

## **Menentukan Kesahan Dan Kebolehpercayaan Item Pentaksiran Mata Pelajaran Rombak Rawat Sistem Klac Dengan Menggunakan Model Pengukuran Rasch**

Muhammad Amirul Abdullah  
Politeknik Mersing  
amirul\_kkbd@yahoo.com

Turina Tumeran  
Politeknik Mersing  
turinatumeran@yahoo.com

### **Abstract**

*This study was carried out to determine the validity and reliability of assessment items used in Continuous Assessment in the “Rombak Rawat Sistem Klac” subject. Item analysis is conducted to assess the psychometric characteristics of assessment items in order to be used as a valid assessment item with high consistency. All dichotomous items were analyzed based on Rasch Measurement Model based on Item Response Theory (IRT) using Winstep software. This quantitative study was administered pioneering to 65 respondents who attended the “Sijil Servis Kenderaan Ringan” at Bandar Darulaman Community College. The findings show that 2 items show negative PTMEA-Corr values, where those items do not measure the topic or there are respondents who respond to the contradictory topic. Meanwhile, 6 items detected are outside the range of MNSQ *infit* and *outfit* values. The overall analysis results show that 7 out of 30 assessment items are proposed to be repaired or removed as an assessment items for the “Rombak Rawat Sistem Klac” subject. Good psychometric characteristics in an assessment items after going through an items analysis process can be used as a valid instruments for that subject.*

**Keywords:** rombak rawat sistem klac, item response theory, model rasch

### **Abstrak**

Kajian ini dijalankan bagi menentukan kesahan dan kebolehpercayaan item-item pentaksiran yang digunakan dalam Penilaian Berterusan (PB) dalam mata pelajaran Rombak Rawat Sistem Klac. Analisis item dijalankan untuk membuat penilaian ke atas ciri-ciri psikometrik item-item pentaksiran supaya dapat digunakan sebagai item pentaksiran yang sah dan mempunyai ketekalan yang tinggi. Item-item pentaksiran yang berbentuk dikotomi dianalisis berasaskan Model Rasch berdasarkan *Item Response Theory* (IRT) dengan menggunakan perisian Winstep. Kajian berbentuk kuantitatif ini ditadbir secara rintis kepada 65 responden yang mengikuti Kursus Sijil Servis Kenderaan Ringan di Kolej Komuniti Bandar Darulaman. Dapatan kajian mendapati 2 item menunjukkan nilai *PTMEA-Corr* negatif, di mana item-item tersebut tidak mengukur topik yang berkenaan atau terdapat responden yang memberi respon yang bercanggah dengan topik yang diukur. Sementara itu, 6 item dikesan berada di luar julat nilai *MNSQ infit* dan *outfit*. Hasil analisis keseluruhan mendapati 7 daripada 30 item pentaksiran dicadangkan untuk dibaiki atau disingkirkan sebagai item pentaksiran mata pelajaran Rombak Rawat Sistem Klac. Ciri-ciri psikometrik yang baik dalam sesuatu item pengukuran setelah melalui proses analisis item boleh digunakan sebagai item pentaksiran bagi PB mata pelajaran berkenaan.

**Kata kunci:** Rombak Rawat Sistem Klac, Item Response Theory, Model Rasch

### **1.0 Pengenalan**

Salah satu pertimbangan yang paling utama dalam pembangunan sesuatu ujian ialah memastikan ujian yang dibina tersebut bersifat unidimensi. Unidimensi bermaksud sesuatu item yang dibina perlulah

mengukur satu keupayaan tunggal (Siti Rahayah, 2008). Ciri-ciri unidimensi ini kerap kali diabaikan dalam pembinaan sesuatu instrumen. Akibatnya, semasa pengukuran terhadap sesuatu konstruk, didapati dapataannya menunjukkan pada tahap yang rendah. Ini disebabkan oleh sesuatu konstruk yang diukur tersebut dipengaruhi oleh faktor-faktor lain. Oleh itu, isu unidimensi dalam sesuatu item perlu disokong dengan keteguhan model pengukuran yang digunakan dalam sesuatu kajian. Walaupun prosedur-prosedur penentuan unidimensi dalam sesuatu item agak kompleks dan rumit, namun, ianya tidak wajar diabaikan. Model Rasch menyediakan prosedur ringkas untuk tujuan menyiasat atau menentukan ciri-ciri unidimensi sesuatu item dalam ujian atau instrumen (Linacre, 2007; Bond, & Fox, 2007).

