

Penilaian Awal Alat Ukuran Keupayaan Kognitif Umum
(Preliminary Assessment Of The General Mental Ability Measurement)

Noorashikeen Mohamed
Wan Shahrazad Wan Sulaiman*
Fatimah wati Halim

*Program Psikologi, Pusat Kajian Psikologi dan Kesejahteraan Manusia,
Fakulti Sains Sosial dan Kemanusiaan, Universiti Kebangsaan Malaysia*

*Corresponding e-mail: [shara@ukm.edu.my]

The objective of this research is to assess the validity and reliability of the general mental ability instrument that was developed based on Western norms. The instrument comprise 153 items obtained from the International Cognitive Ability Resources (ICAR) categorised into three dimension (3D) rotation, abstract reasoning, figural analogies, letter-number series, matrix reasoning and verbal reasoning. Forty-five (45) masters-programme students from the Human and Social Science Faculty at a local university undertook the 1-hour test. Utilising the *Winsteps* software, the Rasch measurement analysis comprising summary statistics, unidimensionality and item-person measures were used to test the validity and reliability of the ICAR. The research findings indicated that the ICAR has good psychometric properties that is acceptable as a measure of general mental ability. Nonetheless, the findings also indicated that the ICAR was difficult relative to the respondents' ability. As such, improvements and selection of appropriate item-difficulty level are required to portray a more accurate level of the general mental ability in the local context.

Key words: assessment, validity, reliability, general mental ability, RASCH measurement model.

Keupayaan kognitif umum atau *general mental ability* (GMA) menentukan kemampuan seseorang individu untuk belajar dan melaksanakan tugas dengan baik (Asfar, Born, Oostrom & Vugt 2019). Turut dikenali sebagai faktor *g*, ia ditakrifkan sebagai kemampuan untuk memahami idea-idea kompleks, beradaptasi secara efektif dengan persekitaran, belajar daripada pengalaman, terlibat dalam berbagai-bagai bentuk penaakulan dan mengatasi rintangan melalui daya fikir (Neisser et al. 1996). Menyoroti sejarah terdahulu, konsep *g* mula diperkenalkan oleh Charles Spearman, seorang ahli psikologi Inggeris, pada awal abad ke-20. Faktor *g* merupakan satu pemboleh ubah yang merangkumkan dapatan korelasi positif di antara beberapa tugas kognitif yang berbeza; seterusnya memberi gambaran bahawa prestasi seseorang individu dalam sesuatu tugas kognitif itu boleh dibandingkan dengan prestasinya dalam tugas kognitif yang lain (Kamphaus et al. 2005). Oleh itu, boleh dikatakan bahawa faktor *g* adalah satu bentuk gagasan atau dimensi yang dibangunkan untuk mengukur keupayaan kognitif dan kecerdikan individu.

Dalam dunia pekerjaan, keupayaan kognitif umum mempunyai bukti empirikal yang menyokongnya sebagai peramal yang lebih baik terhadap prestasi kerja (Schmidt 2002; Schmidt & Hunter 2004; Ones et al. 2012). Faktor *g* juga berupaya meramalkan tahap pekerjaan dan prestasi dalam pekerjaan yang dipilih oleh seseorang (Schmidt & Hunter 2004). Selain itu, faktor *g* bukan sahaja merupakan peramal yang paling penting terhadap prestasi kerja (Reeve & Hakel 2002) malah menjadi elemen yang penting serta popular dalam proses pemilihan oleh sesebuah organisasi di Amerika Syarikat dan Eropah (Salgado et al. 2003; Salgado 2017).

Ujian *aptitude* tidak sama dengan ujian kecerdasan. *Aptitude* bermaksud suatu bentuk atau sifat kecenderungan, bakat dan kebolehan semula jadi individu (Snow 1976; Barret 2009 dlm Nik Hairi & Rusyda Helma 2020). Turut dirujuk sebagai ujian bakat atau kebolehan, ujian *aptitude* cuba meramalkan potensi yang dimiliki oleh seseorang dengan tujuan untuk memperolehi faedah daripada latihan akademik atau vokasional yang akan diberikan (Azizi Yahaya 2008). Dalam konteks pembelajaran, ujian *aptitude* berperanan sebagai mekanisme untuk mengesan perbezaan atau tahap kebolehan semula jadi individu dalam aspek berkaitan abstrak, numerikal, daya tumpu, verbal dan ruang (Snow 1976; Reschly & Carol 2000; Barret 2009, 2012 dlm Nik Hairi & Rusyda Helma 2020) secara spesifik. Ini bermaksud, ujian *aptitude* bertujuan untuk menentukan tahap kebolehan seseorang individu dalam melakukan tugas tertentu. Manakala, ujian kecerdasan pula, yakni faktor *g*, merupakan sejenis ujian kemampuan yang lebih luas lagi dan menilai keseluruhan aspek keupayaan mental (Azizi Yahaya 2008). Ujian kecerdasan direka untuk kecerdasan umum atau keupayaan mental global yang mendasari prestasi dalam semua bidang. Kecerdasan merangkumi pelbagai aspek kebolehan termasuk penyelesaian masalah, pemikiran logik, memori, pengetahuan dan keupayaan untuk menyesuaikan diri dengan persekitaran yang berubah-ubah.

Rata-rata organisasi di Malaysia menggunakan ujian *aptitude* untuk mengukur potensi seseorang individu (Nik Hairi & Rusyda Helma 2020). Kekangan dalam penggunaan ujian *aptitude* disebabkan sifatnya yang spesifik dan sesuai untuk mengukur tugas atau jawatan tertentu sahaja. Sedangkan faktor *g* adalah bersifat umum yang memberi gambaran kepada keupayaan mental secara keseluruhan. Bagi organisasi yang

mempunyai kekangan daripada segi kos dan masa, penggunaan alat ukuran keupayaan kognitif umum adalah lebih praktikal berbanding ujian *aptitude* kerana proses penyesuaian alat ukuran dengan tugas atau jawatan yang hendak dinilai tidak perlu dilakukan, atau ia hanya melibatkan proses penyesuaian yang minimum. Namun, alat ukuran yang khusus mengukur faktor *g* adalah terhad di Malaysia dan kebanyakannya berasal dari Barat serta berdasarkan norma masyarakat di sana. Oleh itu, kesesuaian aplikasinya di Malaysia perlu dikaji terlebih dahulu.

