Malaysian Journal of Social Sciences and Humanities (MJSSH) (e-ISSN : 2504-8562)
2022, Volume 7, Issue 10, e001806
DOI: https://doi.org/10.47405/mjssh.v7i10.1806

Analisis Model Rasch Penilaian Akhir Bahasa Arab Kolej Vokasional

(Vocational College Arabic Test Analysis with Rasch Measurement)

Mior Ahmad Kamil Mohd Lot[1*], Ahmad Zamri Khairani[2]

[1]Pusat Pengajian Ilmu Pendidikan, Universiti Sains Malaysia, 11800 Gelugor, Pulau Pinang, Malaysia.
Email: miorahmadkamil@student.usm.my
[2]Pusat Pengajian Ilmu Pendidikan, Universiti Sains Malaysia, 11800 Gelugor, Pulau Pinang, Malaysia.
Email: ahmadzamri@usm.my

CORRESPONDING AUTHOR (*):
Mior Ahmad Kamil Mohd Lot
(miorahmadkamil@student.usm.my)

KATA KUNCI:
Analisis Item
Pengukuran Model Rasch
Bahasa Arab Kolej Vokasional

KEYWORDS:
Item Analysis
Rasch Measurement
Vocational College Arabic Course

CITATION:
Mior Ahmad Kamil Mohd Lot & Ahmad Zamri Khairani. (2022). Analisis Model Rasch Penilaian Akhir Bahasa Arab Kolej Vokasional. *Malaysian Journal of Social Sciences and Humanities (MJSSH), 7*(10), e001806. https://doi.org/10.47405/mjssh.v7i10.1806

ABSTRAK

Kajian ini bertujuan menganalisis item Penilaian Akhir kursus bahasa Arab Kolej Vokasional, Kementerian Pendidikan Malaysia di enam buah Kolej Vokasional di Negeri Perak. Analisis item melibatkan ujian bertulis kertas penilaian akhir sesi pengajian 2020 melalui penglibatan 244 orang responden bagi kertas ELA3312 dan 245 orang responden bagi kertas ELA4322. Analisis deskriptif model Rasch daripada perisian Winstep 3.71.0.1 digunakan bagi menilai statistik item melalui analisis unidimensi ujian, kesesuaian item, polariti item, indeks pemisahan dan kebolehpercayaan item dan pelajar serta analisis bias DIF item dalam menentukan kesesuaian 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 atau dilibatkan sebagai item bank soalan bahasa Arab. Nisbah kesukaran item juga adalah bersesuaian dengan konstruk pembelajaran kursus bahasa Arab Kolej Vokasional. Melalui dapatan yang diperoleh diharapkan dapat membantu pengkaji merungkap penghasilan item yang lebih baik sekaligus meningkatkan kualiti pentaksiran dan penilaian bahasa Arab di Kolej Vokasional.

ABSTRACT

This study aims to analyse the final assessment test of the Vocational College Arabic course, Ministry of Education Malaysia in six Vocational Colleges in the State of Perak. Item analysis involves the written test of the final assessment paper of the 2020 academic session through the involvement of 244 respondents for the ELA3312 paper and 245 respondents for the ELA4322 paper. Descriptive analysis of the Rasch model from the Winstep

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

3.71.0.1 software was used to evaluate item statistics through unidimensional analysis of the test, item fit, item polarity, separation and reliability index of item and person as well as item DIF analysis in determining the suitability of the items involved as the final test. This analysis aims to confirm the quality of the items involved from a statistical point of view as well as detect the suitability of the items to be used or modified based on the content standards to achieve the learning outcomes. Findings show that most of the items meet the Rasch requirements. Items are suitable to be reused as test items or included as Arabic bank items. The difficulty ratio of the items is also in accordance with the learning construct of the Vocational College Arabic course. Through the findings obtained it is hoped to help the researcher reveal the development of better items as well as improve the assessment proses of Arabic course in Vocational Colleges.

Sumbangan/Keaslian: Kajian ini menyumbang kepada literatur semasa berkenaan pengukuran model Rasch terhadap item ujian terutama dalam konteks bahasa Arab Kolej Vokasional. Dapatan serta perbincangan daripada kajian ini diharapkan membantu proses pembinaan instrumen penilaian yang baik sekali gus membantu proses PdPc yang lebih berkualiti.

1. Pengenalan

Transformasi pendidikan vokasional atau TVET merupakan proses pendidikan dengan hala tuju penyediaan tenaga kerja mahir negara. Penawaran TVET adakah selaras dengan keperluan industri serta menyumbang kepada pertumbuhan ekonomi berasaskan pengetahuan, kemajuan teknologi serta mobiliti tenaga kerja global (KPM, 2019). Penawaran TVET di Kolej Vokasional menjadi titik tolak dalam penghasilan tenaga kerja negara yang mapan dan berkualiti tinggi adalah berdasarkan kerangka pembelajaran berasaskan hasil atau *outcome based education* (OBE) serta selaras dengan hasrat KPM bagi menjadikan pendidikan vokasional sebagai satu cabang pendidikan perdana dalam kalangan pelajar di Malaysia.

Pembelajaran bahasa Arab sesuatu yang tidak asing di Malaysia malah bahasa Arab telah bertapak lama bermula dari peringkat rendah hingga ke peringkat tertiari. Bahasa Arab mendapat tempat di Malaysia apabila akta pendidikan mentakrif ianya sebagai salah satu bahasa asing yang boleh dipraktikkan dan di ajar dengan tujuan menyediakan individu pengetahuan, kemahiran serta peningkatan kendiri individu pada masa hadapan (Akta Pendidikan, 1996). Pembelajaran bahasa Arab di Kolej Vokasional melibatkan penawaran sebagai kursus elektif bahasa kedua pada tahun akhir pengajian Diploma Vokasional Malaysia (DVM). Kursus bahasa Arab di Kolej Vokasional memberi penumpuan terhadap empat kemahiran asas berbahasa (dengar, tutur, tulis dan baca) di samping pembelajaran kaedah asas bahasa Arab seperti kata nama, kata kerja, kata hubung, kata ganti nama dan sebagainya untuk persediaan pelajar berkemahiran bahasa Arab bagi pasaran pekerjaan yang lebih luas.