Dalam Model Rasch, Wright dan Stone (1979) menegaskan bahawa kesan kualiti ujian boleh ditunjukkan apabila empat syarat berikut telah dipatuhi: 1) penggunaan item yang sah untuk mentakrifkan konstruk; 2) definisi konstruk yang jelas untuk diukur dan selaras dengan jangkaan-jangkaan teori; 3) keupayaan item yang konsisten dengan tujuan pengukuran; dan 4) penggunaan pola jawapan yang sah. Bagi memastikan item yang sah dalam ujian, tiga garis panduan turut disenaraikan iaitu: 1) memastikan bahawa item bekerja dalam arah yang sama untuk mengukur konstruk; 2) menjamin sesuatu item itu menyumbang dalam cara-cara yang bermakna dan berguna untuk mengukur konstruk; dan 3) memastikan bahawa ujian secara keseluruhannya adalah mengukur konstruk.

Dalam Model Pengukuran Rasch, kebolehpercayaan individu dan item dapat menunjukkan sejauh mana keserasian item kepada model Rasch dan indeks pemisahan individu dan item. Nilai kebolehpercayaan melebihi 0.8 adalah pada tahap yang baik (Bond, & Fox, 2007).

Tujuan kajian ini adalah untuk menentukan kesahan dan kebolehpercayaan item-item pentaksiran yang digunakan dalam Penilaian Berterusan (PB) dalam mata pelajaran Rombak Rawat Sistem Klac. Analisis item dijalankan untuk membuat penilaian ke atas ciri-ciri psikometrik item-item pentaksiran supaya dapat digunakan sebagai item pentaksiran yang sah dan mempunyai ketekalan yang tinggi.

## **2.0 Permasalahan**

Pembangunan sesuatu instrumen atau ujian melibatkan proses pembinaan item dan proses pengukuran. Isu-isu psikometrik seperti kebolehpercayaan dan kesahan serta kualiti item-item merupakan perkara yang paling penting untuk diberi perhatian (Ruhizan Mohd. Yasin, Faizal Amin Nur Yunus, Ridzwan Che Rus, Azmi Ahmad, dan Mohd Bekri Rahim, 2015)). Siti Rahayah Ariffin, Bishanani Omar, Anita Isa dan Sharida Sharif (2010) menyatakan bahawa skor ujian lazimnya digunakan untuk menghasilkan sesuatu laporan prestasi atau dapatan hasil sesuatu ujian. Ini bermakna skor ujian sering dijadikan satu gambaran terhadap keupayaan seseorang individu tertentu. Namun demikian, amalan tersebut kurang tepat kerana keupayaan individu dalam peperiksaan tidak boleh dianggarkan dengan tepat melalui skor ujian. Ini kerana skor sesuatu ujian tidak dapat menentukan perbezaan di antara setiap keupayaan individu yang menduduki ujian tersebut (Bond, & Fox, 2007).

Dalam sesuatu ujian, skor individu dan dan skor item-item perlu diberi perhatian serius dalam peperiksaan. Skor individu dalam peperiksaan merujuk kepada peratus jawapan yang betul, manakala skor item-item pula merujuk kepada indeks kesukaran (nilai- $p$ ). Skor individu dalam peperiksaan yang berprestasi rendah dibandingkan dengan individu yang berprestasi tinggi. Sebagaimana skor item, indeks kesukaran yang lebih rendah menunjukkan item yang mudah, manakala indeks kesukaran yang lebih tinggi menunjukkan item yang sukar. Masters dan Wright (1984) menegaskan bahawa adalah sangat penting agar skor individu dalam peperiksaan dapat diterjemahkan dengan tepat menggunakan skor item.

