A. (2024). Analisis Butir Situational Judgement Test Kompetensi Kepemimpinan Karyawan BUMN dengan Rasch Model. *JPPP - Jurnal Penelitian Dan Pengukuran Psikologi*, 13(2), 127–137. <https://doi.org/10.21009/jppp.132.06>
- Liang, C.-J., Lin, Y.-L., & Huang, H.-F. (2013). Effect of Core Competence on Organizational Performance in an Airport Shopping Center. *Journal of Air Transport Management*, 31, 23–26. <https://doi.org/10.1016/j.jairtraman.2012.11.005>
- Lievens, F., & Motowidlo, S. J. (2015). Situational Judgment Tests: from Measures of Situational Judgment to Measures of General Domain Knowledge. *Industrial and Organizational Psychology*, 9(1), 3–22. <https://doi.org/10.1017/iop.2015.71>
- Lievens, F., & Sackett, P. R. (2007). Situational Judgment Tests in High-Stakes Settings: Issues and Strategies with Generating Alternate Forms. *Journal of Applied Psychology*, 92(4), 1043–1055. <https://doi.org/10.1037/0021-9010.92.4.1043>
- Mintawati, H., Albert, J., Riyana, R., Safitri, A. N., & Melanda, I. (2023). Strategi Pengembangan Sumber Daya Manusia Melalui Asesmen Kompetensi Karyawan PT PLN Indonesia Power Palabuhan Ratu PGU. *PRINCIPAL: Journal of Managerial, Leadership, Supervision and Human Resources*, 1(2).
- Napitupulu, D. (2018). Analisis Model Rasch pada Instrumen Keberterimaan Marka Optik Penanda Jarak Aman Kendaraan. *Jurnal Penelitian Transportasi Darat/Jurnal Penelitian Transportasi Darat*, 19(1), 37. <https://doi.org/10.25104/jptd.v19i1.604>
- Novriyanti, E., & Arthur, R. (2024). Analisis Kualitas Butir Soal Ujian Tengah Semester Biologi Umum Menggunakan Model Rasch. *JagoMIPA Jurnal Pendidikan Matematika Dan IPA*, 4(4), 718–733. <https://doi.org/10.53299/jagomipa.v4i4.927>
- Nurlina, F. (2019). Pelaksanaan Asesmen Berbasis Kompetensi di Rumah Sakit Swasta Tipe C Kota Tasikmalaya. *Jurnal Mitra Kencana Keperawatan Dan Kebidanan*, 3(2), 32–40.
- Nuryanti, S., Masykuri, M., & Susilowati, E. (2018). Analisis Iteman dan Model Rasch pada Pengembangan Instrumen Kemampuan Berpikir Kritis Peserta Didik Sekolah Menengah Kejuruan. *Jurnal Inovasi Pendidikan IPA*, 4(2), 224–233. <https://doi.org/10.21831/jipi.v4i2.21442>
- Pratama, D. (2020). Analisis Kualitas Tes Buatan Guru melalui Pendekatan Item Response Theory (IRT) Model Rasch. *Tarbawy: Jurnal Pendidikan Islam*, 7(1), 61–70. <https://doi.org/10.32923/tarbawy.v7i1.1187>
- Rahman, T., Hamdu, G., Putri, A. R., & Merliana, A. (2022). Pelatihan Teknis Analisis Instrumen Penilaian Melalui Pemodelan Rasch bagi Guru Sekolah Dasar. *Jurnal Abdidas*, 3(6), 1029–1037. <https://doi.org/10.31004/abdidas.v3i6.724>
- Retnawati, H. (2014). *Teori Respons Butir dan Penerapannya : Untuk Peneliti, Praktisi Pengukuran dan Pengujian Mahasiswa dan Pascasarjana*. Nuha Medika.
- Rost, J., & Davier, M. von . (1994). A conditional Item-Fit Index for Rasch Models. *Applied Psychological Measurement*, 18(2), 171–182. <https://doi.org/10.1177/014662169401800206>
- Setiawati, I., & Hindriana, A. F. (2020). Pengembangan Asesmen Kompetensi Pedagogik pada Mahasiswa Calon Guru Biologi. *JPBIO (Jurnal Pendidikan Biologi)*, 5(1), 1–11. <https://doi.org/10.31932/jpbio.v5i1.531>
- Siregar, & Panjaitan, A. (2022). A Systematic Literature Review : The Role of Classical Test Theory and Item Response Theory in Item Analysis to Determine the Quality of Mathematics Tests. *Jurnal Fibonacci: Jurnal Pendidikan Matematika*, 2(2), 29. <https://doi.org/10.24114/jfi.v2i2.31056>

- Smith, T. (2024). Iscale: Unveiling the Digital Addiction Quotient - A Comprehensive 6-Item Measure for Internet Compulsion. *Caribbean Journal of Multidisciplinary Studies*, 3(1), 127–149. <https://cjms.utt.edu.tt/ojs/index.php/cjms/article/view/56>
- Soojung, C., Park, E., & Choi, Y. (2018). The Psychometric Properties of The Childhood Health Assessment Questionnaire (CHAQ) in Children with Cerebral Palsy. *BMC Neurology*, 18(1). <https://doi.org/10.1186/s12883-018-1154-9>
- Sugitri, P., & Fitri, S. (2024). Penggunaan Rasch Model untuk Analisis Instrumen Adult Attachment Scale. *Journal of Education Research*, 5(2), 2338–2345. <https://doi.org/10.37985/jer.v5i2.971>
- Sugiyono. (2019). *Metode Penelitian Kuantitatif Kualitatif dan R dan D (2nd ed.)*. Alfabeta.
- Sukarelawan, M. I., Indratno, T. K., & Ayu, S. M. (2024). *N-Gain vs Stacking*.
- Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial (Edisi Revisi)*. Trim Komunikata.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Trim Komunikata.
- Taylor, I. (2007). *A Practical Guide to Assessment Centres and Selection Methods: Measuring Competency for Recruitment and Development*. Kogan Page.
- Widyaningsih, S. W., Sebayang, S. R., & Siswanti, S. (2023). Analisis Kualitas Butir Soal Fisika Pada Penilaian Akhir Semester Gasal Kelas X IPA SMA YPJ Oikoumene Manokwari. *Al-Khazini: Jurnal Pendidikan Fisika*, 3(2), 130–139. <https://doi.org/10.24252/al-khazini.v3i2.37721>
- Zubairi, A. M., & Kassim, N. L. A. (2006). Classical and Rasch Analyses of Dichotomously Scored Reading Comprehension Test Items. *Malaysian Journal of ELT Research*, 2(1), 1–20. <https://meltajournals.com/index.php/majer/article/view/663>