Sungguhpun penelitian terhadap kebolehpercayaan dan kesahan boleh dilakukan, namun timbul satu lagi isu berhubung dengan kos pembelian alat ukuran, pentadbiran dan pelaporan ujian yang melibatkan kos tinggi. Di Malaysia, kebanyakan organisasi yang menawarkan perkhidmatan pentadbiran ujian tidak menyatakan kos yang terlibat secara terbuka. Namun, merujuk kepada *Psychological Assessment Resources Inc.* atau lebih dikenali sebagai PAR (<https://www.parinc.com>), penerbit ujian psikologi tersohor di Amerika Syarikat, kos satu kit ujian boleh mencecah sehingga RM1000 lebih (USD240) bergantung kepada jenis-jenis ujian psikologi. Kos ini belum mengambil kira caj kepakaran, kos pelaporan dan sebagainya. Manakala ujian yang digunakan oleh Suruhanjaya Perkhidmatan Awam (SPA) tidak boleh diakses oleh pihak luar kerana ia digunakan untuk proses pengambilan jawatan untuk perkhidmatan awam. Maka integriti ujian tersebut perlu dipelihara dan dihadkan penggunaannya agar tidak disalahgunakan oleh pihak-pihak tertentu. Menyedari hal ini, pengkaji berhasrat untuk membangunkan alat ukuran keupayaan kognitif umum yang boleh digunakan oleh pengurus sumber manusia dalam membantu membuat keputusan dalam aspek pengambilan, penempatan, kenaikan

pangkat, latihan, kemajuan kerjaya dan pembangunan sendiri.

Tujuan dan Objektif Kajian

Kajian ini bertujuan untuk menguji kebolehpercayaan dan kesahan alat ukuran *International Cognitive Ability Resources* (ICAR) dengan menggunakan analisis pengukuran Rasch. Secara khusus, objektif kajian ini adalah untuk:

1. Mengukur alat ukuran menggunakan ujian kebolehpercayaan *item-person* dalam analisis Rasch.
2. Mengukur alat ukuran menggunakan ujian keseragaman dimensi (*unidimensionality assessment*) dalam analisis Rasch.
3. Mengukur alat ukuran menggunakan ujian kesesuaian *item-individu* (*item-individual fit*) dan nilai piawai *z* (*z-standard value*) dalam analisis Rasch.

Model Pengukuran Rasch

Model pengukuran Rasch, atau lebih dikenali sebagai model Rasch, merujuk kepada satu prinsip atau teknik yang membolehkan sesuatu pengukuran dibuat terhadap sifat pendam (*latent trait*). Berdasarkan kepakaran matematikanya, Georg Rasch (1960) membuktikan bahawa model yang diciptanya dapat memenuhi kriteria atau standard pengukuran yang sama seperti analisis dalam bidang sains fizikal (Azrilah et al. 2017). Menurut Linacre (2010), model yang dipopularkan oleh Ben Wright ini boleh menunjukkan kemampuan individu berdasarkan data mentah yang berbentuk dikotomi. Melalui model analisis yang dikenali sebagai *One Parameter Logistic* (IPL) (Olsen 2003), George Rasch membangunkan sebuah model pengukuran yang berupaya menentukan hubungan antara tahap keupayaan individu dan tahap kesukaran

item (Sumintono & Widhiarso 2013, Hudiya et al. 2017).

Dalam penyelidikan sains sosial, model Rasch ini merupakan satu-satunya kaedah pengukuran piawai yang mudah digunakan bagi menentukan kesahihan sesuatu alat ukuran yang dibentuk (Azrilah et al. 2017) bagi mengukur sifat terpendam (*latent trait*). Penilaian kesahihan ini tidak tertakluk kepada nilai *Cronbach* alfa sahaja; penilaian terperinci turut dilakukan ke atas sifat terpendam atau konstruk yang diukur dengan memastikan ia mengukur objektiviti spesifik dalam dimensi yang sama, di samping meminimalkan ketakjelasan dalam pengukuran. Selain itu, model Rasch juga berupaya membuat jangkaan ketepatan yang realistik dan menunjukkan pembuktian kualiti item bagi mengukuhkan lagi kesahihan alat ukuran yang dibentuk (Azrilah et al. 2017; Azrilah 2011 dalam Mona et. al 2016).

Bergantung pada jenis penyelidikan, model Rasch memperkenalkan logaritma nilai kebarangkalian (*probabilistic value*) untuk menghasilkan kaedah pengukuran dengan fungsi logit. Pengukuran logit bermula pada 0 dan berada pada skala pemisah dengan taburan jarak yang sekata antara sela (*interval*). Prinsip yang digunakan oleh model ini menyerupai pengukuran lain yang membentuk model yang sesuai dengan data kajian (Edelen & Reeve 2007). Dalam kajian ini, analisis model Rasch menunjukkan wujudnya hubungan hierarki antara respons individu dan tindakan atau item yang dicerap. Setiap item mempunyai ukuran logit tersendiri yang menentukan tahap kesukaran soalan. Oleh itu, kajian ini menggunakan teorem model pengukuran Rasch untuk menentukan kredibiliti data di samping memperincikan lagi *Cronbach* alfa bagi kebolehpercayaan dan kesahan data yang diperoleh.