Walaupun skor ujian digunakan dengan meluas sebagai ukuran keupayaan, skor ujian individu dalam peperiksaan mempunyai beberapa kekurangan kerana ia berada dalam skala bilangan yang tidak mempunyai interval yang sama (Siti Rahayah Ariffin, 2008). Bond dan Fox (2007) menyatakan bahawa skor ujian boleh menganggangkan keupayaan individu secara hierarki, namun ianya tidak dapat menentukan bagaimana keupayaan ini berbeza daripada individu yang lain. Skor ujian tidak dapat membezakan dengan tepat antara individu yang lebih berkebolehan dengan yang kurang berkebolehan. Ianya hanya boleh menunjukkan bahawa pertambahan skor ujian dari 50 kepada 55 tidak serupa dengan pertambahan daripada 5 kepada 10 atau 85 kepada 90. Perbezaan dalam kebolehan antara dua skor, umpamanya 40 dan 45 tidak sama dengan perbezaan di antara 25 dan 30. Perbezaan di antara dua skor turut tidak boleh dipercayai apabila skema penskoran yang berbeza digunakan.

Oleh itu, adalah sangat penting agar skor ujian diubah kepada skala interval yang sama, di mana: “aritmetik yang boleh dibuat, satu nombor yang boleh ditambah dan ditolak, dan perbezaan yang mana boleh digandakan dan dibahagikan dengan keputusan yang mengekalkan makna numeriknya” (Masters dan Wright, 1984). Perubahan ini dipanggil sebagai kalibrasi ujian, di mana parameter item (kesukaran item) dan parameter individu dalam peperiksaan (kebolehan) dianggarkan supaya mereka boleh diletakkan pada satu skala tunggal. Prosedur kalibrasi ujian ini melibatkan dua peringkat iaitu peringkat anggaran parameter item dan peringkat anggaran parameter individu. Kedua-dua peringkat ini dilakukan secara berulang kali (*iteratively*) sehingga satu ketetapan anggaran parameter yang stabil diperolehi.

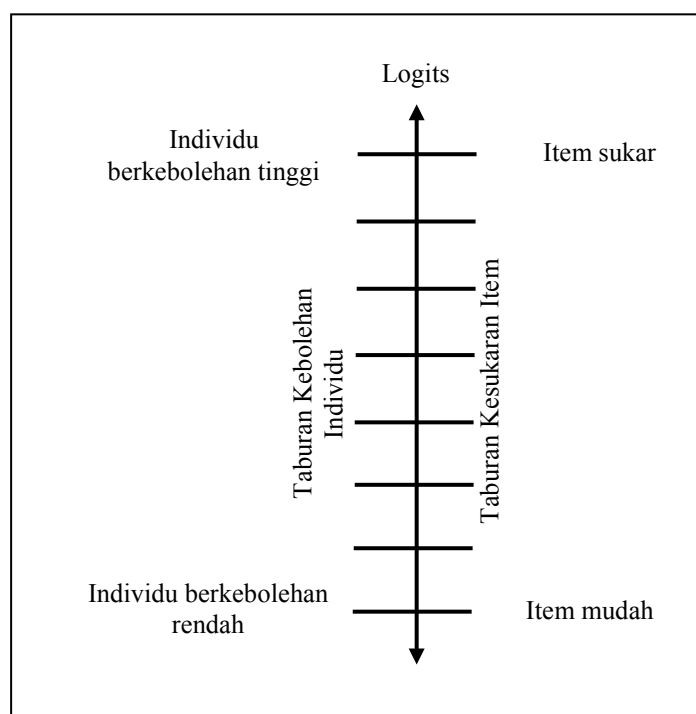
Kesukaran item ialah kesukaran yang terlindung dikira daripada bilangan individu yang berjaya dalam item tersebut. Kebolehan individu (ukuran individu) ialah anggaran kepada kebolehan individu yang terlindung berasaskan pencapaian individu dalam suatu tentuan item. Ianya dikira daripada jumlah keseluruhan item di mana individu tersebut berjaya menjawab dalam ujian. Kesukaran item dan keupayaan individu dalam peperiksaan diukur menggunakan satu unit logit. Log natural yang menghuraikan jumlah item yang betul berbanding jumlah yang tidak betul dipanggilkan sebagai logit (Bond & Fox, 2007).

