

**KESAHAN DAN KEBOLEHPERCAYAAN INSTRUMEN
RUBRIK ANALITIKAL BAHASA ARAB BAGI TAHAP
PENGUASAAN MURID SEKOLAH KEBANGSAAN**

Hariatul Hafidzah Mahmad Khory¹, *Muhammad Azhar Zailani¹,
Mohd Nazri Abdul Rahman¹ & Harris Shah Abdul Hamid²

ABSTRACT

This study examines the validity and reliability of the items in the Arabic Language Analytical Rubric Level One (BATS). The rubric was developed based on qualitative data obtained through purposive sampling involving outstanding teachers in the teaching and learning of Arabic at primary schools. The validity and reliability of this analytical rubric were analyzed using the Multi-Facet Rasch Model, assisted by Facet software. The BATS Analytical Rubric comprised 99 items (48 under the Language Skills dimension and 51 under the Exemplary dimension). The Exemplary dimension items and the Language Skills dimension items were assessed by six and seven experts respectively. Content validity was tested by examining the polarity of the items, while reliability was assessed by examining item functionality in terms of reliability and the item-person separation index. All items demonstrated positive polarity values (Language Skills: PMC between 0.52 - 0.84; Exemplary: PMC between 0.73 - 0.83). The reliability values for both items and individuals across the two dimensions were at a very high level (>0.9). The person separation index for the Language Skills dimension was 5.64, and 5.72 for the Exemplary dimension. The item separation index for the Language Skills dimension was 3.48, while the Exemplary dimension was 4.08. Based on these findings, the BATS Analytical Rubric is considered to have a high level of validity and reliability, making it suitable for use as a guide for teachers in determining students' mastery levels in Level One Arabic Classroom Assessment at national schools. In conclusion, the reliability analysis and construct validity have contributed to developing a high-quality and meaningful BATS Analytical Rubric.

Keywords: *Validity and Reliability, Arabic Language Education, School Based Assessment*

[1]
Universiti Malaya
azhar@um.edu.my

[2]
University College MAIWP
International

PENGENALAN

Saban tahun, penambahbaikan demi penambahbaikan dilakukan bagi memastikan dasar kurikulum terkini kekal relevan. Pelaksanaan Pentaksiran Bilik Darjah (PBD) turut diperkasakan untuk kemaslahatan murid-murid di negara ini. Dalam pelaksanaan PBD di Malaysia, terdapat 3 cara pelaksanaan guru iaitu: (1) Pentaksiran untuk Pembelajaran, (2) Pentaksiran untuk Penentuan Tahap Penguasaan Murid dan (3) Pentaksiran untuk Kemenjadian Murid. Cara kedua agak berbeza dan tidak pernah dibincangkan dalam mana-mana kajian lepas kerana Tahap Penguasaan (TP) ialah kaedah pengukuran PBD dalam konteks Malaysia (Arumugham, 2020). Guru akan merujuk kepada Dokumen Standard Kurikulum dan Pentaksiran (DSKP) dan membuat pertimbangan untuk merancang dan melaksanakan aktiviti pengajaran dan pentaksiran yang bersesuaian mengikut keperluan murid kelas diajar (Hariatul Hafidzah et al., 2021; Zuhairi, 2017). Seterusnya, guru akan membuat keputusan atas penguasaan pembelajaran murid dengan menentukan Tahap Penguasaan Murid berpandukan Standard Prestasi dalam DSKP.

Penentuan Tahap Penguasaan Murid semasa PBD memerlukan guru menguasai aspek pengukuran, penilaian dan pentaksiran dengan baik agar murid dapat ditaksir secara benar dan adil berdasarkan prestasi mereka. PBD berfokus kepada perkembangan potensi setiap individu secara optimum dan tidak berfungsi membandingkan prestasi antara murid (Danielson, 2019; Faezah Saliman & Zamri, 2023; Nor Azizah et al., 2020). Sehubungan dengan itu, para guru menggunakan pertimbangan mereka berpandukan Standard Prestasi bagi menentukan penguasaan pembelajaran murid (Kementerian Pendidikan Malaysia, 2018). Pertimbangan guru dalam PBD adalah penting kerana polisi pendidikan terkini menjadikan pentaksiran sebagai alat pengajaran utama dan menjadi asas pelbagai mekanisme berakuntabiliti (Martínez et al., 2009). Oleh itu, aspek pertimbangan perlu diberi perhatian dan dikuasai dengan baik oleh guru dalam melaksanakan PBD. Para guru memerlukan satu kriteria yang jelas bagi menentukan TP murid agar ia selari dengan perkembangan pembelajaran semasa murid.

PERNYATAAN MASALAH

Guru yang kompeten berupaya merealisasikan dasar-dasar pendidikan negara dengan baik. Kajian kompetensi Guru Bahasa Arab sekolah rendah mendapati guru mempunyai kompetensi yang tinggi dalam aspek pedagogi dan profesionalisme tetapi perlu meningkatkan kompetensi mereka dalam tiga aspek iaitu aspek penilaian, kurikulum dan bahasa (Abdul Razif Zaini et al., 2020). Hal ini menunjukkan kompetensi pengajaran dan pentaksiran guru adalah tidak selari sedangkan pelaksanaan Pentaksiran Bilik Darjah memerlukan kedua-duanya bergerak seiring (Black & Wiliam, 2018; Hariatul Hafidzah et al., 2021; Wan Mohd Zuhairi, 2017). Justeru, para Guru Bahasa Arab memerlukan lebih bimbingan dan panduan dalam aspek penilaian dan pentaksiran, penguasaan kurikulum atau pengetahuan dalam bidang pengajarannya dan bahasa.