Dalam kajian ini, model Rasch melalui perisian *Winsteps* versi 3.72.3 digunakan untuk analisis tapisan data (*data screening*). Hasil analisis memaparkan jadual dan output berbentuk grafik untuk memudahkan penelitian ke atas kesesuaian pengukuran data. Jadual ringkasan statistik (*summary statistic*) yang mengandungi indikator penting (seperti nilai *Cronbach* alfa, kebolehpercayaan item, kebolehpercayaan responden dan ralat piawai) diteliti untuk menentukan sama ada data itu sesuai untuk proses analisis selanjutnya. Data yang melepasi tapisan awal seterusnya dianalisis dengan lebih terperinci dengan memfokuskan kepada nilai *infit* atau *outfit* MNSQ bagi kedua-dua pengukuran item dan responden (*item and person measure*) serta aspek unidimensionaliti item.

Metod

Kajian ini merupakan kajian tinjauan yang menggunakan teknik persampelan mudah. Seramai 45 pelajar siswazah dari Pusat Psikologi, Fakulti Sains Sosial dan Kemasyarakatan di sebuah universiti awam terlibat dalam kajian ini. Memandangkan kajian ini merupakan kajian rintis dan sampel minimum seramai 30 orang (Linacre 1994) sudah memadai untuk menghasilkan pengukuran yang agak stabil secara statistik (dlm Azrilah et al. 2017), pengkaji hanya mentadbir kepada satu kelompok pelajar daripada bidang pengajian yang sama. Selain itu, kajian ini menggunakan soal selidik sebagai alat ukuran kajian. Sebagaimana yang ditunjukkan dalam Jadual 1, alat ukuran kajian ini terdiri daripada dua bahagian. Bahagian A yang mengandungi item demografi responden, sementara Bahagian B mengandungi item yang mengukur konstruk keupayaan kognitif umum.

Jadual 1

Bahagian dalam Soal Selidik

Bahagian		Bilangan Item
Bahagian A	Keterangan Diri Peserta	9
Bahagian B	Keupayaan Kognitif Umum	
	• Putaran 3 Dimensi	23
	• Taakulan Abstrak	90
	• Analogi Bentuk	8
	• Siri Abjad-Nombor	9
	• Taakulan Matriks	11
	• Taakulan Lisan	12

Setelah meneliti beberapa set alat ukuran keupayaan kognitif umum antaranya pengukuran *Raven's Standard Progressive Matrices* (Raven's SPM), *Intelligenz Struktur Test*, *The Wonderlic Test* dan *International Cognitive Ability Resources* (ICAR), pengkaji memutuskan untuk mengadaptasi item-item daripada laman sesawang ICAR. Keputusan ini dibuat setelah mengambil kira isu-isu seperti kos, hak cipta dan sebagainya. *International Cognitive Ability Resources* (ICAR) merupakan satu entiti penyelidikan domain awam (*public domain*) yang dikelolakan oleh pakar-pakar daripada beberapa buah universiti terkemuka di luar negara. ICAR diwujudkan dengan tujuan untuk menggalakkan penyelidikan yang lebih luas terhadap faktor kebolehan kognitif dalam bidang psikologi dan sains sosial yang lain di samping membantu penilaian neuropsikologi dalam kajian dan amalan perubatan. Di samping itu, pelbagai jenis format yang disediakan oleh pasukan ICAR memberi ruang yang luas untuk penyelidik menentukan jenis dan format yang bersesuaian dengan tujuan kajian.

Pengkaji menggunakan empat (4) daripada tujuh (7) jenis item asal yang dibangunkan oleh pasukan teras ICAR bagi tujuan kajian ini. Tiga lagi jenis item digugurkan kerana ia perlu ditadbir secara atas talian. Pengguguran item-item tersebut tidak menjejaskan keseluruhan ujian kerana item-item dalam ICAR dibangunkan secara berasingan oleh pakar-pakar yang

berbeza. Selain itu, pengkaji turut memasukkan dua (2) lagi jenis item yang disumbangkan oleh pengkaji lain iaitu analogi bentuk (Blum, Holling, Galibert & Forthmann 2016; Pasukan ICAR 2014) dan taakulan abstrak (Mahalingam 2016; Pasukan ICAR 2014). Item-item sumbangan lain tidak dipilih kerana kurang sesuai untuk digunakan dalam kajian ini. Kesemua item ICAR ini boleh diakses di laman sesawang <https://icar-project.com/>.

Model pengukuran Rasch digunakan dalam kajian ini kerana kemampuannya mengesan kebolehan dan kemampuan setiap calon yang menjawab item dalam soal selidik di samping mengukur tahap kesukaran setiap item secara serentak. Perkara ini penting bagi membantu proses pemurnian alat ukuran kajian secara sistematik dan telus (Karim et al. 2014). Sehubungan itu, boleh diringkaskan bahawa hasil analisis Rasch dapat membantu pengkaji meneliti dan menangani isu-isu yang mendasari aspek kebolehpercayaan item dan kesahan konstruk dalam alat ukuran kajian. Kesannya, pemilihan item yang benar-benar sesuai untuk digunakan dalam konteks tempatan dapat dilakukan dengan baik dan sistematik.

Keputusan dan Perbincangan

Aplikasi model Rasch yang tepat memerlukan beberapa kriteria asas

dipenuhi. Maka, langkah pertama yang dilaksanakan merupakan pengujian ke atas data yang diperoleh untuk menentukan aspek kebolehpercayaan dan kesahan serta penilaian terhadap kesesuaian data untuk proses analisis lanjut. Penting juga untuk menentukan sama ada sifat terpendam atau konstruk dalam alat ukuran kajian mengukur objektiviti spesifik dalam dimensi yang sama. Ini boleh dicapai melalui penilaian terhadap tiga parameter utama dalam model Rasch merangkumi elemen ringkasan statistik untuk pengesahan kebolehpercayaan dan kawalan data, unidimensionaliti untuk kesahan konstruk dan nilai *infit/ outfit* MNSQ item-responden (*item-person*) sebagai indikator kesesuaian (*fit*) pengukuran. Hasil analisis menunjukkan bahawa model Rasch dapat memenuhi kriteria-kriteria di atas.