Rajah 1 memberi gambaran berkaitan skala pengukuran daripada proses kalibrasi yang menunjukkan taburan kebolehan individu dan taburan kebolehan item. Individu berkebolehan tinggi terletak di sebelah atas (kiri) garis lurus, manakala individu berkebolehan rendah terletak di sebelah

bawah (kiri). Item yang sukar dipersetujui terletak di sebelah atas (kanan), manakala item yang mudah dipersetujui terletak di sebelah bawah (kanan).

Siti Rahayah Ariffin (2008) menyatakan bahawa oleh kerana keupayaan individu dalam peperiksaan adalah bersesuaian dengan kesukaran-kesukaran item, maka, skor yang diukur daripada ujian kalibrasi boleh digunakan untuk menganggar keupayaan individu dalam peperiksaan. Selain daripada itu, oleh kerana skor yang diukur berada dalam skala interval, di mana setiap julat adalah sama, maka, perbezaan keupayaan yang diperiksa boleh ditentukan. Dua perkara penting yang perlu dipertimbangkan ialah andaian unidimensi dan statistik kesesuaian. Unidimensi menganggap item-item dalam ujian mengukur satu keupayaan tunggal.

Model pengukuran Rasch menyediakan prosedur yang ringkas untuk mengukur unidimensi sesuatu ujian dengan mengambil kira faktor unidimensi telah dihitung oleh perisian. Seterusnya, statistik kesesuaian (*fit statistics*) dengan mengambil kira nilai purata kuasa dua (*mean-square*, MNSQ) dan statistik kesesuaian yang seragam (*standardized fit statistics*, Zstd) boleh digunakan bagi mengesan sama ada data yang terkumpul itu menunjukkan percanggahan dengan model Rasch atau sebaliknya (Linacre, 2007). Dengan kata lain, statistik kesesuaian menunjukkan sejauh mana ketepatan atau kebolehramalan data berpadanan dengan Model Pengukuran Rasch.



**Rajah 1:** Skala pengukuran daripada proses kalibrasi

Oleh itu, dapatan kajian-kajian lepas seperti kajian Ruhizan et al. (2015) yang mengkaji kesahan dan kebolehpercayaan dalam pembinaan item *Learning Transfer*, Siti Rahayah et al. (2010) mengkaji kesahan dan kebolehpercayaan item-item kecerdasan pelbagai pelajar IPT; dan pengujian instrumen Malaysian Generic Skills (MyGSI) dengan jelas membuktikan

bahawa pengujian item-item dalam sesuatu instrumen perlu dilakukan dengan teliti agar instrumen yang dibina berupaya menguji apa yang patut diuji serta sah digunapakai dalam pengujian. Maka, penambahbaikan dari semasa ke semasa dalam pengujian item ujian, item pentaksiran, serta apa jua item-item sesuatu instrumen perlu dijalankan dengan betul dan sistematik dari aspek item mahupun individu yang menjawab atau memberi respon terhadap item yang dibina. Kesimpulannya, kesahan dan kebolehpercayaan item merupakan dua perkara yang sentiasa perlu diberi perhatian dalam apa jua situasi pembinaan item-item sesuatu instrumen.

### 3.0 Objektif

Objektif kajian ini adalah seperti berikut:

- a. Menentukan kesahan item pentaksiran mata pelajaran Rombak Rawat Sistem Klac dengan menggunakan Model Pengukuran Rasch;
- b. Menentukan kebolehpercayaan item pentaksiran mata pelajaran Rombak Rawat Sistem Klac dengan menggunakan Model Pengukuran Rasch;
- c. Menentukan indeks pengasingan dalam pembinaan item pentaksiran mata pelajaran Rombak Rawat Sistem Klac dengan menggunakan Model Pengukuran Rasch;
- d. Menentukan polarity item pentaksiran mata pelajaran Rombak Rawat Sistem Klac dengan menggunakan Model Pengukuran Rasch;
- e. Menentukan kesesuaian item (*item fit*) dalam item pentaksiran mata pelajaran Rombak Rawat Sistem Klac dengan menggunakan Model Pengukuran Rasch; dan
- f. Menentukan peta item-individu (*item-person map*) dalam item pentaksiran mata pelajaran Rombak Rawat Sistem Klac dengan menggunakan Model Pengukuran Rasch.

