
Penilaian Tahap Kesukaran Item Ujian Akhir Sesi Akademik (UASA) Matematik Tingkatan 1

Saralah Sovey¹, Mohd Effendi Ewan Mohd Matore²

¹SMK Sungai Ramal, Kajang Selangor, Malaysia
E-mail: saralahsovey@gmail.com

² Research Centre of Education Leadership and Policy, Faculty of Education, The National University of Malaysia, UKM
Bangi, Selangor, Malaysia
E-mail: effendi@ukm.edu.my

Abstrak

Kajian ini bertujuan menganalisis item Ujian Akhir Sesi Akademik matapelajaran Matematik di kalangan murid tingkatan 1 di sebuah sekolah terpilih di Negeri Selangor. Kajian terhadap analisis item soalan Matematik menggunakan model Rasch adalah terhad. Analisis item ini melibatkan ujian bertulis kertas penilaian akhir sesi pengajian 2022 melalui penglibatan 71 orang pelajar. Analisis deskriptif model Rasch daripada perisian Winstep 3.71.0.1 digunakan bagi menilai statistik item melalui analisis unidimensi ujian, kesesuaian item, polariti item, dan analisis tahap kesukaran item yang dilibatkan sebagai ujian akhir. Analisis ini bertujuan bagi mengesahkan kualiti item penilaian yang dilibatkan dari sudut statistik sekali gus mengesan kesesuaian item tersebut untuk digunakan atau diubahsuai berdasarkan standard kandungan sekali gus mencapai hasil pembelajaran yang ditetapkan. Dapatan menunjukkan sebahagian besar item adalah memenuhi keperluan pengukuran model Rasch. Item-item adalah sesuai digunakan semula sebagai item ujian. Nisbah kesukaran item kurang bersesuaian dengan konstruk pembelajaran. Melalui dapatan yang diperoleh diharapkan dapat membantu pengkaji meningkatkan kualiti pentaksiran dan penilaian matematik di sekolah.

Kata Kunci: Item; Matematik; Penilaian; Rasch; Tahap Kesukaran

Abstract

This study aims to analyse the items of the final academic session of the mathematics subject among form one students in a selected school in the State of Selangor. Research on the analysis of Mathematics question items using the Rasch model is limited. Item analysis involves a set of multiple-choice question test of the final assessment paper of the 2022 study session through the involvement of 71 students. Descriptive analysis of the Rasch model from the Winstep 3.71.0.1 software was used to evaluate item statistics through unidimensional analysis of the test, item fit, item polarity, and analysis of the difficulty level of the items. This analysis aims to verify the quality of the evaluation items involved from a statistical point of view, thus detecting the suitability of the item to be used or modified based on the content standard, thus achieving the set learning outcomes. Findings show that most of the items meet the measurement requirements of the Rasch model. The items are suitable for re-use as test items. The difficulty ratio of the items is less compatible with the learning construct. Through the findings, it is hoped that researchers can improve the quality of mathematics assessment and evaluation in schools.

Keywords: Assessment; Difficulty Level; Item; Mathematics; Rasch

I. PENGENALAN

Ujian pencapaian merupakan pengukuran untuk mengesan pencapaian pengajaran dan pembelajaran, meninjau kemajuan pembelajaran dan mendiagnosis masalah pembelajaran pelajar. Pengukuran telah sebatu dalam proses pendidikan dan membantu pendidik untuk mencari formula yang strategik. Justeru, kaedah statistik yang mantap adalah penting untuk menilai kualiti item ujian. Maka, objektif utama ujian pencapaian adalah untuk mengukur hasil pembelajaran di samping menyalurkan maklumat terhadap keputusan yang bermakna

(Davies, 1977). Pengukuran dan penilaian meramalkan kejayaan murid dan mendiagnosis kesukaran murid kerana kejayaan dan kegagalan seseorang murid dalam kelas boleh diramal. Maka, kesukaran murid perlu diberi keutamaan untuk pemulihan. Dengan mengetahui kejayaan dan kesukaran murid, guru dapat menumpukan perhatian kepada titik-titik yang memerlukan peningkatan atau pemulihan.

Seterusnya, dari aspek pengajaran, pengukuran dan penilaian memfokuskan pemurnian proses pengajaran sepertimana yang dinyatakan oleh Gronlund [1] iaitu ujian pencapaian digunakan untuk menyokong dan mengukuhkan aspek-aspek

61

Received: 06 June 2023

Revised: 20 June 2023

Accepted: 30 June 2023

dalam proses pengajaran. Sehubungan itu, ciri psikometrik soalan ujian harus difokus [2] bagi memperincikan kebolehan pelajar dalam sesuatu mata pelajaran dan menyalurkan maklumat kepada guru tentang prestasi seseorang pelajar. Umumnya, guru menggunakan skor peperiksaan untuk menilai kemajuan pelajar dan penguasaan pengetahuan dalam kalangan pelajar [3]. Kadangkala berlaku salah tanggapan terhadap skor seseorang pelajar yang diperolehi dalam sesuatu peperiksaan. Contohnya dua orang pelajar telah mendapat markah yang sama di dalam satu peperiksaan, namun salah seorang daripadanya dianggap lebih pandai kerana berjaya menjawab betul pada item sukar [4].