Maklumat ringkasan statistik item dan responden (*person*) dilaporkan dalam Jadual 2 dan 3. Analisis ini dihasilkan daripada 45 responden yang terdiri daripada pelajar sarjana Fakulti Sains Sosial dan Kemanusiaan di universiti awam. Sejumlah 6695 data poin terhasil daripada interaksi yang dilakukan terhadap analisis ke atas 45 responden dengan 153 item, yang turut menghasilkan Khi-kuasa dua bernilai 6307.15 dengan 6500 darjah kebebasan pada $p=.9957$. Hasil ringkasan 153 item yang diukur mengesahkan data yang tersebar dari 2.53 hingga ke (-3.45) = 5.98 logit, dan nilai kebolehpercayaan 0.84, adalah baik bagi kedua-duanya. Menurut Linacre (2011), had yang boleh diterima untuk item soal selidik adalah: $0.4 < \text{nilai korelasi ukuran titik (point measure correlation)} < 0.8$; $0.5 < \text{nilai outfit Mean Square (MNSQ)} < 1.5$; dan nilai *outfit z-standard* < 2.0 .

Jadual 2

Ringkasan Statistik Pengukuran 153 Item (Summary of 153 Measured Item)

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	11.4	44.3	.00	.43	1.00	.0	1.02	.0
S.D.	8.2	.8	1.11	.14	.04	.5	.13	.6
MAX.	40.0	45.0	2.52	1.01	1.11	1.5	1.53	1.5
MIN.	1.0	41.0	-3.45	.30	.87	-1.8	.68	-1.8
REAL RMSE	.45	TRUE SD	1.02	SEPARATION	2.25	ITEM	RELIABILITY	.84
MODEL RMSE	.45	TRUE SD	1.02	SEPARATION	2.27	ITEM	RELIABILITY	.84
S.E. OF ITEM MEAN = .09								

UMEAN=.0000 USCALE=1.0000
 ITEM RAW SCORE-TO-MEASURE CORRELATION = -.95
 6695 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 6307.15 with 6500 d.f. p =.9557
 Global Root-Mean-Square Residual (excluding extreme scores): .3903

Seterusnya, taburan jawapan daripada responden (rujuk Jadual 2) adalah dari (-.57) hingga (-2.27) = 2.84 logit. Idealnya, nilai kuasa pemisah melebihi 3.00 menggambarkan kemampuan mengasingkan individu yang lebih baik. Namun, dalam kajian ini nilai kuasa pemisah responden (*Person Separation*) hanya 1.35 yang memberi gambaran item-item sedia ada tidak dapat membezakan kemampuan individu dengan baik.

Menurut Fisher (dalam Mona et al. 2016), nilai skala pemisah individu dan item untuk pengukuran kritikal yang dianggap baik adalah di antara 3 hingga 5 logit. Semakin tinggi nilai kuasa pemisah, semakin baik kemampuan pengasingan individu yang digambarkan. Tambahan lagi, Ghazali (2008) turut menyatakan bahawa nilai alpha 0.60 dalam kajian sosial sains sudah dianggap sesuai, seperti yang diamalkan oleh beberapa saintis di

lapangan. Atas dasar ini, kebolehpercayaan 0.64 dan Cronbach alfa 0.66 menunjukkan data yang masih boleh dipercayai, maka data tersebut boleh digunakan untuk analisis selanjutnya.

Oleh itu, boleh disimpulkan hasil analisis Rasch menunjukkan bahawa item-item ICAR ini perlu dinilai semula untuk meningkatkan kuasa pemisah dan kualitinya.

Jadual 3

Ringkasan Statistik Pengukuran 45 Responden (Summary of 45 Measured Person)

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	38.1	150.8	-1.32	.21	1.00	-.1	1.02	.0
S.D.	8.3	4.9	.36	.02	.10	1.0	.22	1.1
MAX.	58.0	153.0	-.57	.26	1.22	1.7	1.77	2.9
MIN.	20.0	122.0	-2.27	.19	.75	-3.2	.67	-2.3
REAL RMSE	.22	TRUE SD	.29	SEPARATION	1.35	PERSON RELIABILITY	.64	
MODEL RMSE	.21	TRUE SD	.29	SEPARATION	1.39	PERSON RELIABILITY	.66	
S.E. OF PERSON MEAN = .05								

PERSON RAW SCORE-TO-MEASURE CORRELATION = .98
 CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .66

Tahap maksimum pengukuran item seperti yang tertera di Jadual 2 adalah +2.52 logit (SE:1.01) berbanding dengan pengukuran maksimum *Person* hanya -0.57 logit (SE:0.26) yang ditunjukkan menerusi Jadual 3. Julat skala item yang boleh dimanfaatkan pula adalah daripada -2.27 hingga +2.52 logit selebar 4.79 logit. Walaupun kebolehpercayaan dan julat skala item yang baik, terdapat jurang yang ketara, iaitu selebar 1.95 logit, di antara item dan responden (-.57 logit hingga +2.52 logit). Ini memberi gambaran bahawa item-item yang dibangunkan mempunyai tahap kesukaran lebih tinggi berbanding keupayaan (*ability*) responden. Buktinya wujud lompong yang luas selebar 3.09 logit dari -.57 logit hingga +2.52 logit di mana terdapat lambakan item-item yang tidak dijawab oleh responden. Sebaliknya, analisis menunjukkan jumlah item mudah yang kurang mencukupi di mana ukuran minimum item ialah -3.45 logit, berbanding responden pada -2.27. Masih wujud lompong selebar 1.08 logit. Justeru, boleh disimpulkan bahawa pengkaji perlu meneliti dan menyusun semula item-item dalam kedua-dua julat tersebut untuk

meminimalkan lambakan item yang sukar dan menambah bilangan item mudah.

Berdasarkan tinjauan kajian terdahulu, data ini memerlukan analisis lanjut yang bertujuan untuk menentukan sama ada item-item mengukur objektiviti spesifik yang sama atau tidak. Keperluan di atas dicapai menerusi analisis unidimensionaliti di mana langkah ini mengenal pasti varians dan perbezaan data dengan meneliti kelompok item yang berkongsi pola yang sama.