Struktur pengajian bahasa Arab di Kolej Vokasional adalah merangkumi empat konstruk atau standard pembelajaran iaitu, huruf Arab, perkataan Arab, ayat pendek serta perbualan ringkas. Silibus bahasa Arab semester 3 memberi penumpuan terhadap

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

konstruk huruf, perkataan dan ayat pendek manakala di semester 4, penumpuan adalah menjurus kepada perkataan, ayat pendek dan perbualan ringkas bahasa Arab. Namun konstruk penilaian adalah berkisar kepada tiga standard sahaja iaitu, huruf, perkataan dan ayat pendek sahaja kerana perbualan ringkas dikategorikan sebagai konstruk ayat pendek. Jadual 1 merujuk konstruk tersebut berdasarkan skop kandungan pembelajaran seterusnya mencapai hasil pembelajaran yang ditetapkan.

Jadual 1: Konstruk Pembelajaran Bahasa Arab Kolej Vokasional

Bil	Konstruk	Huraian	Contoh	Hasil Pembelajaran
1	Huruf	29 huruf-huruf *Hījāīyyah*.	ك ت ب / ق ر أ *kāf, tā', bā' / qāf, rā', alif*	-Menggunakan huruf Arab melalui penulisan dan lisan dengan betul mengikut kedudukan yang dinyatakan. -Menggunakan perkataan bahasa Arab bagi membentuk ayat pendek dengan betul
2	Perkataan	gabungan huruf-huruf *Hījāīyyah* yang membina perkataan seperti kata nama atau kata kerja.	كتب / قرأ *kātābā / qārāā*	
3	Ayat pendek	gabungan perkataan-perkataan yang membina ayat yang memberi maksud.	كتب احمد كتاب / قرأ محمد كتاب. *kātābā ahmad kitābun / qārāā muhammad kitābun*	
4	Perbualan ringkas	gabungan ayat-ayat yang membentuk perbualan ringkas.	- mengikut tema atau kandungan pembelajaran berkaitan.	

Penilaian bahasa Arab di Kolej Vokasional adalah berlandaskan *Outcome-based Education* (OBE) yang meliputi aspek kognitif, psikomotor dan afektif pelajar yang dirangkum melalui pentaksiran berterusan dan penilaian akhir. Penilaian merupakan sebahagian daripada kitaran OBE bertujuan memastikan hasil yang ditetapkan pada sesuatu pembelajaran dapat dicapai melalui kaedah yang tepat. Kaedah penilaian yang autentik mampu menilai konstruk pembelajaran melalui analisis perbandingan antara pelajar (Damit & Omar, 2018). Penilaian yang salah memberi justifikasi yang tidak tepat terhadap pencapaian pelajar. Kesahan pengukuran adalah penting kerana ia adalah faktor utama yang menyumbang kepada keabsahan hasil yang diperoleh (Agensi Kelayakan Malaysia, 2014) seterusnya menyumbang kepada perolehan maklumat yang membantu perkembangan proses pembelajaran dan pengajaran (Marina & Jamil, 2015).

1.1. Objektif Kajian

Secara umum, kajian ini bertujuan menganalisis item penilaian akhir bahasa Arab semester 3 dan semester 4 Kolej Vokasional, KPM. Bagi mencapai tujuan tersebut, beberapa analisis berpandukan model pengukuran Rasch akan dijalankan bagi mengetahui ciri-ciri psikometrik item penilaian akhir bahasa Arab. Pemerhatian akan diberikan terhadap beberapa prinsip asas Rasch iaitu unidimensi ujian, kesesuaian item dengan andaian model, indeks pemisahan dan kebolehpercayaan yang tinggi, DIF item ujian serta padanan antara pelajar-item yang baik.

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

2. Sorotan Literatur

Model Rasch merupakan teori andaian yang menjadi sebahagian daripada Teori Respons Item atau *Item Response Theory* (IRT). IRT membuat pengukuran berdasarkan tiga parameter pengukuran iaitu a) diskriminasi item, b) kesukaran item dan c) tekaan berbanding model Rasch yang memberi penumpuan analisis melalui *parameter b* berdasarkan corak respons item. Kerangka pengukuran model Rasch adalah tidak asing di dalam dunia pendidikan kerana kemampuan model dari sudut pengesahan pembinaan instrumen, analisis kualiti serta analisis pencapaian individu melalui pengukuran dalam menghasilkan dapatan yang empirikal (Boone et al., 2014). Pelbagai kajian lalu menjadikan model pengukuran Rasch sebagai pengesahan item atau ujian yang lebih berkualiti seperti pembinaan bank item (Samsudin et al., 2020b; Hasni, 2020), kalibrasi item (Samsudin et al., 2020a; Hasni, 2014), pengesahan soal selidik (Mahamud & Othman, 2019; Yasin et al., 2015) mahupun ujian menulis esei (Isnani et al., 2019).

Model Rasch berkemampuan menjelaskan kesahan instrumen, pengesanan kualiti item serta pembinaan skala pengukuran bercirikan saintifik yang cekap dan dipercayai berdasarkan fungsi logaritma asas pengukuran menggunakan unit *log odds unit* atau *logits* atas satu skala linear (Azrilah et al., 2017; Boone, 2016, 2020; Sumintono, 2018). Menurut Bond et al. (2021), unidimensi, kesesuaian item, andaian kesukaran item/kebolehan pelajar serta kebolehpercayaan pengukuran merupakan prinsip asas yang perlu diberi penumpuan dalam pengukuran model Rasch. Berdasarkan model Rasch, aras kesukaran item ditentukan oleh respons pelajar terhadap item. Item sukar dijangka mendapat respons betul oleh pelajar berkebolehan tinggi berbanding pelajar berkebolehan sederhana atau rendah. Selain prinsip asas tersebut, beberapa analisis lain seperti fungsi item terbeza (DIF) dan peta pelajar-item (*Wright map*) memberi gambaran pengukuran terhadap item mahupun kebolehan pelajar.

2.1. Prinsip Asas Model Rasch

Analisis unidimensi ujian. Analisis unidimensi merujuk analisis yang dilakukan bagi menilai keseragaman pengukuran yang dilakukan adalah mengukur konstruk yang tunggal (Yu, 2020) atau menguji satu kebolehan khusus (Kolen & Brennan, 2014). Faktor residual yang kecil, rawak dan tidak menunjukkan sesuatu konstruk khusus merupakan bukti unidimensi model Rasch (Sick, 2010). Item yang berada dalam julat *Infit MNSQ* dan *Outfit MNSQ* yang diterima juga menggambarkan kesahan unidimensi ujian (Boone, 2020).