### 4.0 Metodologi

Kajian ini dijalankan dengan menggunakan pendekatan kuantitatif. Responden dipilih secara rawak bertujuan dan pengkaji secara sengaja memilih sampel untuk memenuhi kehendak kajian (Mohd. Najib, 2003). Othman Talib (2013) menyatakan bahawa pemilihan sampel dengan menggunakan teknik bertujuan ini merupakan pemilihan berdasarkan tujuan mendapatkan individu dengan kriteria yang ditetapkan oleh penyelidik. Penyelidik yang menggunakan teknik persampelan ini sudah menetapkan tujuan dan sasaran sampel mereka. Maka, dalam kajian ini, sampel terdiri daripada 64 pelajar yang telah dan sedang mempelajari mata pelajaran Rombak Rawat Sistem Klac di Kolej Komuniti Bandar Darulaman. Sampel yang dipilih hanyalah khusus kepada pelajar-pelajar yang telah atau sedang mempelajari mata pelajaran tersebut sahaja kerana kajian terhadap item soalan ini memerlukan pengetahuan responden terhadap item-item yang dikaji. Sekiranya teknik persampelan bertujuan tidak digunakan, respon responden terhadap item-item yang dikaji tidak menggambarkan keupayaan sebenar individu yang terlibat dalam kajian. Kebarangkalian responden

menjawab secara tekaan atau telahan semata-mata adalah tinggi. Sebanyak 30 item berbentuk dikotomi diedarkan kepada responden. Data dianalisis dengan menggunakan perisian Winstep versi 3.69.1.11 berpandukan Model Pengukuran Rasch.

## 5.0 Analisis dan perbincangan

### 5.1 Kebolehpercayaan dan indeks pengasingan

**Jadual 1:** Analisis kebolehpercayaan dan indeks pengasingan

Topik / LnP	Bil. Item	Pengukuran Item		Pengukuran Individu	
		Kebolehpercayaan	Indeks Pengasingan	Kebolehpercayaan	Indeks Pengasingan
LnP 01.02 (MSK)	20	0.90	3.08	0.63	1.06
LnP 01.03 (MRK)	10	0.94	3.94	0.63	1.06

Jadual 1 menunjukkan hasil statistik menggunakan Model Pengukuran Rasch. Statistik menunjukkan bagaimana model tersebut digunakan untuk menentukan indeks pengasingan item dan indeks pengasingan individu, dan juga kebolehpercayaan item serta kebolehpercayaan individu. Indeks pengasingan individu dan indeks pengasingan item yang melebihi nilai 2.0 dikategorikan sebagai baik (Linacre, 2007). Manakala, nilai 1.5 hingga 2.0 dikategorikan sebagai tidak produktif. Nilai kebolehpercayaan melebihi 0.8 diterima sebagai kebolehpercayaan yang tinggi, nilai kebolehpercayaan di antara 0.6 hingga 0.8 berada pada kategori yang boleh diterima, dan nilai di bawah 0.6 adalah tidak diterima untuk nilai kebolehpercayaan (Bond & Fox, 2007). Dapatan kajian mendapati indeks pengasingan individu bagi Topik Penyelenggara Sistem Klac (MSK) ialah 1.06. Sementara itu, nilai indeks pengasingan item bagi topik tersebut ialah 3.08. Bagi Topik Merombak Rawat Komponen Klac (MRK), nilai indeks pengasingan individu ialah 1.06, manakala indeks pengasingan item ialah 3.94. Nilai indeks pengasingan yang melebihi 2.0 adalah pada tahap yang baik (Linacre, 2007).