Analisis Keseragaman Dimensi (Unidimensionality Analysis)

Penilaian keseragaman dimensi atau unidimensional ternyata penting bagi memastikan objektif pengukuran adalah mantap dan boleh dicapai. Siti Rahayah (2008) menyatakan bahawa unidimensi merujuk kepada pengukuran yang hanya dilakukan pada satu konstruk sahaja dalam satu kajian. Analisis keseragaman dimensi atau unidimensional dalam model Rasch mengukur varians bukan rawak yang dijumpai dalam data yang menjelaskan dimensi tunggal kesukaran serta kebolehan (Sick 2010, Hudiya et al. 2017). Oleh itu,

analisis Rasch dengan aplikasi teknik *Residual Principal Component Analysis* (PCA) bertujuan untuk mengenal pasti keupayaan alat ukuran untuk “mengukur dalam satu dimensi yang seragam dengan aras gangguan atau penarik yang boleh diterima” (Azrilah et al. 2017).

Bagi maksud kajian ini, analisis keseragaman dimensi dilakukan untuk meneliti sama ada item-item yang dibina mengukur dimensi yang sama secara konsisten. Ini merupakan langkah kedua selepas penilaian ke atas jadual ringkasan statistik. Bagi aspek keseragaman dimensi ini, Linacre (2011) menyatakan bahawa

nilai varians kasar yang dijelaskan oleh pengukuran mestilah melepasi batas minimum 40%, diikuti dengan nilai varians yang tidak dijelaskan kurang dari 15% dalam *Kontras 1*. Kriteria ini perlu dipenuhi dengan kekuatan sekurang-kurangnya 3 item. Selari dengan pandangan Linacre tentang keperluan memenuhi syarat keseragaman dimensi dengan pencapaian minimum varians 40%, Conrad, Conrad, Dennis, Riley dan Funk (2009) turut berpendapat bahawa pencapaian varians yang melebihi 30% masih boleh diterima dan dianggap sebagai keseragaman dimensi pengukuran yang sederhana.

Jadual 4

Piawaian Varians Residual (dalam Unit Eigenvalue) Bagi Keseluruhan Item

Penerangan	Unit Eigenvalue	Empirikal	Model yang Dijangka
Varians kasar yang dijelaskan oleh pengukuran	37.8	20.0%	19.7%
Varians kasar yang tidak dijelaskan oleh pengukuran	151.0	80.0%	80.3%
Varians yang tidak dijelaskan dalam kontras 1	8.2	4.4%	
Varians yang tidak dijelaskan dalam kontras 2	7.6	4.0%	
Varians yang tidak dijelaskan dalam kontras 3	6.9	3.7%	
Varians yang tidak dijelaskan dalam kontras 4	6.8	3.6%	
Varians yang tidak dijelaskan dalam kontras 5	6.3	3.3%	

Berdasarkan kepada analisis keseluruhan item di Jadual 4, varians kasar yang dijelaskan oleh item ICAR adalah 19.7%, dengan varians yang tidak dapat dijelaskan dalam kontras pertama sebanyak 8.2 dan 4.4%. Walaupun keperluan keseragaman alat ukuran sekurang-kurangnya 20 peratus telah tercapai, diikuti dengan varians yang tidak dijelaskan dalam kontras 1 adalah terkawal dengan baik dan tidak melebihi nilai siling 15%, pengukuran ini masih tidak melepasi batas minimum yang

melibatkan 40 peratus keperluan Rasch. Justeru, boleh disimpulkan bahawa item dalam soal selidik KKU ini mempunyai konstruk multidimensional dan perlu dinilai secara berasingan. Perisian *Winsteps* diprogramkan untuk mengukur arah satu dimensi pada satu masa. Oleh itu, untuk memperbaiki isu nilai peratusan yang rendah pada model Rasch, pengkaji perlu menganalisis setiap konstruk secara berasingan. Dalam kajian ini, analisis Rasch perlu dilakukan secara berasingan

mengikut jenis item yang ada dalam alat ukuran ICAR. Hasil analisis Rasch

mengikut jenis item adalah seperti yang tertera dalam Jadual 5.

Jadual 5

Piawaian Varians Residual (dalam Unit Eigenvalue) Berdasarkan Jenis Item

Penerangan	Unit Eigenvalue		Empirikal	Model yang Dijangka
Item Taakulan Abstrak				
Varians kasar yang dijelaskan oleh pengukuran	21.3		19.1%	18.7%
Varians kasar yang tidak dijelaskan oleh pengukuran	90.0		80.9	81.3%
Varians yang tidak dijelaskan dalam Kontras 1	5.8		5.2%	
Varians yang tidak dijelaskan dalam Kontras 2	5.2		4.7%	
Varians yang tidak dijelaskan dalam Kontras 3	4.6		4.1%	
Varians yang tidak dijelaskan dalam Kontras 4	4.2		3.8%	
Varians yang tidak dijelaskan dalam Kontras 5	4.1		3.78	
Item Putaran 3D				
Varians kasar yang dijelaskan oleh pengukuran	2.7		10.6%	11.4%
Varians kasar yang tidak dijelaskan oleh pengukuran	23.0		89.4%	88.6%
Varians yang tidak dijelaskan dalam Kontras 1	2.7		10.5%	
Varians yang tidak dijelaskan dalam Kontras 2	2.4		9.5%	
Varians yang tidak dijelaskan dalam Kontras 3	1.9		7.4%	
Varians yang tidak dijelaskan dalam Kontras 4				
Varians yang tidak dijelaskan dalam Kontras 5				
Item Analogi Bentuk				
Varians kasar yang dijelaskan oleh pengukuran	2.1		21.0%	21.7%
Varians kasar yang tidak dijelaskan oleh pengukuran	8.0		79.0%	78.3%
Varians yang tidak dijelaskan dalam Kontras 1	1.7		16.6%	
Varians yang tidak dijelaskan dalam Kontras 2	1.3		12.9%	
Varians yang tidak dijelaskan dalam Kontras 3	1.3		12.5%	
Varians yang tidak dijelaskan dalam Kontras 4	1.2		11.5%	
Varians yang tidak dijelaskan dalam Kontras 5	1.0		9.9%	
Item Siri Abjad dan Nombor				
Varians kasar yang dijelaskan oleh pengukuran	4.1		34.0%	32.2%
Varians kasar yang tidak dijelaskan oleh pengukuran	8.0		66.0%	67.8%
Varians yang tidak dijelaskan dalam Kontras 1	1.7		14.4%	
Varians yang tidak dijelaskan dalam Kontras 2	1.5		12.1%	
Varians yang tidak dijelaskan dalam Kontras 3	1.3		10.3%	
Varians yang tidak dijelaskan dalam Kontras 4	1.2		9.6%	
Varians yang tidak dijelaskan dalam Kontras 5	9		7.7%	
Item Taakulan Lisan				
Varians kasar yang dijelaskan oleh pengukuran	10.8		49.5%	45.8%
Varians kasar yang tidak dijelaskan oleh pengukuran	11.0		50.5%	54.2%
Varians yang tidak dijelaskan dalam Kontras 1	2.0		9.4%	
Varians yang tidak dijelaskan dalam Kontras 2	1.8		8.1%	