Oleh itu, analisis statistik perlu dilakukan supaya penilaian yang tepat boleh dilakukan berdasarkan kebolehan pelajar yang sebenar. Dalam kajian Pomeranz et al. dan Gulten et al telah menggunakan model Rasch dalam mengukur kebolehpercayaan dan kesahan instrumen kajian mereka . [5, 6]. Model ini mempunyai maklumat tambahan yang tidak diperolehi di dalam Teori Pengukuran Klasik. Ini kerana model Rasch menyediakan metodologi yang lebih tepat dan terperinci untuk mengenalpasti sesuatu instrumen pada tahap item (item level) dan trait pendam (ability of student). Justeru, model Rasch yang merupakan salah satu model dalam Teori Respons Item (Item Response Theory, IRT) dikatakan lebih sesuai digunakan untuk menganalisis soalan-soalan aneka pilihan, kerana model ini mempunyai dua ciri keistimewaan: pertama, model Rasch melibatkan parameter yang paling sedikit, oleh itu ia lebih mudah diaplikasikan; kedua, model Rasch mempunyai objektif yang spesifik iaitu membenarkan pemisahan lengkap item dan keupayaan anggaran [7].

Selain itu, pengukuran Rasch menilai beberapa ciri psikometrik lain termasuklah unidimensi(unidimensionality), hierarki item (item hierarchy), kebolehpercayaan individu (person reliability) dan statistik individu pemisah (person separation statistic). Oleh kerana kurangnya kajian terhadap analisis item soalan Matematik menggunakan model Rasch, maka pengkaji menggunakan pendekatan model ini untuk mengenalpasti ciri psikometrik yang terdapat di dalam soalan Ujian Akhir Sesi Akademik 2022 Matematik tingkatan satu. Kajian ini turut penting untuk memaparkan sejauh mana kebolehan matematik murid bagi keseluruhan kandungan Matematik Tingkatan 1 digambarkan dalam satu skala yang sama melalui penggabungan ujian yang dijalankan. Maklumat ini penting untuk memberi maklumat kepada guru dan murid tentang bidang pembelajaran, bab dan standard kandungan yang mudah atau sukar dikuasai. Hal ini dapat membantu guru dalam merancang pengajaran di bilik darjah dan membantu murid memberi penekanan kepada bahagian yang sukar dikuasai oleh mereka. Di

samping itu, maklumat item Matematik Tingkatan 1 merangkumi semua bab dan standard kandungan yang digambarkan dalam satu skala ini turut mampu memberikan bukti empirikal yang berguna kepada penggubal kurikulum yang telah menyusun Dokumen Standard Kurikulum dan Pentaksiran (DSKP) Matematik Tingkatan 1. Secara teorinya, kajian ini juga penting untuk membuktikan penggunaan model Rasch daripada Teori Respons Item (IRT) mampu mencirikan item dengan baik.

II. OBJEKTIF KAJIAN

Kajian ini adalah bertujuan untuk menyemak ciri psikometrik soalan Matematik melalui perspektif pengukuran Rasch. Pengkaji membincangkan bagaimana model Rasch boleh berguna kepada pendidik untuk memahami ciri psikometrik dalam instrumen. Kajian ini akan menjawab soalan-soalan berikut:

1. Adakah item ujian Matematik Tingkatan 1 sepadan dengan model Rasch?
2. Sejauh manakah item ujian Matematik Tingkatan 1 mampu menghasilkan skor pelajar secara konsisten?
3. Sejauh manakah item membezakan pelajar mengikut kebolehan mereka?

III. METODOLOGI KAJIAN

Dokumen yang digunakan adalah soalan ujian matematik. Soalan ini merupakan soalan ujian matematik berpusat yang disediakan oleh Kementerian Pelajaran Malaysia (KPM) untuk ujian akhir sesi akademik. Kertas soalan ini mengandungi 20 soalan aneka pilihan meliputi setiap topik dalam silibus tingkatan satu. Walaubagaimanapun, item A18 disingkirkan kerana mempunyai ralat. Kajian ini bertujuan melihat kesesuaian item penilaian akhir yang ditadbir kepada pelajar tingkatan satu sesi 2022. Seramai 109 sampel murid terlibat dalam analisis ini. Jadual 1 merujuk demografi responden yang dilibatkan dalam kajian ini yang dipilih berdasarkan pensampelan bertujuan.