Kesesuaian item dengan jangkaan model Rasch. Statistik *Infit MNSQ* dan *Outfit MNSQ* digunakan bagi mengesan kesesuaian item dengan keperluan model Rasch. Saranan julat adalah bergantung kepada bentuk ujian (Bond & Fox, 2015), julat 0.7 hingga 1.3 merupakan julat kebiasaan yang digunakan (Bond et al., 2021; Sbeglia & Nehm, 2018; Shamsuddin, 2014) namun julat yang lebih besar sekitar .5 hingga 1.5 juga diterima (Linacre, 2012). Item yang menjauhi julat .5 -1.5 *logits* sepatutnya digugurkan daripada analisis kerana tidak berfungsi seperti yang dikehendaki model Rasch (Boone, 2020). Nilai z-std juga memberi maklumat keserasian item dengan jangkaan model. Item dengan julat nilai *z-std* ± 2 hendaklah diperhatikan. Item menjauhi nilai *z-std* 2.00 *logits* dianggap terkeluar daripada pengukuran. Namun analisis z-std boleh diabaikan sekiranya julat *infit/outfit* MNSQ telah dipenuhi.

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Polariti Item. Pola Korelasi Pengukuran atau *Point Measure Correlation* (PMC) bertujuan menilai item yang menjurus kepada penilaian satu konstruk. Nilai polariti item seharusnya berada dalam julat positif (Azrilah et al., 2017; Linacre, 2012; Sumintono, 2018). Polariti item yang positif menunjukkan keserasian item adalah sejajar dengan kebolehan pelajar (Bond & Fox, 2015). Menurut Boone (2020), item yang memiliki nilai polariti melebihi .30 *logits* membuktikan pengukuran konstruk tunggal dipenuhi. Nilai negatif menunjukkan bahawa item tersebut perlu digugurkan daripada pengukuran. Namun ralat kemasukan data boleh menyebabkan nilai item menjadi negatif (Boone, 2020).

Indeks pemisahan dan kebolehpercayaan yang tinggi. Di dalam model Rasch, indeks pemisahan dan kebolehpercayaan merupakan pembuktian terhadap kesahan sesebuah instrumen yang ditadbir terhadap responden. Indeks pemisahan pula merujuk keupayaan pengukuran yang dibuat bagi mendiskriminasikan pelajar atau item kepada beberapa aras kebolehan pelajar atau kesukaran item. Menurut Bond dan Fox, (2015), indeks pemisahan bagi pelajar dan item adalah seharusnya melebihi nilai < 2.0. Julat indeks kebolehpercayaan adalah daripada 0 hingga 1.00 di mana nilai terbesar adalah lebih baik. Kebolehpercayaan item merujuk item dapat berfungsi sebagaimana jangkaan kedudukan item walaupun ditadbir kepada sampel yang lain (Bond et al., 2021; Bond & Fox, 2015). Nilai indeks kebolehpercayaan bagi pelajar dan item adalah .8 dan .9 dan ke atas (Linacre, 2012).

DIF contrast. Differential Item Functioning (DIF) atau fungsi item terbeza merujuk analisis terhadap item dari sudut bias item terhadap mana-mana kumpulan. Bias item boleh berlaku apabila item hanya berfungsi atau menarik perhatian sebahagian kumpulan sahaja (Sick, 2010). Sebagai contoh, kandungan item berkenaan sukan lelaki lebih menarik minat lelaki berbanding perempuan. Julat bias item adalah sekitar ±.5 *logits* (Azrilah et al., 2017; Bond & Fox, 2015). Nilai DIF adalah penting bagi menjamin keadilan sesebuah ujian supaya tidak bias kepada mana-mana kumpulan (Hasni, 2014).

Padanan item dan pelajar yang baik. Peta pelajar-item merujuk *Wright map* yang memaparkan hierarki kebolehan pelajar dan kesukaran item pada satu skala sela menggunakan unit *logits* (Boone, 2016). Item pada kedudukan atas pada skala *Wright map* menunjukkan aras kesukaran yang tinggi berbanding item pada kedudukan bawah. Begitu juga bagi kedudukan kebolehan pelajar, pelajar pada kedudukan atas menunjukkan pelajar yang berkebolehan tinggi berbanding pelajar pada kedudukan bawah. *Wright map* memberi panduan kepada pengkaji mengesan dapatan andaian kesukaran item berbanding susunan kesukaran yang diharapkan ketika pembinaan item.

3. Metod Kajian

Kajian ini bertujuan melihat kesesuaian item penilaian akhir semester tiga dan empat yang ditadbir kepada pelajar tahun akhir Kolej Vokasional sesi 2020. Instrumen yang digunakan adalah kertas penilaian akhir bahasa Arab ELA3312 dari semester 3 dan ELA4322 dari semester 4. Jadual 2 merujuk demografi responden yang dilibatkan dalam kajian ini daripada 6 buah Kolej Vokasional di Negeri Perak yang dipilih berdasarkan pensampelan bertujuan. Analisis deskriptif terhadap item ujian adalah berpandukan pengukuran model Rasch menggunakan perisian Winstep 3.71.0.1.

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Jadual 2: Taburan Responden Ujian

Ujian	Lelaki	%	Perempuan	%	Jumlah
ELA3312	184	74.80	62	25.20	246
ELA4322	182	74.59	62	25.41	244
Jumlah keseluruhan	366		124		490

Kajian ini merupakan kajian kuantitatif deskriptif adalah berdasarkan kerangka pengukuran Rasch melalui analisis unidimensi, kesesuaian item, polariti item, indeks pemisahan dan kebolehpercayaan item dan pelajar serta *DIF contrast* item dalam menentukan kesesuaian 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 melalui perisian Winstep 3.71.0.1. Prosedur analisis tersebut akan dijalankan secara berasingan bagi kedua-dua ujian yang dilibatkan. Transformasi data adalah dalam unit *logits* yang mengandungi maklumat kesukaran item serta kebolehan pelajar bagi menilai ciri-ciri psikometrik item sekali gus mencapai objektif kajian. Analisis item akan menggunakan tetapan statistik daripada Jadual 3 sebagai panduan bagi interpretasi dapatan analisis.