Merujuk kepada Bond dan Fox (2007), nilai kebolehpercayaan yang melebihi 0.8 adalah diterima dan merupakan nilai yang kuat, manakala nilai yang kurang daripada 0.8 adalah kurang diterima. Jadual 1 menunjukkan kebolehpercayaan individu bagi MSK ialah 0.53, manakala nilai kebolehpercayaan item ialah 0.90. Bagi topik MRK, nilai kebolehpercayaan individu ialah 0.53, manakala nilai kebolehpercayaan item ialah 0.94.

Kajian ini memberi fokus kepada pengasingan item. Nilai pengasingan item merujuk kepada bilangan strata kesukaran item yang diperolehi daripada set ujian yang dikendalikan. Nilai indeks pengasingan item dalam kajian ini memenuhi cadangan Linacre (2007) iaitu melebihi nilai 2.0.

## 5.2 Polariti item

**Jadual 2(a):** Analisis PTMEA item LnP 01.02

Topik/ Item	K201	K202	K203	K204	K205	K206	K207	K208	K209	K210
LnP 01.02 (MSK)	0.15	0.52	<b>-0.26</b>	0.29	0.31	0.14	0.46	0.41	0.41	0.41

**Jadual 2(b):** Analisis PTMEA item LnP 01.02

Topik/ Item	U201	U202	U203	U204	U205	U206	U207	U208	U209	U210
LnP 01.02 (MSK)	0.07	0.46	0.33	0.66	0.42	0.26	0.10	0.32	0.26	0.27

**Jadual 3:** Analisis PTMEA item LnP 01.03

Topik/ Item	U301	U302	U303	U304	U305	U306	U307	U308	U309	U310
LnP 01.03 (MRK)	0.38	0.31	0.27	0.08	0.21	0.41	0.45	<b>-0.01</b>	0.01	0.40

Analisis polariti atau keselarian item adalah indikator yang digunakan untuk menunjukkan item-item yang digunakan bergerak dalam satu arah yang dimaksudkan oleh konstruk yang diukur. Nilai positif bagi semua item memberi makna bahawa kesemua item mengukur konstruk atau topik yang sama (Bond dan Fox, 2007). Jika terdapat nilai negatif, data perlu diperiksa semula sama ada item tersebut perlu dibaiki atau digugurkan kerana ianya bercanggah dengan pembolehubah atau topik yang ingin diukur (Linacre, 2007). Jadual 2(a) menunjukkan satu item bernilai negatif (K203), manakala Jadual 3 juga menunjukkan satu item bernilai negatif (U308). Kedua-dua item tersebut perlu diteliti semula sama ada perlu dibaiki atau digugurkan.

## 5.3 Kesesuaian item (*Item fit measure construct*)

**Jadual 4:** Analisis kesesuaian item berdasarkan nilai MNSQ

ENTRY NUMBER	INFIT		OUTFIT		ITEM
	MNSQ	ZSTD	MNSQ	ZSTD	
3	1.29	2.00	2.08	4.20	K203
1	1.01	0.30	0.58	0.00	K201
19	0.97	0.20	0.51	-0.40	U209
8	0.88	-0.20	0.57	-0.80	K208
26	0.85	-0.10	0.44	-0.80	U306
2	0.80	-0.50	0.50	-1.20	K202

Analisis menggunakan model Rasch boleh menganggarkan darjah kesesuaian item yang mengukur sesuatu pembolehubah. Julat yang boleh

diterima untuk statistik MNSQ bagi item dikotomi ialah 0.6 hingga 1.4 (Bond dan Fox, 2007). Sementara itu, nilai Zstd perlu berada di antara -2 dan +2. Namun, jika nilai MNSQ diterima, maka nilai Zstd boleh diabaikan. Linacre (2007) mencadangkan agar item-item yang berada di luar julat perlu diasingkan bagi tujuan pengubahsuaian atau dibaiki sebelum dibuang. Jadual 4 menunjukkan 5 item MSK berada di luar julat yang dicadangkan. Item-item tersebut ialah K203, K201, U209, K208, dan K202. Manakala, 1 item MRK berada di luar julat yang ditetapkan iaitu item U306.