Jantina	Bilangan	Peratus %
Lelaki	38	35
Perempuan	71	65

Analisis deskriptif terhadap item ujian adalah berpandukan pengukuran model Rasch menggunakan perisian Winstep 3.71.0.1. Kajian ini merupakan kajian kuantitatif deskriptif adalah berdasarkan kerangka pengukuran Rasch melalui analisis unidimensi, kesesuaian item, polariti item, indeks pemisahan dan kebolehpercayaan item dalam menentukan kesepadanan item yang

dilibatkan sebagai ujian akhir tersebut. Respons pelajar bagi setiap item akan direkodkan dalam SPSS dan dianalisis berdasarkan prosedur analisis separa kredit model Rasch. Transformasi data adalah dalam unit logits yang mengandungi maklumat kesukaran item. Analisis item akan menggunakan tetapan statistik daripada Jadual 2 sebagai panduan bagi interpretasi dapatan analisis.

Jadual 2 Analisis model Rasch dan tetapan ukuran

Kriteria	Tetapan	Sumber
Kesepadanan Item (<i>Item Fit</i>)		
Infit	0.7 – 1.3	Bond & Fox (2001)
Outfit	0.7 – 1.3	Bond & Fox (2001)
Unidimensionaliti		
PCA	> 20%	Azrilah (2017)
Jangkaan model	> 20%	Linacre (2017)
Unexplained variance in 1st contrast	< 10%	Linacre (2007)
Nilai Eigen	< 5	Linacre (2005)
Polariti Item	> 0	Linacre (2012)

IV. DAPATAN DAN PERBINCANGAN KAJIAN

A. Unidimensi

Analisis statistik item dan model Rasch perlu dipenuhi bagi menjamin pengukuran yang sah. Analisis pertama adalah melalui prosedur *Principal Component Analysis* (PCA) bagi mengesan unidimensi ujian. Unidimensi ujian yang baik menunjukkan keseragaman elemen yang dilibatkan dalam pengukuran adalah menjurus kepada pengukuran satu konstruk atau satu elemen. Instrumen yang tidak mencapai tahap unidimensi akan menyebabkan kekeliruan kepada responden dan berkemungkinan mengukur dimensi lain yang tidak dijangkakan. Berdasarkan analisis PCA elemen yang perlu diberi perhatian adalah daripada nilai eigenvalue unit bagi *Raw variance explain by*

measure dan *Unexplained variance in 1st contrast*. Menurut Azrilah et al., model Rasch memerlukan nilai *Raw variance explain by measure* melebihi 40% bagi menjamin pengukuran menepati keperluan model namun kadar 20% adalah memadai kerana tidak melemahkan pengukuran yang dibuat [8]. Nilai *Unexplained variance in 1st contrast* juga seharusnya tidak melebihi nilai kawalan 15%. Jadual 3 menunjukkan varians residual bagi UASA Matematik Tingkatan 1. Nilai *Raw variance explain by measure* adalah 31.1% dan nilai eigen tersebut adalah sesuai kerana berada di dalam julat pengukuran. Nilai *Unexplained variance in 1st contrast* juga adalah baik dengan nilai eigen adalah 1.9. Nilai faktor yang kecil menunjukkan kesahan pengukuran yang unidimensi [9] dan mengesahkan dimensi kedua tidak wujud dengan jelas.

Jadual 3 Analisis Unidimensi

	Nilai Eigen	Empirical	Modeled
Total raw variance in observations	27.6	100.0%	100.0%
Raw variance explained by measures	8.6	31.1%	30.9%
Raw variance explained by persons	2.7	9.7%	9.6%
Raw Variance explained by items	5.9	21.4%	21.3%
Raw unexplained variance (total)	19.0	68.9%	69.1%
Unexplained variance in 1st contrast	1.9	7.0%	

B. Kesepadanan Item

Dalam penentuan kesesuaian item, perhatian akan diberikan kepada beberapa analisis iaitu, kesukaran item, infit/outfit MNSQ serta polariti item. *Measure* dalam Winstep merujuk andaian tentu ukuran parameter seperti kesukaran item dan dilaporkan dalam unit logits. Semakin tinggi nilai *measure* semakin sukar item tersebut dan sebaliknya. Jadual 4 menunjukkan taburan julat kesukaran item bagi kertas UASA matematik tingkatan 1. Nilai kesukaran terendah adalah pada julat -6.90. Analisis kesepadanan ini bertujuan untuk memastikan item yang dibina memenuhi ciri-ciri kesepadanan yang ditetapkan oleh Model Rasch. Sebagai suatu model bercirikan kebarangkalian, model Rasch menganggap nilai *mean square* (MNSQ) infit dan outfit bersamaan atau menghampiri 1 dianggap mematuhi model Rasch.