Penentuan tahap penguasaan (TP) menggunakan kaedah pemerhatian, penulisan dan lisan adalah subjektif, sekali gus menambah beban guru dan menimbulkan isu ketidakadilan dalam penilaian (Mustaqim, 2023). Perbezaan pertimbangan guru ini mewujudkan ketidaksamaan pengertian dan interpretasi guru tentang standard yang merupakan panduan utama dalam amalan pentaksiran. Hal ini mengakibatkan wujudnya ancaman kepada konsistensi pertimbangan ketika mentaksir pembelajaran murid. Situasi ini merupakan antara penyebab penilaian guru ke atas murid tidak selari dengan prestasi sebenar murid tersebut.

Ketidaktepatan pentaksiran guru menyebabkan murid kurang bermotivasi untuk belajar serta menjejaskan kredibiliti guru selaku pentaksir (Noh & Matore, 2019). Ketidakselarasan PBD memberikan kesan buruk kepadanya sebagai kayu ukur kemenjadian murid (Noel Jimba et al., 2021). Hal ini kerana penentuan guru terhadap tahap penguasaan murid akan dilaporkan dan didokumenkan untuk pelbagai tujuan sekali gus memberi impak kepada prestasi murid.

Bagi mengatasi isu ini, satu rubrik analitikal Bahasa Arab Tahap Satu telah dibina. Dalam kajian ini, penyelidik menguji dan mengesahkan kandungan Rubrik Analitikal BATS sebagai panduan pertimbangan guru berdasarkan Model Pengukuran Rasch. Proses ini penting bagi mengesahkan kualiti item yang dibina agar dapat memberi penjelasan kepada guru untuk membuat pertimbangan semasa menentukan Tahap Penguasaan Murid bagi mata pelajaran Bahasa Arab Tahap Satu di sekolah kebangsaan.

OBJEKTIF KAJIAN

Kajian ini bertujuan menguji dan mengesahkan kandungan panduan pertimbangan guru dalam bentuk Rubrik Analitikal Bahasa Arab Tahap Satu berdasarkan Model Pengukuran Rasch.

SOALAN KAJIAN

1. Sejauh manakah panduan pertimbangan guru dalam bentuk Rubrik Analitikal Bahasa Arab Tahap Satu boleh dipercayai dan sah berdasarkan perbezaan ketegasan penilai dan ketekalan penilaian pakar?
2. Sejauh manakah panduan pertimbangan guru dalam bentuk Rubrik Analitikal Bahasa Arab Tahap Satu boleh dipercayai dan sah berdasarkan kedudukan dan ketekalan item?

METODOLOGI KAJIAN

Penyelidik telah menganalisis dan menyusun perincian pertimbangan TP peserta-peserta kajian berbantuan Standard Prestasi untuk membina Rubrik Analitikal Bahasa Arab Tahap Satu (BATS). Pertimbangan lima peserta kajian yang dipilih secara pensampelan bertujuan dalam kalangan guru cemerlang Bahasa Arab dikumpul menggunakan pendekatan kualitatif melalui temu bual, pemerhatian bilik darjah dan bukti dokumen. Seterusnya, data tersebut disaring dan dianalisis secara komprehensif dan menemui dua tema yang membentuk dua dimensi Rubrik Analitikal iaitu Dimensi Kemahiran Bahasa dan Dimensi Mithali. Terdapat dua kaedah penghasilan item iaitu kaedah induktif dan kaedah deduktif (Hinkin, 1998; 1995). Kaedah induktif ialah menghasilkan item berdasarkan temu bual responden manakala melalui kaedah deduktif pula, item dihasilkan melalui tinjauan literatur atau menggunakan soal selidik sedia ada (Wan Noor Hanim, 2021). Dalam kajian ini, pengkaji menggunakan kaedah induktif iaitu membina item instrumen BATS berdasarkan temu bual mendalam ke atas 5 peserta kajian.

Instrumen

BATS mempunyai dua dimensi berjumlah 99 item. Dimensi pertama ialah kemahiran bahasa yang mengandungi 48 item manakala dimensi kedua ialah mithali yang mengandungi 51 item. Pembinaan item rubrik ini telah disemak oleh tiga orang pakar bahasa dan pakar kandungan.

Pakar yang telah dilantik telah menilai kualiti item berdasarkan empat kriteria iaitu:

- i. Ketepatan – Tugas yang terdapat pada item mewakili dimensi dan boleh menguji dimensi yang diuji.
- ii. Munasabah – Pernyataan pada item boleh dipercayai dan munasabah berkaitan tugas yang dilakukan.
- iii. Kejelasan – Struktur ayat dan laras bahasa yang digunakan adalah sesuai, jelas, mudah difahami dan tidak mengelirukan responden yang akan menggunakan instrumen ini.
- iv. Ringkas – Pernyataan yang digunakan pada item adalah ringkas dan mudah difahami oleh responden daripada sudut istilah, penggunaan perkataan dan struktur ayat.

Penilaian setiap kriteria menggunakan skala penarafan empat mata iaitu Skala 1 mewakili sangat tidak sesuai, skala 2 mewakili tidak sesuai, skala 3 mewakili sesuai dan skala 4 mewakili sangat sesuai. Di samping itu, pakar juga diminta mengesahkan aras kesukaran item yang dibina dan mengemukakan pendapat sekiranya terdapat penambahbaikan terhadap item.