Penerangan	Unit Eigenvalue		Empirikal	Model yang Dijangka
Varians yang tidak dijelaskan dalam Kontras 3	1.5		7.0%	
Varians yang tidak dijelaskan dalam Kontras 4	1.4		6.4%	
Varians yang tidak dijelaskan dalam Kontras 5	1.1		5.2%	
Item Taakulan Matriks				
Varians kasar yang dijelaskan oleh pengukuran	3.0		21.2%	20.9%
Varians kasar yang tidak dijelaskan oleh pengukuran	11.0		78.8%	79.1%
Varians yang tidak dijelaskan dalam Kontras 1	2.1		15.1%	
Varians yang tidak dijelaskan dalam Kontras 2	1.6		11.8%	
Varians yang tidak dijelaskan dalam Kontras 3	1.4		9.8%	
Varians yang tidak dijelaskan dalam Kontras 4	1.2		8.9%	
Varians yang tidak dijelaskan dalam Kontras 5	1.2		8.3%	

Berdasarkan Jadual 5 di atas, hasil analisis Rasch menunjukkan hanya satu jenis item sahaja (taakulan lisan) yang benar-benar memenuhi keperluan batas minimum 40%, dengan varians kasar pengukuran yang dijelaskan sebanyak 45.8%. Manakala, tiga lagi jenis item mencapai keperluan keseragaman alat ukuran sekurang-kurangnya 20% tetapi tidak melepasi batas minimum Rasch 40%. Dua lagi jenis item langsung tidak memenuhi kriteria asas keseragaman dimensi. Sehubungan itu, pengkaji harus meneliti semula setiap jenis item dan mengeluarkan item-item yang dianggap tidak berkualiti berdasarkan nilai *infit* dan *outfit mean square* (MNSQ).

Selain daripada teknik *Residual Principal Component Analysis* (PCA), keseragaman dimensi juga boleh disemak melalui aspek kekutuban item atau *item polarity* dengan meneliti nilai *point measure correlation* (PMC). Kekutuban item merupakan indikator awal kepada kesahan konstruk di mana nilai PMC yang positif menunjukkan item adalah selari dengan konstruk yang diukur dan bekerja kearah pengukuran konstruk tunggal (Bond & Fox 2015). Manakala nilai PMC item yang negatif membawa maksud yang sebaliknya dan perlu dibaiki atau digugurkan. Seterusnya, nilai kekutuban item yang baik dan memenuhi kriteria adalah melebihi 0.30 (Linacre 2011; Kline et al. 1999). Justeru, nilai kekutuban item positif yang tinggi

adalah lebih baik kerana ia menggambarkan keupayaan item untuk membezakan kemampuan di antara responden di samping mengukur konstruk tunggal. Dalam kajian ini, nilai PMC majoriti item tidak memenuhi kriteria dalam model Rasch (sila rujuk Lampiran 1). Ini mengesahkan lagi bahawa alat ukuran ICAR ini adalah bersifat multidimensional dan perlu dianalisis secara berasingan mengikut jenis item.

Pengukuran Item (*Item Measures*)

Antara kelebihan model pengukuran Rasch terletak pada keupayaannya untuk menyemak dan mengenal pasti kedudukan tepat bagi setiap item. Kesesuaian item (*item fit*) boleh dinilai melalui penelitian ke atas *mean-square* (MNSQ), nilai piawai *z* (*z-standard*) dan pola korelasi pengukuran (*point measure correlation*). Semakan kesesuaian item bermula dengan memastikan nilai *mean-square* (MNSQ) item yang tidak melebihi jumlah purata (*mean*) MNSQ dengan SD. Secara asasnya, nilai MNSQ dalam konteks Rasch merujuk kepada ketepatan atau kebolehan data sepadan dengan model yang mana akan memberikan nilai ideal sepadan = 1. Hasil analisis daripada perisian *Winsteps*, setiap kedudukan item disenaraikan mengikut tahap respons yang diperoleh daripada kajian yang dijalankan. Susunan item dalam senarai adalah berdasarkan

kepada bilangan responden tertinggi dan terendah dalam memberikan respons kepada sesuatu item. Item yang mempunyai kadar bilangan pilihan tertinggi seharusnya berada di aras yang paling rendah; manakala item yang mempunyai kadar bilangan pilihan yang terendah dan sukar dijawab perlu berada di aras yang paling tinggi. Oleh itu, item pada peringkat pertama boleh dikatakan sebagai item yang paling sukar dijawab oleh responden.

Kajian ini telah merujuk kualiti pengukuran dengan julat MNSQ pada kolom kedua-dua kolom *infit* dan *outfit* di mana $MNSQ = y$, $0.8 < y < 1.2$ adalah dikenalpasti dalam pengukuran *high stakes* yang produktif (Bond & Fox 2015).