Julat kesepadanan yang ditetapkan adalah berdasarkan nilai infit MNSQ dan outfit MNSQ ($0.7 < x < 1.3$) [10]. Namun indeks outfit MNSQ perlu diberi perhatian terlebih dahulu berbanding infit bagi menentukan kesesuaian item yang mengukur sesuatu konstruk atau *latent variable* [8]. Sekiranya nilai MNSQ lebih daripada 1.5 logits, maka ia memberi maksud item mengelirukan. Jika nilai MNSQ kurang daripada 0.5 logits pula ia menunjukkan item terlalu mudah dijangka oleh individu [11]. Selain itu, nilai outfit ZSTD dan infit

ZSTD perlu juga berada dalam lingkungan -2 hingga +2 [12], namun jika nilai outfit dan infit MNSQ diterima, indeks ZSTD boleh diabaikan [13]. Jika syarat ini tidak ditepati, maka item boleh dipertimbangkan untuk disingkirkan atau dimurnikan. Untuk data dari ujian aneka pilihan yang memenuhi andaian nilai MNSQ di antara 0.8 dan 1.2 boleh diterima [14]. Nilai-nilai MNSQ di bawah 0.7 menunjukkan jawapan -jawapan terlalu konsisten. Nilai- nilai MNSQ yang melebihi 1.3 menunjukkan jawapan-jawapan adalah terlalu rambang daripada yang diramal oleh model.

Semua item menunjukkan nilai infit MNSQ adalah antara 0.75 hingga 1.13 dan outfit MNSQ adalah antara 0.49 hingga 1.72. Walaubagaimanapun, item 7 dan 20 tidak memenuhi outfit yang ditetapkan. Ini menunjukkan item ini mempunyai masalah dan harus diperbaiki kerana item terlalu mudah dijangka oleh individu [11]. Item lain dapat mengukur konstruk yang sama dan menghasilkan pengukuran yang bersifat *additive* (*additivity of measures*) dan kebarangkalian pelajar jawab dengan betul terhadap sesuatu item tidak bergantung kepada item-item lain dalam penilaian ini [15].

Selanjutnya, analisis polariti item merupakan analisis awal untuk memastikan kesemua item bergerak sehalu secara positif untuk mengukur pemboleh ubah yang dikaji [16]. Analisis ini dijalankan untuk mengukur hubungan antara

item dalam mengukur setiap konstruk. Jika nilai PTMEA adalah tinggi, bermaksud item tersebut berupaya untuk membezakan kemampuan antara individu dan mengukur konstruk yang ingin diukur. Sebaliknya jika nilai adalah negatif, item yang dibangunkan tidak mengukur konstruk yang ingin diukur [17] di mana pembina soalan harus memeriksa semula item sama ada item perlu diperbaiki atau disingkirkan kerana indicator ini menunjukkan terdapatnya item atau individu yang memberi respon yang bercanggah dengan pembolehubah [17].

Jadual 4 menunjukkan semua item berada di dalam julat (0.24 hingga 0.56). Polariti item ujian

yang positif menunjukkan keserasian item adalah sejajar dengan kebolehan pelajar namun bagi kesahan konstruk, polariti item seharusnya memiliki nilai korelasi >0.3 logits. Nilai polariti item (PTMEA) yang diperolehi sebaiknya memiliki nilai positif dan melebihi 0.3 logits [4]. Namun demikian, item 7(0.28), item 11(0.29), item 13(0.24) dan item 20(0.25) menunjukkan nilai polariti dibawah 0.30 yang menunjukkan item tidak memenuhi kriteria [18]. Walaubagaimanapun, Linacre berpendapat, nilai positif adalah memadai kerana ianya menunjukkan item yang ditadbir adalah seiring dengan kebolehan pelajar [19].

Jadual 4 Analisis Kesepadanan Item

ENTRY	MEASURE	INFIT	ZSTD	OUTFIT	ZSTD	PTMEA
A3	-2	0.88	-1.05	0.85	-0.72	0.49
A14	-1.88	0.97	-0.24	0.97	-0.09	0.42
A6	-1.53	0.99	-0.07	0.96	-0.21	0.42
A2	-1.06	0.96	-0.44	0.91	-0.71	0.46
A4	-1.01	1.03	0.37	1.06	0.54	0.4
A19	-0.79	0.96	-0.45	1	0.01	0.45
A9	-0.56	1	0.05	1.03	0.31	0.41
A8	-0.47	1.08	0.96	1.06	0.53	0.36
A1	-0.24	1.11	1.28	1.18	1.31	0.31
A12	-0.1	1.13	1.39	1.12	0.88	0.31
A16	0.3	0.98	-0.19	0.91	-0.44	0.42
A10	0.47	1	0.02	1	0.06	0.39
A15	0.47	0.98	-0.1	0.99	0.01	0.4
A5	0.83	0.96	-0.23	0.78	-0.87	0.43
A17	0.83	0.75	-1.87	0.72	-1.14	0.56
A13	1.1	1.1	0.66	1.21	0.75	0.24
A11	1.18	1.08	0.53	0.99	0.07	0.29
A20	1.71	0.95	-0.14	1.72	1.57	0.25
A7	2.75	0.97	0.04	0.49	-0.69	0.28