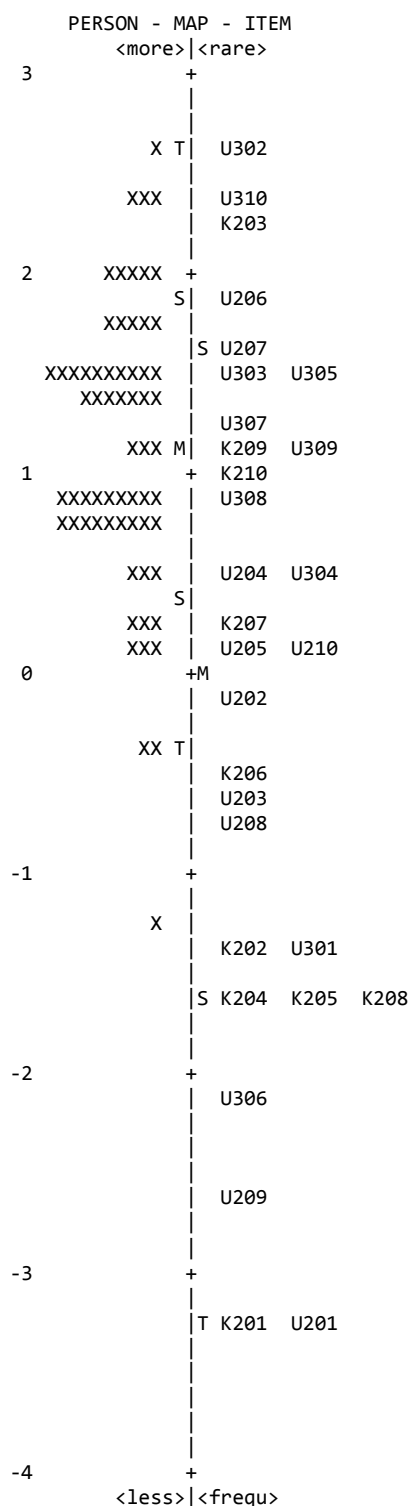
#### **5.4 Peta item-individu (*Item-person map*)**

Peta item-individu digunakan untuk menunjukkan sama ada ujian atau instrumen yang ditadbir itu sesuai dengan kebolehan sampel atau menepati sasaran (objektif) pengujian tersebut dilaksanakan. Garisan putus di tengah menggambarkan lokasi kebolehan individu dan ukuran kesukaran item pada skala logit. Pada bahagian kiri garisan putus-putus adalah kedudukan individu; di sebelah kanan adalah kedudukan item; di bahagian atas-kiri adalah kedudukan individu dengan kebolehan tinggi dan di bahagian atas-kanan adalah kedudukan item dengan ukuran kesukaran tinggi. Dua huruf M pada garisan putus-putus tersebut menunjukkan ukuran kebolehan min (purata) bagi individu (kiri) dan item (kanan). Huruf S di bahagian kanan adalah bermaksud satu sisihan piawai (*standard error*, S.E) bagi individu dan di bahagian kanan pula menunjukkan sisihan piawai bagi item. Huruf T pula menunjukkan dua sisihan piawai bagi individu (kanan) dan dua sisihan piawai bagi item (kiri).

Rajah 2 menunjukkan item K201 dan U201 adalah item yang sangat mudah dipersetujui dengan nilai ukuran kedua-dua item tersebut mencatatkan -3.30 logit. Sementara itu, item U302 merupakan item yang sangat sukar dipersetujui dengan nilai ukuran 2.63 logit. Hanya seorang responden yang berkebolehan sangat tinggi, dan 3 orang responden yang berkebolehan sangat rendah.