Jadual 1: Contoh Item dalam Rubrik Analitikal Bahasa Arab Tahap Satu (BATS)

Dimensi	Kemahiran Bahasa	Mithali
TP		
TP2	Murid dapat mengajuk sebutan dengan bimbingan guru	Murid lambat menyebut apabila diminta oleh guru

TP5 Murid dapat membezakan harakat panjang Murid sukarela membantu rakan pendek bacaan, bacaan huruf sakinah dan menyebut perkataan musyaddadah dengan betul

Peserta Kajian

Rubrik analitikal ini diedarkan kepada 7 pakar dimensi Kemahiran Bahasa dan 6 pakar dimensi Mithali untuk dinilai dan akhirnya diuji dan disahkan menggunakan Teori Pengukuran Rasch. Pakar penilai kajian ini terdiri daripada pakar pengukuran dan penilaian di Institut-institut Pengajian Tinggi, pensyarah bidang pedagogi dan Bahasa Arab di Institut Pendidikan Guru Malaysia, penggubal DSKP Bahasa Arab KSSR, serta Guru Bahasa Arab yang berpengalaman dan cemerlang dalam pengajaran di sekolah rendah.

Prosedur Analisis Data

Dalam kajian ini, data kajian yang diperoleh daripada penilai direkodkan dalam perisian *Microsoft Excel*. Seterusnya, analisis dilakukan dengan menggunakan Model Pengukuran Rasch Multi Facet iaitu aplikasi *Facets* versi 3.67.1 bagi analisis kandungan dan analisis semakan ciri-ciri psikometri Instrumen BATS. Standard pengukuran yang digunakan adalah berdasarkan soalan kajian seperti jadual:

Jadual 2: Standard Pengukuran

Tujuan	Kriteria	Maklumat statistik
Menilai kebolehppercayaan instrumen	Indeks Kebolehppercayaan item dan individu	<ul style="list-style-type: none"> • ≥ 0.80 (Fisher, 2007)
Menilai ketegasan penilai	Indeks pengasingan item dan individu	<ul style="list-style-type: none"> • ≥ 2 (Fisher, 2007)
Menilai ketegasan penilai	Pemetaan kesukaran item - individu	<ul style="list-style-type: none"> • Nilai logit • Semakin rendah nilai logit pengukuran, semakin mudah penilai bersetuju terhadap item tersebut (Eckes, 2015)
Menilai kualiti item	Kesepadanan item	<ul style="list-style-type: none"> • Nilai outfir MNSQ 0.5 – 1.5 • Nilai $Z_{std} \pm 2$ (Boone et. al., 2014) • Nilai logit
	Polariti item	<ul style="list-style-type: none"> • Nilai PTMea. Corr. dalam positif (Bond & Fox, 2015)

Kesahan kandungan membincangkan sejauh mana item yang terkandung dalam instrumen menguji dimensi yang dikaji dan ketepatan item yang digunakan (Baghaei & Amrahi, 2011 dalam Wan Norhanim, 2021). Kesahan kandungan sangat berguna dalam pembangunan instrumen kerana ia boleh mengurangkan kesilapan semasa membina instrumen dan meningkatkan kemungkinan untuk mendapatkan indeks kesahan dimensi yang baik. Dalam kajian ini, instrumen item rubrik yang dibina diserahkan kepada pakar bahasa untuk tujuan semakan struktur ayat dan penggunaan tatabahasa yang betul.

Analisis Kesahan Kandungan Instrumen

Respons yang diterima daripada pakar dianalisis menggunakan perisian *Facets* bagi model Rasch pelbagai faset atau *Multiple Facet Rasch Model* (MFRM). Antara kelebihan Model Rasch ialah dapat mengenal pasti respons luar jangka (*unexpected responses*) daripada panel pakar, mengenal pasti pertimbangan yang tidak sesuai serta dapat meramalkan data yang hilang berdasarkan corak tindak balas sistematik (Fahmina et al., 2019). Penilaian pakar telah dianalisis berdasarkan kepada dua aspek iaitu tahap ketegasan dan ketekalan dalaman penilai. Analisis Rasch pelbagai faset memberikan maklumat mengenai kebolehppercayaan dan indeks pengasingan penilai (*rater*), kebolehppercayaan dan indeks pengasingan item, kriteria penilaian instrumen yang telah dicapai dan tidak dicapai pada setiap item, dan respons penilai yang tidak dijangka.

Analisis Kesahan Dimensi Instrumen

Analisis kesahan dimensi dijalankan untuk menyemak sejauh mana item yang dikemukakan dalam instrumen menggambarkan dimensi yang diukur. Pembuktian instrumen ini mengandungi kesahan dimensi berdasarkan prinsip pengukuran Rasch iaitu ciri unidimensi instrumen yang disemak melalui analisis komponen prinsipal. Analisis ini dijalankan untuk menyemak bilangan dimensi yang terkandung dalam instrumen.