C. Peta Item-Individu

Pemetaan item-individu bercirikan satu bentuk taburan kesukaran item yang disepadankan dengan taburan kebolehan murid di sepanjang skala logits pada satu kontinum pengukuran dari aras paling mudah kepada paling sukar secara visual [12]. Kebolehan pelajar dinilai dalam logit, iaitu unit yang ditukar dari skala ordinal kepada skala selang [4]. Jika item tidak diukur dalam skala logit dan hirarki item, skor mentah yang diperolehi pelajar tidak dapat membezakan antara pelajar yang lain. Maka, peta pelajar-item (Wright map) memberi gambaran pengukuran terhadap item mahupun kebolehan pelajar kerana kesukaran item dan kebolehan pelajar dibandingkan secara langsung. Keistimewaan ini dikenali sebagai *Mapping* di

mana kebolehan pelajar dan kesukaran item diwakili oleh graf yang diberi nama *item-by-person map*. Kedua-duanya iaitu item dan individu akan diukur dalam skala logit yang sama, dan juga menunjukkan samada item sesuai dengan kebolehan pelajar. Individu dan item diletakkan secara bersebelahan dan dipisahkan dengan satu garisan putus-putus. Di bahagian kiri garisan adalah kedudukan individu, manakala di sebelah kanan garisan adalah kedudukan item. Jika min bagi kebolehan individu dan kesukaran item adalah sama, maka item-item dalam instrumen adalah berpadanan dengan persepsi sampel kajian [20].

Tujuan pemetaan adalah untuk menunjukkan hubungan antara keupayaan pelajar dengan aras kesukaran item [4]. Pemetaan menunjukkan item yang sangat mudah dipersetujui