TABLE 12.2 DIF ITEM SKR3102.sav ZOU476WS.TXT Apr 26 21:33 2017  
 INPUT: 64 PERSON 30 ITEM MEASURED: 64 PERSON 30 ITEM 2 CATS WINSTEPS 3.69.1.11



**Rajah 2:** Pemetaan kesukaran item dan kebolehan individu

## 6.0 Rumusan

Setelah analisis data dijalankan, dapatan-dapatan kajian yang diperolehi berpandukan Model Pengukuran Rasch perlu diteliti satu persatu bagi memastikan kesahan dan kebolehpercayaan instrumen pentaksiran dipatuhi. Item-item pentaksiran yang dicadangkan untuk diberi perhatian, sama ada perlu dibaiki atau digugurkan daripada instrumen pengukuran bagi mata pelajaran Rombak Rawat Sistem Klac. Jadual 5 menunjukkan rumusan keseluruhan item-item tersebut. 5 daripada 20 item yang mewakili Topik MSK perlu dibaiki atau digugurkan. Manakala, 2 daripada 10 item yang mewakili Topik MRK perlu dibaiki atau turut digugurkan.

**Jadual 5:** Rumusan Keseluruhan Item Yang Digugurkan

Topik / LnP	Item	Pengguguran Item			Jumlah Item	Item Digugurkan	Item Dikekalkan
		PTMEA	Infit	Outfit			
01.02 (MSK)	LnP K203	-0.26	1.29	2.08	20	5	15
	K201	0.15	1.01	0.58			
	U209	0.26	0.97	0.51			
	K208	0.41	0.88	0.57			
	K202	0.52	0.80	0.50			
01.03 (MRK)	LnP U306	0.41	0.85	0.44	10	2	8
	U308	-0.01	1.24	1.32			

Kesimpulannya, kesahan dan kebolehpercayaan adalah perkara yang utama untuk dipertimbangkan dalam pembinaan sesuatu instrumen pentaksiran. Secara keseluruhan, 8 item digugurkan daripada instrumen pentaksiran setelah mengambil kira faktor kesahan dan kebolehpercayaan. Dengan menggunakan Model Pengukuran Rasch, tahap kesukaran item dan kebolehan responden dapat ditentukan dengan tepat. Penambahbaikan dalam pembinaan item pentaksiran perlu dilakukan bagi memastikan item yang dibina dapat mengukur sesuatu topik, dimensi atau konstruk. Dengan yang demikian, dapatan-dapatan kajian yang dibincangkan dalam kajian ini telah menjawab kesemua tujuan dan objektif kajian.

## Rujukan

Bond, T.G. & Fox, C.M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Second Edition. New Jersey: Lawrence Erlbaum Associates Publishers London.

Linacre, J.M. (2007). *A User's Guide to WINSTEPS Rasch-model Computer Programs*. Chicago: MESA Press.

Masters, G.N., & Wright, B.D. (1984). *The essential process in a family of measurement models*. *Psychometrika*, 49: 529-544.

Mohd Najib Abd Ghafar. (2003). *Rekabentuk Tinjauan Soal Selidik Pendidikan*. Skudai, Johor: Penerbit Universiti Teknologi Malaysia.

Othman Talib. (2013). *Asas Penulisan Tesis Penyelidikan & Statistik*. Serdang, Selangor: Penerbit Universiti Putra Malaysia.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Denmark's Paedagogiske Institut.

Ruhizan Mohd. Yasin, Faizal Amin Nur Yunus, Ridzwan Che Rus, Azmi Ahmad, & Mohd Bekri Rahim. (2015). Validity and Reliability Learning Transfer Item Using Rasch Measurement Model. *Procedia Social and Behavioral Sciences*, 204: 212-217.

Siti Rahayah Ariffin, Bishanani Omar, Anita Isa & Sharida Sharif. (2010). Validity dan reliability multiple intelligent item using Rasch Measurement Model. *Procedia Social and Behavioral Sciences*, 9: 729-733.

Siti Rahayah Ariffin, Rodiah Idris & Noriah Mohd Ishak. (2010). Differential Item Functioning in Malaysian Generic Skills Instrument (MyGSI). *Jurnal Pendidikan Malaysia*, 35(1) : 1-10.

Siti Rahayah Ariffin. (2008). *Inovasi dalam Pengukuran dan Penilaian Pendidikan*. Universiti Kebangsaan Malaysia.

Wright, B.D. & Stone, M.H. (1979). *Best Test Design*. Chicago, IL: MESA Press.