Jadual 3: Tetapan Statistik Analisis Item

Statistik	Nilai	Sumber
Unidimensi		
Raw variance explained by measures	20% - 40%	Azrilah et al. (2017)
Unexplained variance in 1st contrast	>15%	Azrilah et al. (2017)
Kesesuaian Item		
Infit MNSQ	0.5 - 1.5	Linacre (2012)
Outfit MNSQ	0.5 - 1.5	Linacre (2012)
Polariti item (PMC)	< .0	Linacre (2012)
DIF item	± 0.5	Azrilah et al. (2017)
Indeks Pemisahan dan Kebolehpercayaan		
Pemisahan pelajar	< 2.0	Bond & Fox (2015)
Pemisahan item	< 2.0	Bond & Fox (2015)
Kebolehpercayaan pelajar	< 0.8	Linacre (2012)
Kebolehpercayaan item	< 0.9	Linacre (2012)

4. Hasil dan Perbincangan Kajian

Analisis bagi setiap ujian dijalankan secara berasingan terhadap kedua-dua ujian bahasa Arab Kolej Vokasional menunjukkan padanan 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. 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. (2017), 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. Nilai *Unexplained variance in 1st contrast* juga seharusnya tidak melebihi nilai kawalan 15%. Jadual 4 dan Jadual 5 menunjukkan varian residual bagi ujian ELA3312 dan ELA4322. Nilai *Raw variance explain by measure* adalah 27.5% bagi ujian ELA3312 dan 32.9% bagi

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

ELA4322. nilai *eigen* tersebut adalah sesuai kerana berada di dalam julat pengukuran. Nilai *Unexplained variance in 1st contrast* juga adalah baik bagi kedua-dua ujian dengan nilai *eigen* masing-masing adalah 6.7% bagi ELA3312 dan 9.1% bagi ELA4322. Nilai faktor yang kecil menunjukkan kesahan pengukuran yang unidimensi (Sick, 2010).

Jadual 4: Analisis PCA ELA3312

	Empirical			Modeled
Total raw variance in observations	= 41.4	100.0%		100.0%
Raw variance explained by measures	= 11.4	27.5%		27.2%
Raw unexplained variance (total)	= 30.0	72.5%	100.0%	72.8%
Unexplned variance in 1st contrast	= 2.8	6.7%	9.3%	

Jadual 5: Analisis PCA ELA4322

	Empirical			Modeled
Total raw variance in observations	= 44.7	100.0%		100.0%
Raw variance explained by measures	= 14.7	32.9%		33.3%
Raw unexplained variance (total)	= 30.0	67.1%	100.0%	66.7%
Unexplned variance in 1st contrast	= 4.1	9.1%	13.6%	

Kesesuaian Item. Dalam penentuan kesesuaian item, perhatian akan diberikan kepada beberapa analisis iaitu, kesukaran item, *infit*/*oufit* MNSQ serta polariti item. *Measure* dalam Winstep merujuk andaian tentu ukuran parameter seperti kesukaran item dan dilaporkan dalam unit *logits.* Semakin tinggi nilai *measure* menunjukkan semakin sukar item tersebut dan begitulah sebaliknya. Jadual 6 menunjukkan taburan julat kesukaran item bagi ujian ELA3312 dan ELA4322 dengan masing-masing memiliki satu item melebihi kesukaran 3.00 *logits*. Nilai kesukaran terendah adalah pada julat -2.01 hingga 2.50 *logits* sekitar 1 dan 2 item bagi kedua-dua ujian tersebut. Empat belas item berada di atas min kesukaran item untuk Ujian ELA3312 berbanding 13 item bagi ujian ELA4322.

Jadual 6: Julat Kesukaran Item Ujian

Kesukaran Item (*logits*)	**Jumlah Item**	
	Ujian ELA3312	**Ujian ELA4322**
Atas 3.00	0	1
2.50 hingga 2.99	0	2
2.00 hingga 2.49	0	0
1.50 hingga 1.99	2	3
1.00 hingga 1.49	7	0
.50 hingga 0.99	1	0
.00 hingga .49	5	7
-0.01 hingga -0.50	6	7
-0.51 hingga -1.00	2	4
-1.01 hingga -1.50	3	1
-1.51 hingga -2.00	3	3
-2.01 hingga -2.50	1	2
-2.51 hingga -3.00	0	0
Bawah -3.00	0	0
Jumlah Item	30	30

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Menurut Hasni (2014), penentuan item sukar dan mudah ujian boleh dilakukan melalui nilai sisihan piawai kesukaran item. Item dengan nilai kesukaran +1 daripada nilai sisihan piawai kesukaran item ditakrif sebagai item sukar manakala item dengan nilai kesukaran -1 daripada nilai sisihan piawai kesukaran item ditakrif sebagai item mudah dan baki item dirujuk sebagai item sederhana. Melalui kaedah ini, item-item ujian dapat dikelaskan kepada tiga kumpulan item dengan ujian ELA3312 pada nilai sisihan piawai kesukaran 1.21 *logit* dan ujian ELA4322 pada nilai sisihan piawai kesukaran item 1.46 *logits*. Jadual 7 menunjukkan taburan kumpulan kesukaran item mengikut ujian. Analisis kesukaran item ujian ELA3312 menunjukkan item sukar=8, item sederhana=15 dan item mudah=7 manakala ujian ELA4322 menunjukkan item sukar=6, item sederhana=19 dan item mudah=5.

Jadual 7: Aras Kesukaran Item Ujian

Kesukaran Item (*logits*)	**Jumlah Item**	
	Ujian ELA3312	**Ujian ELA4322**
Item sukar	8	6
Item sederhana	15	19
Item mudah	7	5
Jumlah Item	30	30

Berdasarkan Jadual 8, dapatan bagi menilai kesesuaian item menunjukkan bahawa kesemua item daripada ujian ELA3312 dan ELA4322 berada dalam julat item yang munasabah bagi *infit MNSQ* iaitu masing-masing antara 0.85 hingga 1.32 *logits* dan 0.83 hingga 1.21 *logits*. Manakala bagi *outfit MNSQ* item masing-masing menunjukkan julat nilai antara 0.39 *logits* hingga 1.50 *logits* bagi ujian ELA3312 dan julat antara 0.55 *logits* hingga 1.34 logit bagi ujian ELA4322. Berdasarkan Jadual 8, taburan nilai *infit/outfit MNSQ* adalah tinggi dalam julat 0.50 hingga 0.99 bagi kedua-dua ujian.

Jadual 8: Julat *Infit* dan *Outfit MNSQ* Ujian

	ELA3312		**ELA4322**	
	Infit MNSQ	***Outfit MNSQ***	***Infit MNSQ***	***Outfit MNSQ***
1.51 ke atas	0	0	0	0
1.00 hingga 1.50	12	9	13	11
.50 hingga 0.99	18	19	17	19
Bawah .49	0	2	0	0
Jumlah Item	30		30	

Jadual 9 merujuk item yang menjauhi julat *infit/outfit MNSQ* yang dibenarkan. Kedua-dua ujian menunjukkan nilai diterima bagi *infit MNSQ*.