Kebolehpercayaan Instrumen

Analisis kebolehpercayaan digunakan bagi mengukur nilai kebolehpercayaan item dan responden. Analisis ini menganggarkan sejauh mana skor dapat mendiskriminasikan seorang individu dengan individu yang lain (Boone et al., 2014). Indeks kebolehpercayaan item merujuk kepada anggaran ketekalan item pada skalanya walaupun item yang sama diberikan pada sampel berlainan. Fisher (2007) berpendapat nilai kebolehpercayaan melebihi 0.94 dikategorikan cemerlang, nilai 0.93 hingga 0.91 dikategorikan sebagai amat baik dan nilai 0.90 hingga 0.81 adalah baik.

Nilai pengasingan individu dan item perlu diketahui untuk membuat jangkaan bahawa instrumen yang dibina dapat mengelompokkan individu mengikut tahap keupayaan dan mengelompokkan item mengikut aras kesukaran (Wright & Masters, 1982). Nilai indeks pengasingan yang baik melebihi dua dan menunjukkan item yang terkandung dalam instrumen dapat mendiskriminasi responden mengikut tahap keupayaan dan mengasingkan item mengikut aras kesukaran.

Kualiti Item

Kualiti item yang terkandung dalam instrumen rubrik analitikal BATS disemak melalui analisis kesepadanan item dan juga nilai polariti item menggunakan model Rasch (Azrilah Abul Aziz et al., 2015). Semakan pada parameter statistik dilakukan pada min kuasa dua outfit (outfit MNSQ) dan statistik kesesuaian seragam outfit (outfit Zstd) untuk mengenal pasti sekiranya terdapat percanggahan antara data dengan model Rasch. Nilai yang produktif untuk digunakan dalam pengukuran ialah 0.5 hingga 1.5 (Boone et al., 2014; Wan Noor Hanim, 2022). Nilai yang melebihi 1.5 menunjukkan item tersebut turut mengukur perkara berbeza dengan item lain yang terkandung dalam instrumen. Sebaliknya, nilai 0.50 pula merupakan indikator terdapat pertindihan dimensi dengan item lain. Item yang tidak memenuhi had potongan outfit MNSQ akan dipertimbangkan untuk digugurkan daripada instrumen.

Selain itu, analisis polariti item dijalankan untuk menyemak item yang terkandung dalam instrumen selari dengan pemboleh ubah yang diukur (Linacre, 2010). Menurut Bond dan Fox (2015), nilai korelasi (PTMe.Corr.) bernilai positif menunjukkan indikator item mengukur dimensi yang diuji. Nilai PTMe.Corr. negatif pula menunjukkan item sukar difahami oleh responden dan dicadangkan untuk digugurkan. Semakan polariti item adalah analisis awal untuk tujuan kesahan dimensi instrumen yang dibina (Azrilah Abdul Aziz et al., 2015).

DAPATAN DAN PERBINCANGAN***Laporan Ketegasan Penilai***

Jadual menunjukkan ringkasan ketegasan penilai yang terlibat dalam kajian ini. Terdapat dua kumpulan penilai dalam kajian ini bagi dimensi Kemahiran Bahasa dan Mithali.

Jadual 3: Ringkasan Ketegasan Penilai Kemahiran Bahasa

Penilai	Model		Infit		Outfit		Ptmea.
	Pengukuran	Ralat	MnSq	Zstd	MnSq	Zstd	Corr
C	-2.46	.13	.81	-2.2	.82	-2.1	.73
F	-2.39	.13	1.02	.2	1.05	.6	.52
E	-3.30	.15	1.19	1.9	1.15	1.3	.68
A	-4.54	.2	0.76	-1.7	.58	-2.4	.84
B	-4.63	.2	1.38	2.2	1.34	1.5	.71
G	-4.67	.21	.95	-.3	.84	-.7	.81
D	-5.16	.24	1.06	.3	1.0	0.0	.81

Min : .73

S.D : .10

Pengasingan :