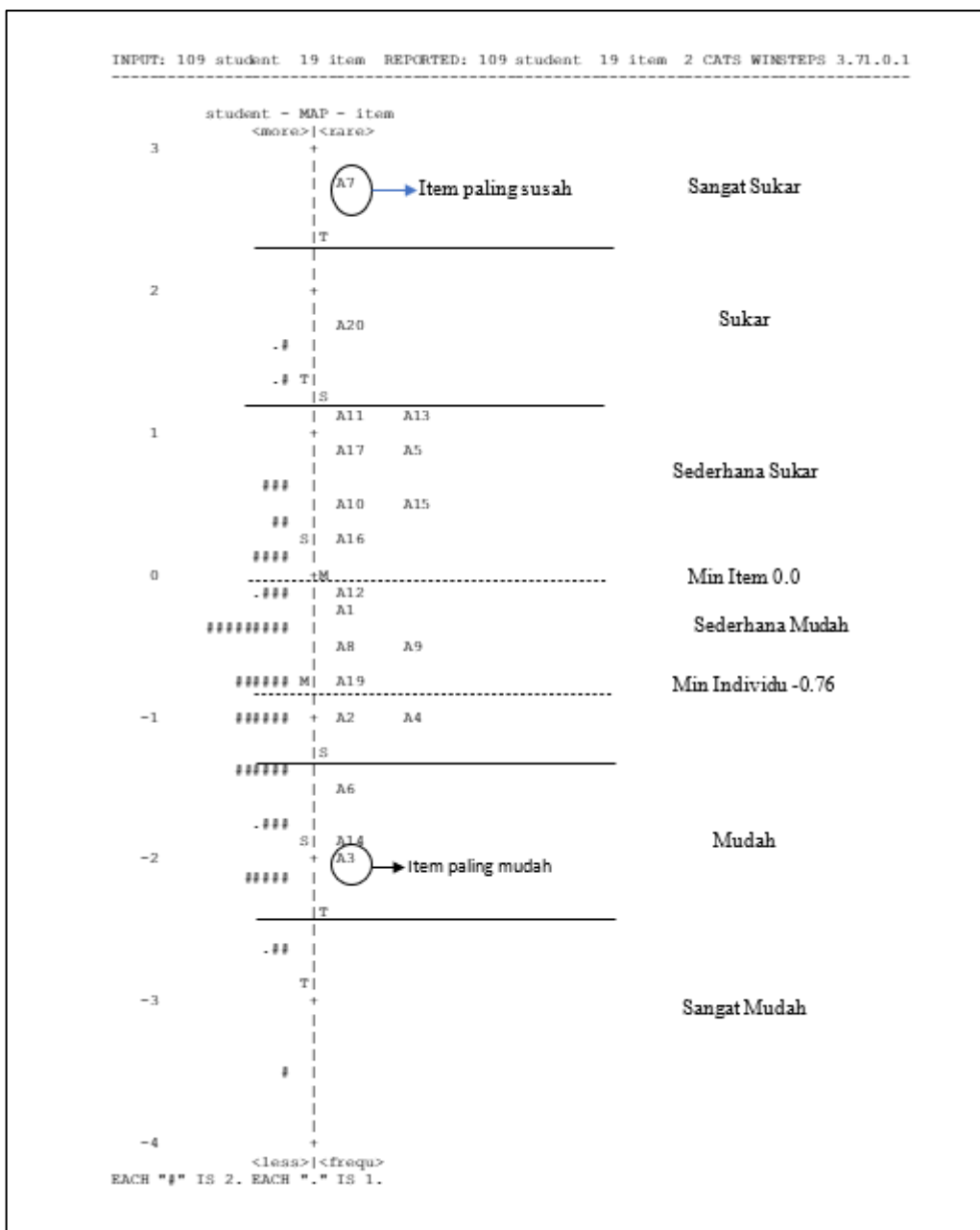
pada kedudukan kanan di bawah peta manakala item yang sangat sukar dipersetujui pada kedudukan kanan di atas peta. Seterusnya, kedudukan bilangan responden yang berkebolehan tinggi adalah di bahagian kiri atas peta dan bilangan responden berkebolehan rendah pada bahagian kiri bawah peta [15]. Lompang pada kalangan taburan item turut dikaji bagi menguji kecukupan item bagi semua tahap kebolehan individu. Apabila item dan kebolehan pelajar berada dalam linear kontinum yang sama, pendidik mempunyai maklumat yang lebih tepat keadaan pelajar di dalam ukuran. Dalam peta tersebut, kedudukan pelajar dan item diletakkan secara bersebelahan dan dipisahkan dengan satu garisan putus. Garisan putus di tengah menggambarkan lokasi ukuran kebolehan individu dan ukuran kesukaran item. Dua huruf M pada garisan putus adalah menunjukkan ukuran kebolehan min(purata) bagi individu (kiri) dan item(kanan). Manakala huruf S di bahagian kiri adalah bermaksud satu sisihan piawai bagi individu dan di bahagian kanan pula menunjukkan satu sisihan piawai bagi item. Huruf T pula adalah menunjukkan dua sisihan piawai bagi individu (kanan) dan dua sisihan piawai bagi item (kiri).

Merujuk kepada Rajah 1, kesemua item berada pada hirarki yang tersusun bermula dari log -2 sehingga log 2.75. Item dikelompokkan kepada enam aras kesukaran yang dibezakan berdasarkan nilai sisihan piawai pembaris logit Rasch. Item yang berada di atas min item dikategorikan sebagai sangat sukar, sukar dan sederhana sukar. Manakala item yang berada di bawah min item dikategorikan sebagai item sederhana mudah, mudah dan sangat mudah. Sembilan item berada di atas min kesukaran item. Secara empirikal, hanya satu item dalam kategori sangat sukar (5%), 1 item (5%) berada dalam kategori sukar, 7 item (37%) berada dalam kategori sederhana sukar, 7 item (37%) berada dalam kategori sederhana mudah, 3 item (16%) berada dalam kategori mudah dan tiada item dikategorikan sangat mudah. Menurut sistem nisbah membina Jadual Spesifikasi Ujian (JSU) nisbah aras kesukaran yang ditetapkan ialah 3:5:2 (senang: sederhana: susah) di mana soalan aras sederhana adalah lebih banyak berbanding soalan aras mudah dan susah. Merujuk kepada analisis yang dijalankan terhadap nisbah item ujian UASA adalah 3:14:2 di mana pemberatan lebih kepada soalan aras sederhana.

Selanjutnya, Rajah 1 memaparkan item 3 adalah item paling senang dijawab oleh pelajar dan item 7 merupakan item paling sukar. Rajah menunjukkan tiada pelajar berkebolehan sangat tinggi sebaliknya ada 6 orang pelajar berkebolehan sangat rendah. Didapati min item pada logit 0 iaitu lebih tinggi daripada min individu, pada skala - 0.76 logit. Perbezaan ini menunjukkan kesukaran item adalah melebihi kebolehan pelajar. Secara empirikalnya, soalan UASA matematik didapati

susah berdasarkan kebolehan pelajar. Pengkaji berpendapat item ini tidak dijawab dengan baik oleh pelajar dengan beberapa alasan iaitu: Matematik bukanlah subjek yang menjadi pilihan kerana sikap pelajar yang negatif terhadap sesuatu mata pelajaran seperti mengantuk, ponteng kelas semasa proses pengajaran dan pembelajaran berlangsung [21], pelajar kurang berkebolehan membuat koordinasi mental menggunakan nombor dan konsep [22] dan pelajar sukar mempelajari Matematik yang dikaitkan dengan perkataan berbanding Matematik yang melibatkan angka dan symbol [23].