Jadual 9 - Item Tersasar daripada Julat Kesesuaian Item

Ujian	**Item dengan nilai julat *infit/outfit MNSQ* tersasar (0.5 < MNSQ < 1.5)**		**Jumlah Item**
	infit MNSQ	***outfit MNSQ***	
ELA3312	-	item1 (.39 *logits*) dan item12 (.40 *logits*)	2
ELA4322	-	-	0
Jumlah Item			2

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Namun julat *outfit MNSQ* terkeluar dikesan pada dua item daripada ujian ELA3312 iaitu item1 (0.38 *logits*) dan item11 (0.40 *logits*) yang menunjukkan variasi terendah di bawah 0.5 *logits*. Nilai *outfit* MNSQ di bawah < 0.5 *logits* tidak melemahkan pengukuran (Linacre, 2012), namun menyebabkan dapatan pekali indeks pemisahan dan kebolehpercayaan yang diperoleh adalah mengelirukan (Azrilah et al., 2017).

Polariti item ujian yang positif menunjukkan keserasian item adalah sejajar dengan kebolehan pelajar namun bagi kesahan konstruk, polariti item seharusnya memiliki nilai korelasi > .30 *logits*. Jadual 10 menunjukkan taburan pola korelasi item berada dalam julat positif bagi kesemua item ujian. Kertas ELA4322 menunjukkan 18 item berada bawah nilai <.30 berbanding hanya 13 bagi kertas ELA4322. Item yang menunjukkan nilai polariti melebihi >.30 bagi ujian ELA 3312 ialah item15, item21, item24, item23, item12, item14, item11, item27, item26, item29, item30 dan item28 manakala bagi ujian ELA4322 ialah item9, item14, item18, item28, item19, item26, item12, item20, item29, item30, item27, item23, item15, item25, item22, item24 dan item21. Menurut Boone, (2020), nilai korelasi melebihi .30 *logits* menunjukkan kesahan konstruk dipenuhi. Linacre (2012) berpendapat, nilai positif adalah memadai kerana ianya menunjukkan item yang ditadbir adalah sejajar dengan kebolehan pelajar. Julat lebih tinggi boleh ditetapkan bagi ujian khas seperti ujian kemasukan atau peperiksaan awam yang memerlukan item dengan kesahan konstruk yang lebih baik (Azrilah et al., 2017).

Jadual 10: Taburan Pola Korelasi Pengukuran Item Ujian

Nilai *Pt-mea Corr*. Item	**Ujian ELA3312**	**Ujian ELA4322**
.00 hingga .19	7	5
.20 hingga .29	11	8
.30 hingga .39	5	10
.40 hingga .49	3	6
.50 hingga .59	4	1
Jumlah Item	30	30

Jadual 11 merujuk indeks pemisahan pelajar dan kebolehpercayaan pelajar bagi ujian ELA3312 dan Jadual 12 bagi ujian ELA4322.

Jadual 11 : Indeks Pemisahan dan Kebolehpercayaan Pelajar ELA3312

	Jumlah Skor	**Bil**	**Kesukaran**	**Ralat Model**	**Infit**		**Outfit**	
					MNSQ	**ZSTD**	**MNSQ**	**ZSTD**
Min	31.7	30.0	1.95	.45	.98	.0	.93	.0
Sisihan Piawai	4.7	.0	.81	.13	.31	1.0	.52	.9
Max	39.0	30.0	4.10	1.01	2.43	4.7	3.41	5.1
Min	12.0	30.0	-.69	.34	.44	-2.2	.26	-2.3
RMSE Sebenar	.53	Sebenar S.D	.72	Pemisahan	1.36	Kebolehpercayaan Pelajar		.65
RMSE Model	.51	Sebenar S.D	.73	Pemisahan	1.44	Kebolehpercayaan Pelajar		.67
S.E Min Item = .06								

Dapatan masing-masing menunjukkan nilai 1.42 *logits* dan .67 *logits* bagi ujian ELA3312 manakala 1.72 *logits* dan .75 *logits* bagi ujian ELA4322. Nilai tersebut adalah lebih rendah berbanding keperluan pengukuran yang memerlukan nilai pemisahan melebihi < 2 dan kebolehpercayaan < .8. Nilai kebolehpercayaan pelajar menunjukkan bahawa ujian

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

yang ditadbir hanya mampu mendiskriminasi pelajar kepada 2 atau 3 aras kebolehan. Menurut Linacre (2012), nilai pemisahan yang rendah menunjukkan ujian yang ditadbir tidak mampu untuk membezakan antara prestasi responden. Menurut beliau lagi, nilai kebolehpercayaan yang tinggi adalah bergantung kepada beberapa aspek seperti taburan kebolehan pelajar yang pelbagai, saiz ujian yang lebih panjang atau sasaran ujian terhadap kumpulan tertentu. Nilai *infit/outfit MNSQ* yang rendah juga boleh mengakibatkan dapatan yang rendah bagi indeks pemisahan dan kebolehpercayaan (Azrilah et al., 2017).

Jadual 12: Indeks Pemisahan dan Kebolehpercayaan Pelajar ELA4322

	Jumlah Skor	Bil	Kesukaran	Ralat Model	Infit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
Min	28.3	30.0	1.08	.44	.99	.0	.96	.0
Sisihan Piawai	5.1	.0	.99	.18	.31	1.0	.53	1.0
Max	40.0	30.0	5.94	1.90	1.93	3.0	4.11	4.9
Min	13.0	30.0	-.97	.34	.41	-2.7	.16	-2.0
RMSE Sebenar	.50	S.P Sebenar	.85	Pemisahan	1.72	Kebolehpercayaan Pelajar		.75
RMSE Model	.47	S.P Sebenar	.86	Pemisahan	1.83	Kebolehpercayaan Pelajar		.77
S.E Min Item = .06								

Jadual 13 menunjukkan indeks pemisahan item pada 5.42 *logits* dan kebolehpercayaan item pada .97 *logits* bagi ujian bahasa Arab ELA3312. Manakala bagi ujian ELA4322 (rujuk Jadual 14) mendapati indeks pemisahan item berada pada 8.08 *logits* dan kebolehpercayaan item pada .98 *logits*. Indeks tersebut adalah tinggi kerana melebihi nilai pemisahan < 2.0 dan kebolehpercayaan .9. Nilai indeks pemisahan item yang tinggi menunjukkan jumlah sampel yang dilibatkan adalah mencukupi bagi penentuan hierarki kesukaran item ujian (Linacre, 2012). Dapatan indeks kebolehpercayaan item yang tinggi juga menunjukkan variasi kesukaran item yang pelbagai di samping saiz responden yang dilibatkan adalah mencukupi. Dapatan analisis indeks pemisahan dan kebolehpercayaan pelajar yang rendah menggambarkan kebolehan pelajar adalah setara dalam kedua-dua ujian yang ditadbir atau ujian tidak dapat membezakan antara kumpulan pelajar, kebarangkalian item perlu diubahsuai bagi meningkatkan jumlah item sukar ujian sekaligus menjadikan instrumen ujian lebih sensitif membezakan aras pelajar.