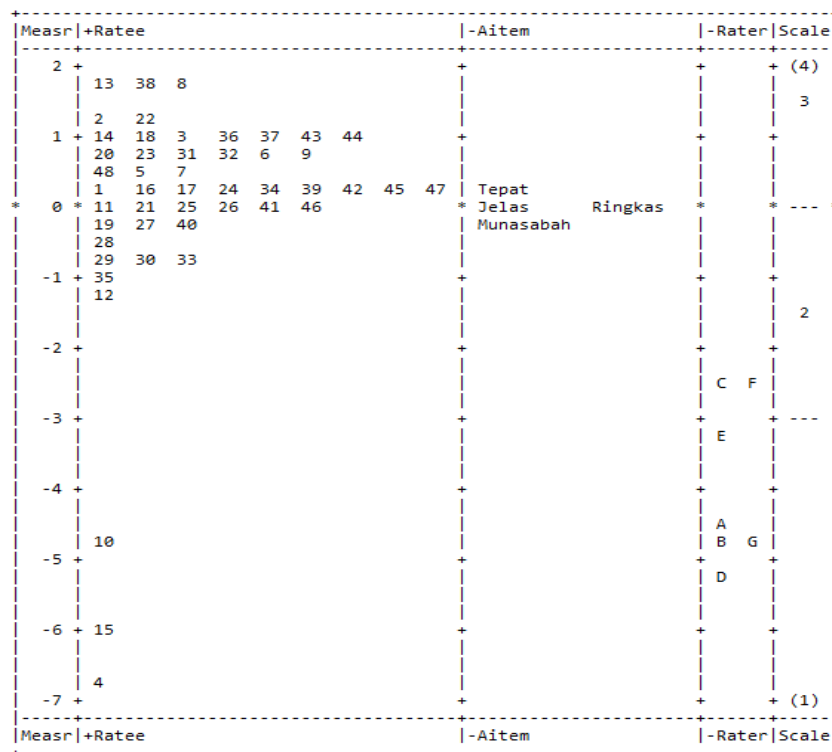
5.64

Strata : 7.85

Indeks Kebolehpercayaan : .97

Bagi dimensi kemahiran bahasa, berdasarkan nilai logit pengukuran pada rajah, Pakar F ialah paling tegas dan pakar D pula paling mudah memersetujui item dalam instrumen rubrik. Nilai statistik bagi infit dan outfit digunakan bagi menilai kesepadanan respons yang dikemukakan oleh penilai. Nilai yang diterima adalah dalam lingkungan 0.5 hingga 1.5. Dapatan menunjukkan kesemua pakar dimensi Kemahiran Bahasa sepadan dengan model iaitu penilai konsisten dalam memberikan rating yang tidak dipengaruhi oleh faktor luaran. Berdasarkan rajah, didapati Pakar B dan Pakar G mempunyai tahap kesukaran yang hampir sama dalam memersetujui item-item yang terkandung dalam instrumen dengan nilai logit berada dalam lingkungan -4.63 hingga -4.67. Hal ini menunjukkan, panel pakar ini mempunyai pandangan yang hampir sama berkaitan item walaupun tiada perbincangan dilakukan.

Kriteria ketepatan merupakan aspek yang sukar mencapai kesepakatan diikuti kriteria jelas dan ringkas. Kriteria munasabah merupakan aspek yang paling mudah mencapai kesepakatan penilai. Nilai pengasingan bagi ketegasan penilai ialah 5.64 dan indeks kebolehpercayaan penilai berada pada kategori sangat baik iaitu 0.97.



Rajah 1: Ringkasan Keputusan Keseluruhan Faset Dimensi Kemahiran Bahasa

Ringkasan ketegasan penilai bagi dimensi Mithali adalah seperti berikut:

Jadual 4: Ringkasan Ketegasan Penilai Mithali

Penilai	Model		Infit		Outfit		Ptmea.
	Pengukuran	Ralat	MnSq	Zstd	MnSq	Zstd	Corr
F	-3.02	.19	.81	-1.4	.8	-1.3	.82
C	-5.16	.16	1.12	1.7	1.04	.4	.73
E	-5.48	.17	.83	-2.5	.74	-2.8	.82
B	-5.60	.17	1.00	.0	.98	-.1	.75
D	-5.71	.17	.98	-.1	.89	-.9	.76
A	-6.15	.18	1.14	1.3	1.29	1.5	.75

Min : .77

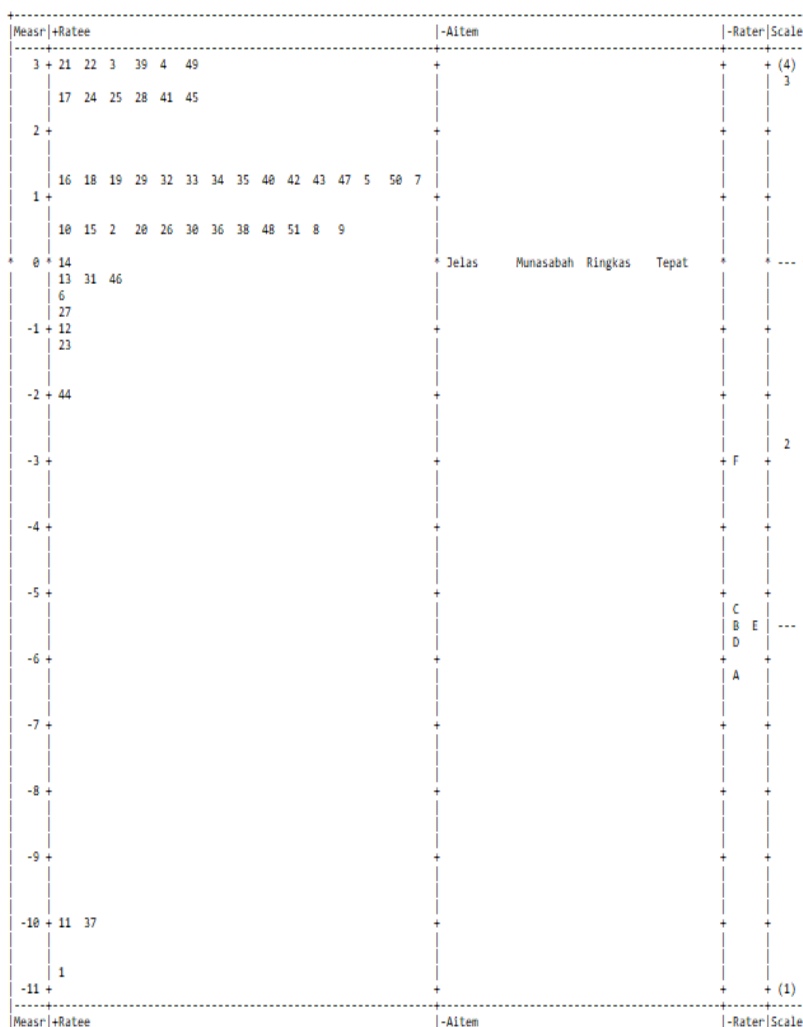
S.D : .03

Pengasingan : 5.72 Strata : 7.96 Indeks Kebolehpercayaan : .97

Bagi dimensi Mithali, berdasarkan nilai logit pengukuran pada rajah, Pakar F ialah paling tegas dan Pakar A pula paling mudah mempersetujui item dalam instrumen rubrik. Nilai statistik bagi infit dan outfit digunakan bagi menilai kesepadanan respons yang dikemukakan oleh penilai. Nilai yang diterima adalah dalam lingkungan 0.5 hingga 1.5. Dapatan menunjukkan kesemua pakar dimensi Mithali sepadan dengan model iaitu penilai konsisten dalam memberikan rating yang tidak dipengaruhi oleh faktor luaran.