Bagi jurang antara item, dua jurang item melebihi .5 logits dapat dikesan iaitu antara item 7 (2.75 logits) dan item 20 (1.71 logits) dengan jurang item = 1.04 logits, dan antara item 20 (1.71 logits) dan item 11 (1.18 logits) dengan jurang item = 0.53 logits. Menurut Boone, jurang antara item tersebut dapat direndahkan melalui pengguguran item atau penggantian kepada item yang lebih baik [24]. Menambahkan item bagi memenuhi jurang antara item atau meminimumkan jurang adalah penting agar pengukuran terhadap kebolehan pelajar dapat dibuat dengan lebih jelas [25].



Rajah I Peta Item-Individu

V. PENUTUP

Statistik kesukaran item merujuk tahap kebolehan pelajar menjawab item yang ditadbir. Item sukar hanya dapat dijawab oleh pelajar yang berkebolehan tinggi berbanding item mudah yang mampu dijawab pelajar berkebolehan rendah dengan betul. Item yang tidak dapat membezakan antara pelajar merupakan item yang tidak sesuai bagi mengukur kebolehan pelajar dan seharusnya digugurkan atau diubahsuai. Pengukuran kualiti ujian bergantung kepada kualiti item yang dilibatkan dalam proses tentu ukuran. Model pengukuran Rasch menyediakan analisis psikometrik item yang baik dalam menentukan

kualiti setiap item. Dapatan akhir menunjukkan bahawa sebahagian besar item adalah memenuhi andaian pengukuran Rasch. Item-item yang telah ditentu ukur boleh dilibatkan semula sebagai item ujian yang baik atau digunakan dalam kajian lanjutan sekiranya diperlukan. Justeru, 20 item per ujian dilihat tidak cukup meminimumkan jurang item dalam padanan item-pelajar. Keperluan ubahsuai item atau perubahan jumlah item dilibatkan dalam ujian boleh dipertimbangkan.

Selain itu, analisis item membantu memperbaiki ujian dan menyalakan item yang tidak efektif di samping mendiagnosis maklumat yang pelajar tahu dan tidak tahu [26]. Pelajar yang berprestasi tinggi, sederhana dan rendah juga dapat dikategorikan mengikut kelompok. Maka, analisis



statistik ini, membantu guru mengenal pasti kelemahan dan kekuatan pelajar dalam sesuatu topik yang diwakili oleh item tersebut. Maka perancangan berarah mengikut kelompok boleh dilakukan untuk meningkatkan kefahaman pelajar terhadap sesuatu subjek yang dipelajari. Melalui analisis statistik ini, pendidik mendapat maklumat mengenai item yang bermutu dan kebolehan pelajar (trait pendam) berdasarkan pola respons yang diberi. Dapatan kajian ini juga penting untuk membolehkan pihak ibu bapa dan pentadbir sekolah bekerjasama dan merancang pendekatan yang sesuai untuk membantu pelajar menguasai mata pelajaran matematik kerana masih terdapat keperluan dalam pendidikan matematik untuk anjakan paradigma yang membolehkan potensi bermakna matematik menjadi lebih nyata kepada pelajar.