Jadual 13: Indeks Pemisahan dan Kebolehpercayaan Item ELA3312

	Jumlah Skor	Bil	Kesukaran	Ralat Model	Infit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
Min	260.4	246.0	.00	.20	1.00	.0	.93	-.2
Sisihan Piawai	73.4	.0	1.10	.09	.11	1.3	.23	1.5
Max	439.0	246.0	1.76	.46	1.32	3.9	1.50	4.9
Min	134.0	246.0	-2.21	.09	.85	-2.1	.39	-2.7
RMSE Sebenar	.22	S.P Sebenar	1.08	Pemisahan	4.89	Kebolehpercayaan Item		.96
RMSE Model	.22	S.P Sebenar	1.08	Pemisahan	4.94	Kebolehpercayaan Item		.96
S.E Min Item = .21								

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Jadual 14: Indeks Pemisahan dan Kebolehpercayaan Item ELA4322

	Jumlah Skor	Bil	Kesukaran	Ralat Model	Infit		Outfit	
					MNSQ	ZSTD	MNSQ	ZSTD
Min	230.4	244.0	.00	.17	1.00	.3	.96	.0
Sisihan Piawai	117.9	.0	1.46	.06	.10	1.4	.22	1.7
Max	430.0	244.0	4.05	.31	1.25	4.5	1.34	4.0
Min	18.0	244.0	-2.30	.09	.83	-2.2	.55	-2.6
RMSE Sebenar	.18	S.P Sebenar	1.45	Pemisahan	8.08	Kebolehpercayaan Item		.98
RMSE Model	.18	S.P Sebenar	1.45	Pemisahan	8.19	Kebolehpercayaan Item		.99
S.E Min Item = .27								

Jadual 15 menunjukkan *Fungsi Item Terbeza (DIF)* melalui pemerhatian terhadap DIF *contrast* item bagi ujian ELA3312 dan ELA4322. DIF item yang melebihi julat ± 0.5 *logits* menunjukkan wujud kecenderungan item terhadap salah satu kumpulan dalam ujian yang ditadbir (Azrilah et al., 2017). Di dalam kajian ini, DIF item akan melihat bias item dari faktor jantina pelajar. Ujian ELA3312 menunjukkan 6 item yang melebihi julat DIF *contrast* ± 0.5 *logits* iaitu item9 (1.49 *logits*), item8 (1.09 *logits*), item2 (0.96 *logits*), item5 (0.96 *logits*), item7 (0.96 *logits*) dan item1 (-0.80 *logits*). Ujian ELA4322 juga menunjukkan 6 item yang terkeluar daripada julat kawalan DIF *contrast* iaitu pada item15 (0.92 *logits*), item12 (0.90 *logits*), item1 (0.75 *logits*), item11 (0.54 *logits*), item6 (-0.58 *logits*) dan item8 (-0.59 *logits*).

Jadual 15: DIF *Contrast* Item Ujian

Ujian	ELA3312	ELA4322	Jumlah Item
< 0.51	5	4	9
± 0.50	24	24	28
> -0.51	1	2	3
Jumlah Item	30	30	60

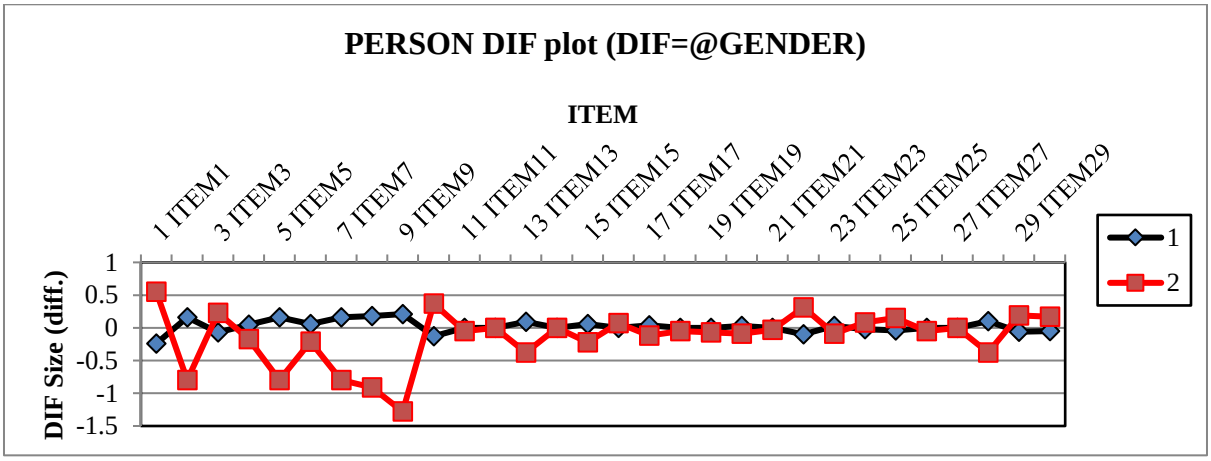
Selain daripada pemerhatian terhadap saiz DIF *contrast* , analisis saiz DIF dan DIF=*t* juga boleh diperhatikan. Item yang menjauhi julat saiz DIF > 0.50 *logits* serta DIF=*t* yang melebihi ± 2.0 *logits* perlu digugurkan. Namun sekiranya nilai DIF *t* item dalam julat kawalan ± 2.0 *logits,* item tersebut tidak perlu digugurkan kerana perbezaan bias adalah tidak signifikan (Azrilah et al., 2017). Rajah 1 menunjukkan saiz DIF bagi ujian ELA3312, paparan graf menunjukkan enam item terkeluar daripada julat saiz DIF dibenarkan iaitu item1, item2, item3, item5, item7 dan item9. Namun item tersebut tidak melebihi julat DIF *t*= ± 2.0 *logits* dan seharusnya boleh dikekalkan.