Berdasarkan rajah, didapati Pakar B dan Pakar E mempunyai tahap kesukaran yang hampir sama dalam mempersetujui item-item yang terkandung dalam instrumen dengan nilai logit yang berada dalam lingkungan -5.60 hingga -5.48. Hal ini menunjukkan panel pakar ini mempunyai pandangan yang hampir sama berkaitan item walaupun tiada perbincangan dilakukan.

Kriteria ketepatan merupakan aspek yang sukar mencapai kesepakatan diikuti kriteria kejelasan dan ringkas. Kriteria munasabah merupakan aspek yang paling mudah mencapai kesepakatan penilai. Nilai pengasingan bagi ketegasan penilai ialah 5.72 dan indeks kebolehpercayaan penilai berada pada kategori sangat baik iaitu 0.98.



Rajah 2: Ringkasan Keputusan Keseluruhan Faset Dimensi Mithali

Berdasarkan analisis dan bacaan peta item, penilai bagi kedua-dua dimensi menunjukkan ketegasan yang berbeza antara satu sama lain. Julat perbezaan antara penilai bagi kedua-dua dimensi adalah seperti berikut:

Jadual 5: Julat Perbezaan antara Penilai bagi Dimensi Kemahiran Bahasa Dan Dimensi Mithali

Dimensi Kemahiran Bahasa			Dimensi Mithali		
Pakar	Logit	Perbezaan	Pakar	Logit	Perbezaan
C	-2.46	.07	F	-3.02	2.14
F	-2.39	.84	C	-5.16	.32
E	-3.30	1.24	E	-5.48	.12
A	-4.54	.09	B	-5.60	.11
B	-4.63	.03	D	-5.71	.44
G	-4.67	.49	A	-6.15	-
D	-5.16	-			

Bagi dimensi Kemahiran Bahasa, julat perbezaan ketegasan antara penilai adalah tidak ketara iaitu dalam lingkungan .03 hingga 1.24. Julat perbezaan antara penilai paling tegas dan paling mudah memersetujui item rubrik ialah 2.7 iaitu antara Pakar F (-2.46) dan Pakar D (-5.16). Nilai perbezaan yang paling dekat adalah antara Pakar B dan Pakar G iaitu .03. Pakar B dan Pakar G masing-masing merupakan pakar ketiga dan kedua yang paling mudah memersetujui item rubrik analitikal yang dikemukakan. Nilai perbezaan yang paling jauh pula ialah 1.24 iaitu antara Pakar E dan Pakar A. Pakar E ialah pakar ketiga tegas dan Pakar A ialah pakar keempat tegas dan sukar memersetujui item rubrik dimensi Kemahiran Bahasa.

Bagi dimensi Mithali pula, julat perbezaan antara penilai adalah lebih besar iaitu antara .11 hingga 2.14. Julat perbezaan antara penilai paling tegas dan paling mudah memersetujui item rubrik ialah 3.13 iaitu Pakar F (-3.02) dan Pakar A (-6.15). Nilai perbezaan yang paling dekat adalah .11 iaitu antara Pakar B dan Pakar D. Pakar B merupakan pakar keempat paling tegas dan pakar D merupakan pakar kelima paling tegas atau paling sukar memersetujui item rubrik yang dikemukakan. Nilai perbezaan yang paling jauh pula adalah antara Pakar F dan Pakar C iaitu 2.14. Penilai F merupakan penilai yang paling tegas dan paling sukar memersetujui item rubrik yang dikemukakan serta menunjukkan nilai perbezaan yang jauh dan ketara berbanding pakar-pakar lain dalam dimensi ini. Selain Pakar F, nilai perbezaan antara pakar dalam dimensi ini ialah .11, .12, .32, .44.

Berdasarkan dapatan kajian, panduan pertimbangan guru dalam bentuk BATS, kesemua pakar dimensi Kemahiran Bahasa dan Mithali sepadan dengan model iaitu penilai konsisten dalam memberikan rating yang tidak dipengaruhi oleh faktor luaran. Justeru, penilaian pakar penilai Rubrik Analitikal BATS adalah sah dan boleh dipercayai untuk menyokong penghasilan Rubrik Analitikal ini.

Laporan Kualiti Item Rubrik Analitikal BATS

Kualiti item disemak berdasarkan dua aspek iaitu nilai logit pengukuran dan nilai statistik kesepadanan. Nilai logit merupakan jumlah rating yang diberikan oleh pakar pada setiap item berdasarkan kriteria yang ditetapkan. Nilai statistik kesepadanan menunjukkan jumlah rating yang tidak terjangka pada setiap item. Item adalah tidak sepadan sekiranya penilai yang tegas memberikan rating yang tinggi sementara penilai yang mudah (lenien) memberikan rating yang rendah pada item yang dinilai. Item yang tidak sepadan akan disemak semula pada kriteria yang tidak memuaskan atau tidak dimasukkan sebagai item dalam instrumen.