Jadual 6 menunjukkan item 117, 121, 65, 107, 19, 53, 137 dan 109 dalam alat ukuran KKU adalah *misfit* dengan nilai outfit MNSQ > 1.2 logit. Bond dan Fox (2015) menyatakan bahawa nilai MNSQ di antara 0.4 sehingga 1.4 sudah boleh menggambarkan ketepatan item; walau bagaimanapun julat ini boleh berubah mengikut objektif dan jenis kajian yang dijalankan. Oleh itu, dalam kajian ini, item 117, 121 65, 107, 19, 53, 137 dan 109 merupakan item *outlier*. Dengan nilai logit tertinggi iaitu +2.52, item 23 dan 117 merupakan yang paling sukar dijawab oleh responden diikuti dengan item 21 (+ 2.51 logit) serta item 1, 32 dan 14 dengan nilai logit +1.80 bagi kesemuanya (sila rujuk Lampiran 2).

Jadual 6

Item Misfit

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	TOTAL MEASURE	MODEL S.E.	INFIIT MNSQ	OUTFIT MNSQ	PT-MEASURE CORR.	EXACT MATCH OBS%	MATCH EXP%	ITEM	
117	1	45	2.52	1.01	1.02	.4	.15	.05	97.8	97.8	AB4
121	3	45	1.37	.60	1.03	.2	.15	.09	93.3	93.3	AB8
65	3	45	1.37	.60	1.04	.2	.15	.09	93.3	93.3	TA42
107	5	43	.76	.48	1.06	.3	.15	.11	88.4	88.4	TA84
19	6	45	.60	.44	1.10	.4	.26	.12	86.7	86.7	P19
53	3	44	1.35	.60	1.05	.3	.15	.09	93.2	93.2	TA30
137	10	44	-.07	.36	1.10	.6	.22	.15	77.3	77.3	TM7
109	2	44	1.78	.73	1.03	.3	.09	.07	95.5	95.5	TA86
34	3	45	1.37	.60	1.04	.3	.10	.09	93.3	93.3	TA11
114	2	45	1.80	.73	1.03	.3	.08	.07	95.6	95.6	AB1
12	3	45	1.37	.60	1.04	.3	.09	.09	93.3	93.3	P12
16	5	45	.81	.48	1.05	.3	.10	.11	88.9	88.9	P16
118	8	45	.25	.39	1.06	.3	.08	.13	82.2	82.2	AB5
42	6	44	.57	.44	1.07	.3	.11	.12	86.4	86.4	TA19
64	3	45	1.37	.60	1.01	.2	.01	.09	93.3	93.3	TA41
87	7	44	.37	.42	1.05	.3	.08	.13	84.1	84.1	TA64
15	3	43	1.31	.60	1.03	.2	.05	.09	93.0	93.0	P15
17	8	44	.23	.39	1.03	.2	.02	.13	81.8	81.8	P17
94	2	44	1.78	.73	1.02	.3	.03	.07	95.5	95.5	TA71
66	13	44	-.43	.33	1.09	.7	.10	.16	70.5	70.5	TA43
3	6	44	.58	.44	1.05	.3	.06	.12	86.4	86.4	P3
8	4	45	1.06	.53	1.02	.2	.01	.10	91.1	91.1	P8
84	5	45	.81	.48	1.02	.2	.01	.11	88.9	88.9	TA61
153	26	45	-1.65	.31	1.11	1.5	.12	.17	48.9	59.6	TL12
26	11	45	-.16	.35	1.06	.4	.05	.15	75.6	75.6	TA3
6	5	45	.81	.48	1.03	.2	.01	.11	88.9	88.9	P6
23	1	45	2.52	1.01	.99	.3	.17	.05	97.8	97.8	P23
21	1	44	2.51	1.01	.98	.3	.20	.05	97.7	97.7	P21
10	3	45	1.37	.60	.98	.1	.23	.09	93.3	93.3	P10

Kesimpulan

Kajian ini bertujuan untuk menguji kesahan dan kebolehpercayaan alat ukuran keupayaan kognitif umum (ICAR) dalam konteks tempatan dengan menggunakan analisis Rasch. Hasil analisis polariti item dan *Principal Component Analysis* (PCA) menunjukkan sebahagian item yang diuji boleh menyumbang kepada kesahan alat ukuran keupayaan kognitif umum setelah dianalisis secara berasingan berdasarkan jenis item. Seterusnya, ujian kebolehpercayaan menunjukkan item-item boleh diterima untuk mengukur keupayaan kognitif umum; namun penambahbaikan dalam aspek keseimbangan di antara item mudah dan item sukar adalah diperlukan. Justeru, penambahan kepada item-item yang mudah diperlukan supaya akan lebih menggambarkan tahap keupayaan kognitif umum yang jitu dan tekal. Hal ini kerana secara purata kajian ini menunjukkan item kajian yang diperoleh daripada laman

sesawang ICAR adalah lebih sukar berbanding dengan kemampuan responden yang menjawab. Limitasi lain bagi kajian ini adalah jumlah sampel yang kecil ($n=45$) dan kekurangan kepelbagaian demografi. Selain itu, kajian replikasi kesahan dan kebolehpercayaan ke atas soal selidik KGU yang melibatkan jumlah sampel yang lebih besar dan pelbagai perlu dijalankan agar dapat membentuk norma keupayaan kognitif umum dalam kalangan sampel tempatan. Walaupun terdapat limitasi terhadap kajian ini, secara umumnya, reka bentuk kajian yang digunakan untuk menilai kebolehpercayaan, kesahan dan kesesuaian item dalam instrumen kajian agar sesuai dengan model adalah disokong oleh dapatan yang diperoleh. Oleh itu, boleh disimpulkan bahawa aplikasi model Rasch amat berguna dalam meningkat dan memperkuat kualiti item untuk mengukur secara berkesan konstruk yang ingin diukur.