Berdasarkan Rajah 2, hanya tiga item daripada enam yang memiliki saiz DIF menjauhi julat saiz DIF > 0.50 logits iaitu item1, item12 dan item15 berbanding analisis DIF *contrast* yang mengeluarkan enam item daripada ujian ELA4322 tersebut. Pemerhatian lanjut terhadap nilai DIF *t* menunjukkan item12 sahaja yang melebihi DIF *t*= ± 2.0 *logits* iaitu pada nilai -2.38 logits. Item12 adalah bias terhadap salah satu jantina responden ujian dan seharusnya digugurkan. Item tersebut merupakan item mudah bagi kumpulan pelajar perempuan berbanding kepada kumpulan pelajar lelaki.

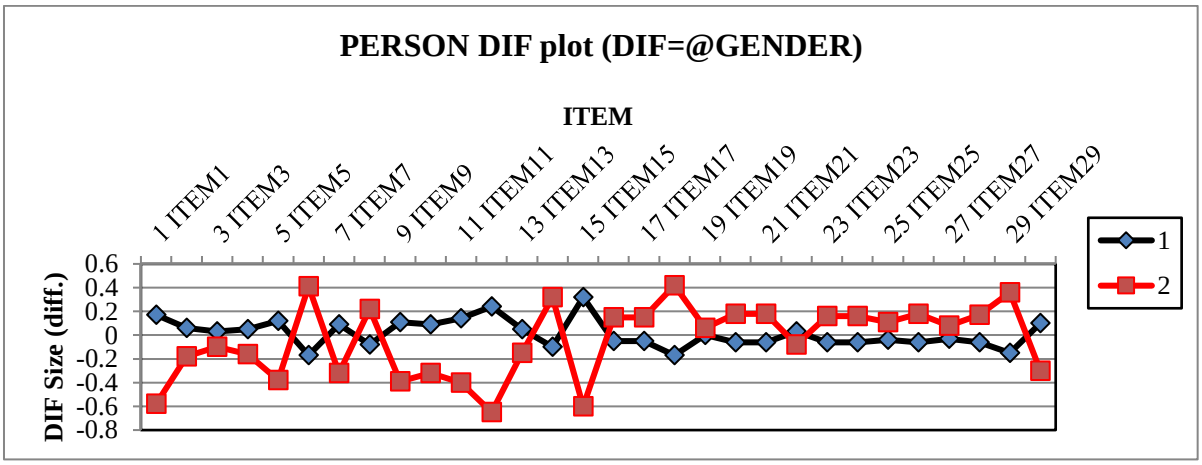
© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Malaysian Journal of Social Sciences and Humanities (MJSSH) (e-ISSN : 2504-8562)