Berdasarkan analisis, item-item ini telah dikategorikan kepada empat kategori berdasarkan nilai logit pengukuran. Item yang mempunyai nilai logit melebihi 1.56 dikategorikan sebagai amat baik, nilai logit antara 0.00 hingga 1.56 dikategorikan sebagai baik, nilai logit <0.00 hingga -1.56 dikategorikan sebagai boleh diterima dan item melebihi nilai logit -1.56 dikategorikan sebagai item yang lemah. Jadual menunjukkan ringkasan kualiti item yang terkandung dalam instrumen rubrik.

Jadual 6: Ringkasan Kategori Item Dimensi Kemahiran Bahasa

Kategori Item	Nilai Logit	Item	Jumlah
Sangat baik	>1.56	38, 13, 8	3
Baik	<1.56 hingga 0.00	2, 22, 3, 14, 18, 36, 37, 43, 44, 6, 9, 20, 23, 31, 32, 5, 7, 48, 1, 17, 34, 39, 47, 24, 16, 42, 45, 25, 26	29
Diterima	<0.00 hingga -1.56	11, 21, 41, 46, 27, 19, 40, 28, 30, 33, 29, 35	13
Lemah	>-1.56	10, 15, 4	3

Min : .00

S.D: 1.68

Pengasingan : 3.48

Strata : 4.97

Indeks Kebolehpercayaan: .92

Berdasarkan jadual ringkasan kategori item dimensi kemahiran bahasa, nilai pengasingan item ialah 3.48 dan indeks kebolehpercayaan item ialah 0.92. Secara keseluruhan, 3 item sangat baik, 29 item dikategorikan sebagai baik, 13 item dikategorikan sebagai boleh diterima dan tiga item yang lemah.

Jadual 7: Ringkasan Kategori Item Dimensi Mithali

Kategori Item	Nilai Logit	Item	Jumlah
Sangat baik	>1.56	3, 4, 21, 22, 39, 49, 17, 24, 25, 28, 41, 45	12
Baik	<1.56 hingga 0.00	5, 7, 16, 18, 19, 32, 33, 34, 35, 40, 42, 43, 47, 50, 30, 2, 8, 9, 10, 15, 20, 26, 36, 38, 48, 51, 14	28
Diterima	<0.00 hingga -1.56	13, 31, 46, 6, 27, 12, 23	7
Lemah	>-1.56	44, 11, 37, 4	4

Min : .65

S.D: 3.30

Pengasingan : 4.08

Strata :5.78

Indeks Kebolehpercayaan: 0.94

Berdasarkan jadual ringkasan kategori item dimensi mithali, nilai pengasingan item ialah 4.08 dan indeks kebolehpercayaan item ialah 0.94. Secara keseluruhan, 12 item dikategorikan sebagai sangat baik, 28 item dikategorikan baik, 7 item dikategorikan sebagai boleh diterima dan 4 item dikategorikan lemah.

KESIMPULAN

BATS memberikan contoh pertimbangan guru dalam aspek kognitif dan tingkah laku serta nilai murid yang menghuraikan TP1 hingga TP6 dalam Standard Prestasi Bahasa Arab mengikut kemahiran bahasa. Dapatan pertimbangan peserta kajian dalam menentukan Tahap Penguasaan Murid bermula TP1 hingga TP6 telah dianalisis dan disusun sebagai Item BATS di sekolah rendah. Dapatan ini menyetengahkan dua dimensi pertimbangan guru dalam penentuan TP murid: dimensi kemahiran bahasa dan nilai. Dimensi kemahiran bahasa memperincikan komponen kemahiran asas bahasa yang ditaksir oleh guru iaitu kemahiran mendengar, bertutur, membaca dan menulis. Dimensi Mithali pula mengemukakan perincian aspek tingkah laku dan nilai murid yang dipertimbangkan oleh guru bagi menentukan TP mereka.

Berdasarkan analisis data yang dijalankan menggunakan Teori Pengukuran Rasch, item-item yang terkandung dalam Rubrik Analitikal BATS mengukur tahap penguasaan murid berdasarkan dimensi mithali dan kemahiran bahasa mempunyai eviden yang menyokong ciri-ciri psikometri. Berdasarkan dapatan kajian, panduan pertimbangan guru dalam bentuk BATS, kesemua pakar dimensi Kemahiran Bahasa dan Mithali sepadan dengan model iaitu penilai konsisten dalam memberikan rating yang tidak dipengaruhi oleh faktor luaran. Justeru, penilaian pakar penilai Rubrik Analitikal BATS adalah sah dan boleh dipercayai untuk menyokong penghasilan Rubrik Analitikal ini. Berdasarkan analisis kualiti item Rubrik Analitikal BATS pula, item-item ini dikategorikan kepada empat kategori berdasarkan nilai logit pengukuran iaitu kategori (1) sangat baik, (2) baik (3) diterima dan (4) lemah.

Secara keseluruhannya, BATS dapat digunakan untuk membantu para guru merancang pengajaran dan melaksanakan pentaksiran dengan lebih kompeten. Penggunaan meluas rubrik analitikal ini dalam kalangan Guru Bahasa Arab dapat mewujudkan keseragaman penentuan tahap penguasaan murid dalam Pentaksiran Bilik Darjah. Hal ini kerana, BATS dapat digunakan sebagai panduan menentukan tahap penguasaan murid dengan kriteria yang lebih jelas dan melaporkan TP yang sepadan dengan prestasi sebenar murid.