RUJUKAN

- [1] N.E . Gronlund, *Constructing Achievement Test.*, 3rd ed. New Jersey: Prentice-Hall, 1982.
- [2] M. S. Nordin, "Pengujian dan Penaksiran di Bilik Darjah.," Kuala Lumpur, 2002.
- [3] G. Sax, *Principles of Educational and psychological measurement and evaluation.*, 3 rd. Belmont, California: Wadworth Publishing, 1989.
- [4] T. G. Bond and C. M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.* New Jersey: Routledge, 2007.
- [5] J. L. Pomeranz, K. L. Byers, M. D. Moorhouse, C. . Velozo, and R. J. Spitznagel, "Rasch Analysis as a Technique to Examine the Psychometric Properties of a Career Ability Placement Survey Subtest.," *Rehabil. Couns. Bull.*, vol. 51, pp. 251–259, 2008.
- [6] G. Erkin, A. H. Elhan, C. Aybay, H. Sirzai, and S. Ozel, "Validity and Reliability of the Turkish translation of the pediatric Evaluation of Disability Inventory (PEDI).," *Disabil. Rehabil.*, vol. 29, no. 16, pp. 1271–1279, 2007.
- [7] R. K. Hambleton and H. J. Rogers, "Detecting potentially biased test items: Comparison of IRT area and Mantel-Haenszel methods.," *Appl. Meas. Educ.*, vol. 2, no. 4, p. 313, 1989, [Online]. Available: https://doi.org/10.1207/s15324818ame0204_4
- [8] A. A. Aziz, M. S. Masodi, and A. Zaharim, *Asas Model Pengukuran: Pembentukan Skala & Struktur Pengukuran (Ketiga).* Penerbit UKM, 2017.
- [9] J. Sick, "Rasch Measurement in Language Education Part 5: Assumptions and requirements of Rasch measurement.," in *Shiken: JALT Testing & Evaluation SIG Newsletter*, 14(October), 2010, pp. 23–29.
- [10] T. G. Bond and C. M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.*, Pertama. Mahwah: Erlbaum, 2001.
- [11] J. M. Linacre, *A User's Guide to WINSTEPS: Ministep Rasch Model Computer Programs.* Chicago: MESA Press, 2007.
- [12] T. G. Bond and C. M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.*, Third. Taylor & Francis: Routledge, 2015. [Online]. Available: <https://doi.org/10.4324/9781410614575>
- [13] J. M. Linacre, "Standard errors: means, measures, origins and anchor values.," *Rasch Meas. Trans.*, vol. 19, no. 3, p. 1030, 2005.
- [14] J. M. Linacre, "Investigating judge local independence.," *Rasch Meas. Trans.*, vol. 11, no. 1, pp. 546–547, 1997.
- [15] S. R. Ariffin, T. S. M. M. N. Othman, and I. M. Zin, "Analisis Kecerdasan pelbagai di kalangan pensyarah di sebuah institusi pengajian tinggi.," *J. Teknol.*, vol. 41, pp. 33–42, 2004.
- [16] J. M. Linacre, *Winsteps Rasch Measurement Computer Program User's Guide.* Beaverton, OR: Winsteps.com, 2015.
- [17] J. M. Linacre, "Winsteps Computer Program Version 3.57.2.," 2003. <http://www.winsteps.com> (accessed Apr. 24, 2016).
- [18] J. C. Nunnally and I. H. Bernstein, *Psychometric Theory.*, Ke-3. New York: McGraw Hill, 1994.
- [19] J. M. Linacre, *A user's guide to WINSTEPS: Rasch Model Computer Programs.* Chicago: MESA Press, 2012.
- [20] B. D. Wright and G. N. Masters, *Rating scale analysis in Rasch measurement.* Chicago: MESA Press, 1982.
- [21] K. Osman, Z. haji Iksan, and L. Halim, "Sikap terhadap Sains dan Sikap Saintifik di

- kalangan pelajar sains,” *J. Pendidik.*, vol. 32, pp. 39–60, 2007.
- [22] Z. A. Bakar and M. R. B. M. Jalil, “Keberkesanan Kaedah petak Sifir Dalam Penguasaan Fakta Asas Darab Dalam Matematik Tahun 4: Satu Kajian di Sekolah Kebangsaan Mersing, Johor.,” Universiti teknologi Malaysia, 2010.
- [23] B. T. Lim, “Penyelidikan mengenai jenis kesilapan dalam menyelesaikan masalah Matematik berayat bagi pelajar tingkatan dua.,” Universiti Teknologi Malaysia, 2000.
- [24] W. J. Boone, “Rasch analysis for instrument development: Why,when,and how?,” *CBE Life Sci. Educ.*, vol. 15, no. 4, pp. 1–7, 2016, [Online]. Available: <https://doi.org/10.1187/cbe.16-04-0148>
- [25] M. A. Samsudin, T. S. Chut, M. E. Ismail, and N. J. Ahmad, “A Calibrated Item Bank for Computerized Adaptive Testing in Measuring Science TIMSS Performance.,” *Eurasia J. Math. Sci. Technol. Educ.*, vol. 16, no. 7, p. 1863, 2020, [Online]. Available: <https://doi.org/10.29333/ejmste/8259>
- [26] A. L.R and G. G, *Psychological Testing and Assessment.*, 12th ed. Boston: Pearson Education, 2006.

AUTHOR’S INFORMATION

<p>First Author: Saralah Sovey</p> 	<p>SMK Sungai Ramal, Kajang Selangor, Malaysia</p> <p>E-mail: saralahsovey@gmail.com</p>
<p>Second Author: Mohd Effendi Ewan Mohd Matore</p> 	<p>Research Centre of Education Leadership and Policy, Faculty of Education, The National University of Malaysia, UKM Bangi, Selangor, Malaysia</p> <p>E-mail: effendi@ukm.edu.my</p>