Rujukan:

- Asfar, D., Born, M. P., Oostrom, J. K., & van Vugt, M. 2019. Psychological individual differences as predictors of refugees' local language proficiency. *European Journal of Social Psychology*, 49(7), 1385-1400.
- Azizi Yahaya. 2008. Kecerdasan. Dlm Hashim, S. & Azizi Yahaya. *Psikologi Pendidikan*. (hlm. 41-78). UTM Press: Fakulti Pendidikan/Jabatan Asas Pendidikan, Universiti Teknologi Malaysia.
- Azrilah Abdul Aziz. 2011. *Rasch Model Fundamentals: Scale Construct and Measurement Structure*: Integrated Advance Planning Sdn Bhd.
- Blum, D., Holling, H., Galibert, M. S., & Forthmann, B. 2016. Task difficulty prediction of figural analogies. *Intelligence*, 56, 72-81.
- Bond, T. & Fox, C.M. 2015. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. New York: Routledge.
- Conrad, K. J., Conrad, K. M., Dennis, M. L., Riley, B.B., & Funk, R. 2009. *Validation of the Behavioral Complexity Scale (BCS) to the Rasch Measurement Model*, GAIN Methods Report 1.1. Chicago, IL: Chestnut Health Systems.

- Edelen, M. O. & Reeve, B. B. 2007. Applying item response theory (IRT) modelling to questionnaire development, evaluation and refinement. *Quality Life Research*, 16, 5-18.
- Fisher, W.P. 2007. Rating scale instrument quality criteria. *Rasch Measurement Transactions*, 21(1), 1095.
- Fox, C. M. & Jones, J. A. 1998. Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology*, 45(1), 30.
- Ghazali Darusalam. (2008). Kesahan dan Kebolehppercayaan Dalam Kajian Kuantitatif and Kualitatif. *Jurnal Institut Perguruan Islam*. April.
- Hudiya Adzhar, Aidah Abdul Karim & Muhammad Uzair Sahrin. 2017. Pembangunan Instrumen Penerimaan E-Pembelajaran Pelajar Pascasiswazah Menggunakan Analisis Rasch. *Jurnal Pendidikan Malaysia*, 42(2), 147-155.
- Hunter, J. E. 1980. Validity generalization for 12,000 jobs: An application of synthetic validity and validity generalization to the General Aptitude Test Battery (GATB). Washington, DC: U.S. Department of Labour, Employment Service.
- Hunter, J. E. & Hunter, R. F. 1984. Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96(1), 72.
- Kamphaus, R. W., Winsor, A. P., Rowe, E. W. & Kim, S. 2005. A history of intelligence test interpretation. In D.P. Flanagan and P.L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd Ed.) (pp 23–38). New York: Guilford.
- Karim, A. A., Shah P. M., Din, R. & Lubis M. A. 2014. Developing information skills test for Malaysian youth students using Rasch analysis. *International Education Studies*, 7(13), 112-122.
- Kline, R. B. 1999. Book review: Psychometric theory. *Journal of Psychoeducational Assessment*, 17(3), 275-280.
- Linacre, J. M. 2011. A User's guide to WINSTEPS Ministeps; Raschmodel Computer Program. Program Manual 3.7.3.
- Mahalingam, V. 2016. Abstract reasoning items. Dlm *The International Cognitive Ability Resource*. <https://icar-project.com/>.
- Mona Isa, Mazlan Abu Bakar, Ibrahim Sipan, Mohamad Sufian Hasim, Ahmad Ezanee Hashim & Mohd Khairul Anuar Jalil. 2016. Measuring Instrument Constructs of Return Factors for Green Office Building Investments Variables Using Rasch Measurement Model. In MATEC Web of Conferences (Vol. 66, p. 00125). EDP Sciences.
- Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N. & Ceci, S. J., et al. 1996. Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77–101.
- Nik Hairi Omar & Rusyda Helma Mohd. 2020. Ujian Aptitud Sebagai Pengukur Kebolehpasaran Mahasiswa Sains Sosial. *Jurnal Psikologi Malaysia*, 34(2), 132-143.
- Olsen, L. W. 2003. Essays on Georg Rasch and his contributions to statistics. (Unpublished PhD Thesis), Copenhagen, Institute of Economics, University of Copenhagen.

- Ones D. S., Dilchert S. & Viswesvaran C. 2012. Cognitive abilities. Dlm N. Schmitt (Ed.) *Oxford library of psychology. The Oxford handbook of personnel assessment and selection* (p. 179–224). Oxford University Press.
- Reeve, C. L. & Hakel, M. D. 2002. Asking the right questions about g. *Human Performance*, 15, 47–74.
- Salgado, J. F. 2017. Using ability tests in selection. In H. W. Goldstein, E. D. Pulakos, J. Passmore, & C. Semedo (Eds.), *Wiley Blackwell handbooks in organizational psychology. The Wiley Blackwell handbook of the psychology of recruitment, selection and employee retention* (p. 115–150). Wiley Blackwell.
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., De Fruyt, F. & Rolland, J. P. 2003. A meta-analytic study of general mental ability validity for different occupations in the European community. *Journal of Applied Psychology*, 88(6), 1068.
- Schmidt, F. L. & Hunter, J. E. 1998. The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274.
- Schmidt, F. L. 2002. The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, 15(1-2), 187-210.
- Schmidt, F. L. & Hunter, J. 2004. General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86(1), 162.
- Sick, J. 2010. Assumptions and requirements of Rasch measurement. *JALT Testing & Evaluation SIG Newsletter*, 14(2), 23-29.
- Siti Rahayah Ariffin. 2008. Inovasi dalam Pengukuran dan Penilaian Pendidikan. Bangi: Fakulti Pendidikan UKM.
- Sumintono, B. & W. Widhiarso. 2014. *Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial*. Edisi Revisi. Jakarta: Trim Komunikata Publishing House.
- The International Cognitive Ability Resource Team. 2014. <https://icar-project.com/>.
- Wright, B. D. & Mok, M. M. 2004. An overview of the family of Rasch measurement models. *Introduction to Rasch Measurement: Journal of Applied Measurement*, 1-24.