RUJUKAN

- Abdul Razif bin Zaini , Noor Shamshinar binti Zakaria, Mohd Rofian Ismail, Mohd Redzaudin Ghazali & Hasmadi Hamdan (2020). Kompetensi guru bahasa Arab sekolah rendah di Malaysia. *Jurnal Sultan Alauddin Sulaiman Shah*, 7(1). e-ISSN: 2289 8042
- Arumugham, K. S. (2020). Kurikulum, pengajaran dan pentaksiran dari perspektif pelaksanaan pentaksiran bilik darjah. *Asian People Journal (APJ)*. <https://doi.org/10.37231/apj.2020.3.1.175>
- Azrilah Abdul Aziz, Mohd Saidudin Masodi, & Zaharim, A. (2015). Asas model pengukuran Rasch, pembentukan skala dan struktur pengukuran. Penerbit UKM.
- Baghaei, P., & Amrahi, N. (2011). *Validation of a multiple choice english vocabulary test with the Rasch Model*. Academy Publisher. <http://ojs.academypublisher.com/index.php/jltr/article/view/5528>

- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. In 3rd Edition (Ed.), Routledge Taylor & Francis Group. <https://doi.org/10.4324/9781410614575>
- Boone, W. J., Yale, M. S., & Staver, J. R. (2014). *Rasch analysis in the human sciences*. Springer. <https://doi.org/10.1007/978-94-007-6857-4>
- Danielson, C. (2019). Assessment for learning—For teachers as well as students. In *The Future of Assessment*. <https://doi.org/10.4324/9781315086545-7>
- Faezah Saliman & Zamri Mahamod (2023). Amalan guru Bahasa Melayu di sekolah kurang murid dalam melaksanakan dan melaporkan Pentaksiran Bilik Darjah. *Jurnal Pendidikan Bahasa Melayu*. <https://spaj.ukm.my/jpbm>
- Fisher, J.W.P. (2007). Rating Scale instrument quality kriteria. <https://www.rasch.org/rmt/rmt211m.htm>
- Hariatul Hafidzah Mahmad Khory, Mohd Nazri Abdul Rahman & Muhammad Azhar Zailani. (2021). Pengurusan pentaksiran bilik darjah mata pelajaran Bahasa Arab berasaskan keperluan pembelajaran murid. *JuPiDi: Jurnal Kepimpinan Pendidikan*, 8(2), 41-57
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, 21(5), 967–988. <https://doi.org/10.1177/014920639502100509>
- Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 1(1), 104–121. <https://doi.org/10.1177/109442819800100106>
- Kementerian Pendidikan Malaysia. (2018). *Panduan Pelaksanaan Pentaksiran Bilik Darjah*. Kementerian Pendidikan Malaysia (Pertama, Vol. 1). Putrajaya.
- Linacre, J. M. (2010). *A user's guide to WINSTEPS: Rasch-model computer program*.
- Martínez, J. F., Stecher, B., & Borko, H. (2009). Classroom assessment practices, teacher judgments, and student achievement in mathematics: Evidence from the ECLS. *Educational Assessment*. <https://doi.org/10.1080/10627190903039429>
- Muhammad Mustaqim bin Roslan. (2023). *Kaedah Pentaksiran Bilik Darjah (PBD) dan kesannya kepada murid tahap satu di Malaysia: Analisis subjek Pendidikan Islam*. Proceeding Conference: International Conference of Islamic Universities (ICIU 2023)At: Universiti Islam Sultan Sharif Ali (UNISSA) - Brunei Darussalam
- Noel Jimbai Balang, Kiah Sakudan & Mohamad Faizal Semion. (2021). *Lestarikan pentaksiran bilik darjah (PBD) demi kemenjadian murid*. Researchgate. https://www.researchgate.net/publication/351135097_LESTARIKAN_PENTAKSIRAN_BILIK_DARJAH_PBD_DEMI_KEMENJADIAN_MURID (diakses pada 20 Mac 2024)
- Noh, M. F. bin M., & Matore, M. E. bin M. (2019). Brunswik's Lens Model: This is how to inspire accurate raters. *Creative Education*. <https://doi.org/10.4236/ce.2019.1012212>
- Nor Azizah Atan, Zaharah Mohamad, Norasyikin Yaacob @ Norshila Shafie. (2020). *Amalan Pentaksiran Bilik Darjah bagi pentaksiran autentik dalam kalangan guru sekolah rendah*. Proceedings of International Conference on the Future of Education, pp. 184-196. Institute of Teacher Education Tuanku Bainun Campus, Penang.
- Sidhu, G. K., Kaur, S., & Chi, L. J. (2018). CEFR-Aligned School-Based Assessment in the Malaysian primary ESL classroom. *Indonesian Journal of Applied Linguistics*, 8, 452-463.
- Wan Noor Hanim Abd Aziz. (2021). *Pembinaan dan penilaian instrumen efikasi guru dalam implementasi pentaksiran bilik darjah berdasarkan pentaksiran rujukan standard*. Disertasi Sarjana. Universiti Malaya
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Mesa Press.
- Zuhairi, W. A. W. M. (2017). *Evidens pentaksiran sekolah dalam pengajaran bahasa Arab sekolah rendah*. Tesis Doktor Falsafah. Universiti Malaya.