Rajah 1 - Saiz DIF Ujian ELA3312



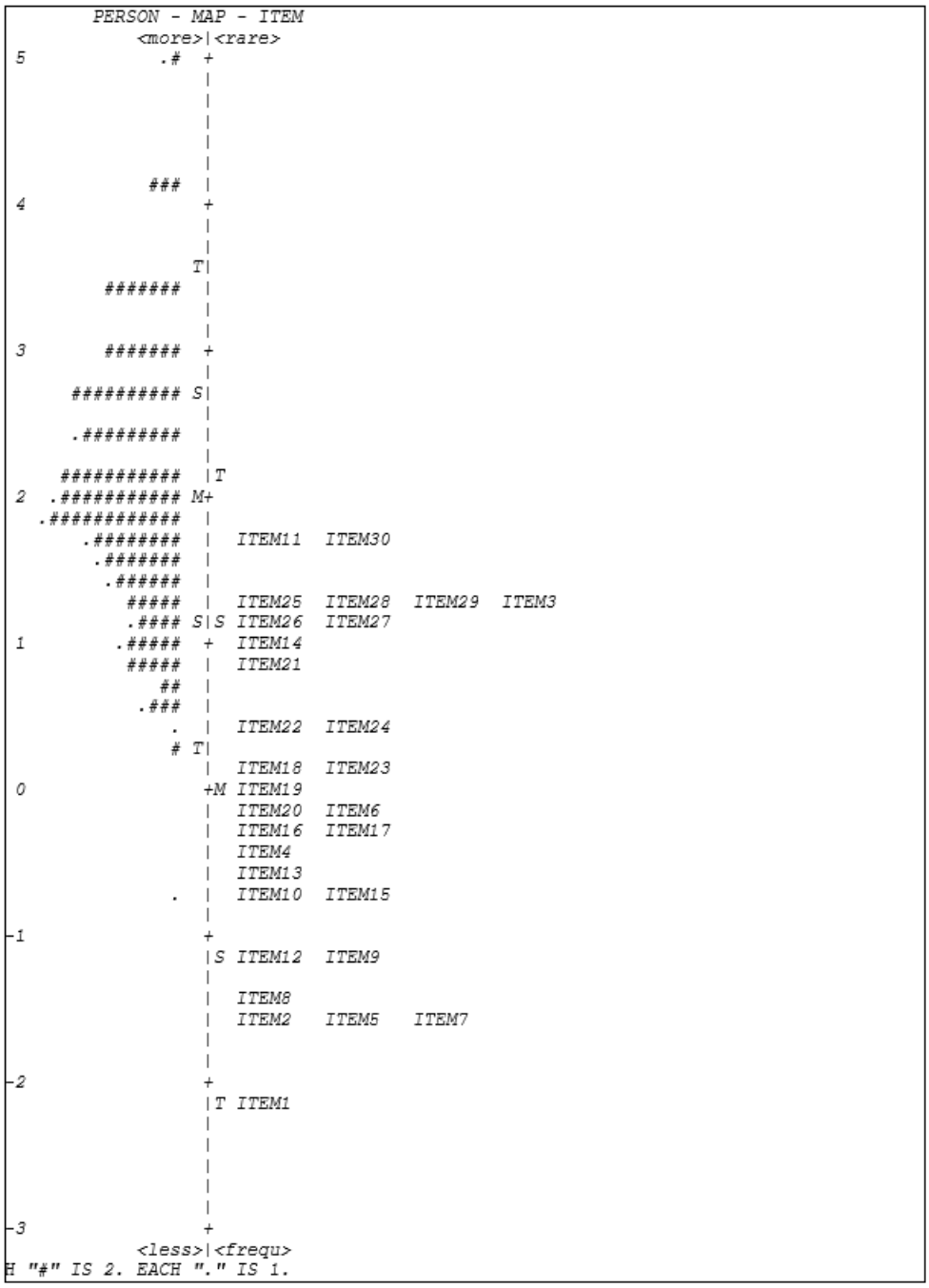
Rajah 2 - Saiz DIF Ujian ELA4322



Rajah 3 merujuk paparan taburan kebolehan pelajar di bahagian kiri peta serta kesukaran item di bahagian kanan peta. Urutan setiap item dan pelajar diletakkan di atas satu skala tunggal melalui garis ukuran sela *logits*. Melalui Rajah 3 tersebut, min kebolehan pelajar adalah lebih tinggi daripada min kesukaran item. Selang min kebolehan dan kesukaran item yang menghampiri dua logits ini menunjukkan item yang tadbir adalah mudah berbanding kebolehan pelajar. Item1 merupakan item termudah dan item11 dan item30 merupakan item paling sukar bagi ujian ELA3312. Daripada Rajah 3 juga, jurang signifikan antara item dapat dilihat. Jurang antara item berturutan melebihi .5 *logits* menunjukkan kesahan pengukuran konstruk tidak dapat dipenuhi oleh item tersebut (Hasni, 2014). Ujian ELA3312 menunjukkan dua jurang antara item yang melebihi .5 *logits* iaitu antara item21 (0.92 *logits*) dan item22 (0.37 *logits*) dengan jurang item =0.55 *logits* serta antara item2 (-1.59 *logits*) dan item1 (-2.21 *logits*) dengan jurang item =0.62 *logits.*

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Rajah 3: Peta Pelajar-Item Ujian ELA3312

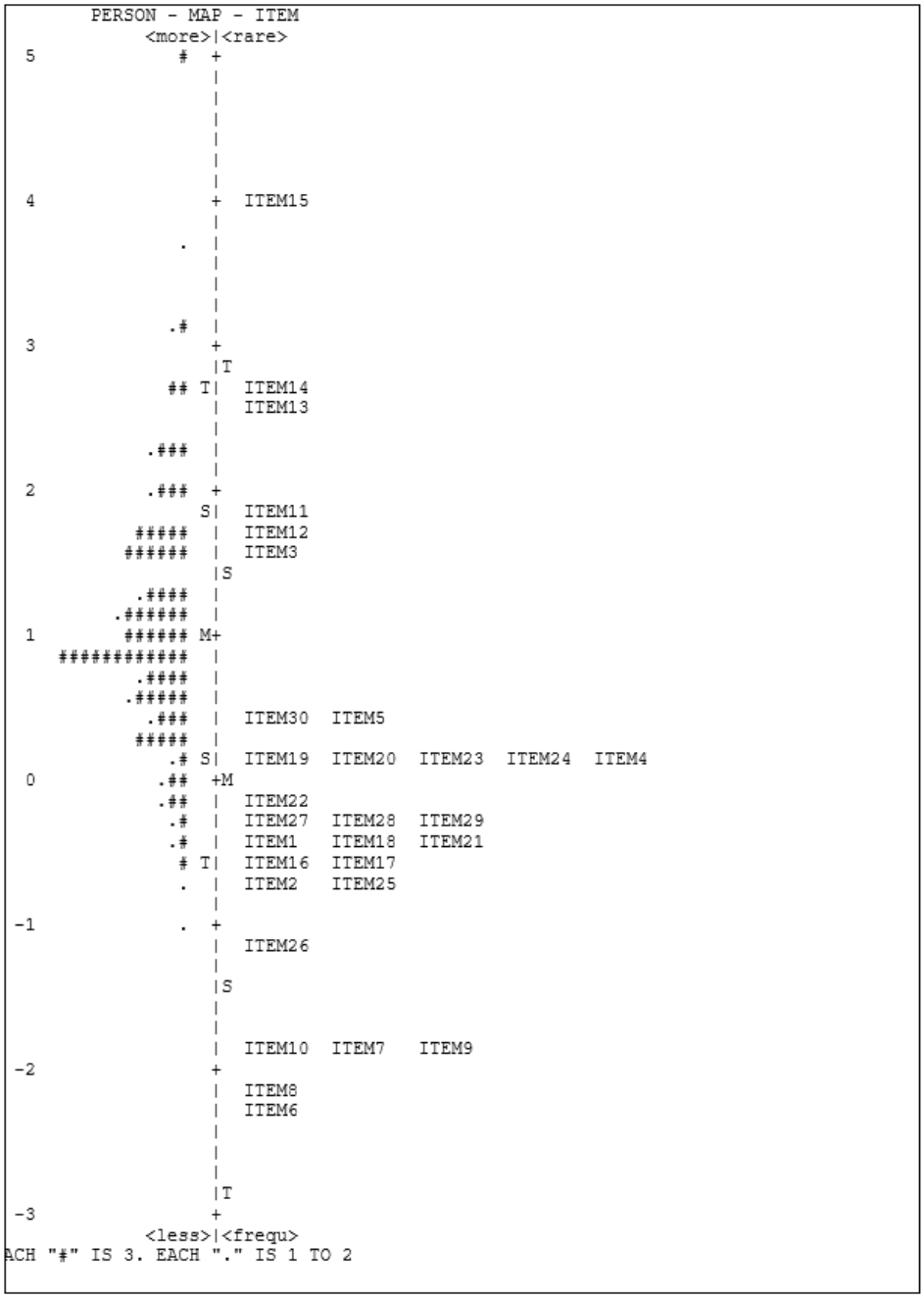


Rajah 4 merupakan paparan peta pelajar-item bagi ujian ELA4322. Nilai min kebolehan pelajar adalah lebih tinggi berbanding nilai min kesukaran item menunjukkan prestasi pelajar yang lebih baik berbanding item dengan item paling mudah adalah item6 dan item15 sebagai item paling sukar. Bagi jurang antara item, ujian ELA4322 memaparkan banyak jurang yang signifikan antara item berbanding ujian ELA3312. Empat jurang item melebihi .5 *logits* dapat dikesan iaitu antara item15 (4.05 *logits*) dan item14 (2.66 *logits*) dengan jurang item = 1.39 *logits*, item13 (2.54 *logits*) dan item11 (1.9 *logits*) dengan jurang item = 0.64 , item3 (1.52 *logits*) dan item30 (0.42 *logits*) dengan jurang item = 1.10 *logits* dan yang terakhir antara item26 (-1.20 *logits*) dan item10 (-1.89 *logits*) dengan jurang item = 0.69 *logits*. Menurut Boone (2016), jurang antara item tersebut dapat diminimumkan melalui pengguguran item atau penggantian kepada item yang

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

lebih baik. Tambahan item bagi memenuhi jurang antara item atau meminimumkan jurang adalah penting agar pengukuran terhadap kebolehan pelajar lebih tepat dapat dibuat (Samsudin et al., 2020a).

Rajah 4: Peta Pelajar-Item Ujian ELA4322



Jadual 16 merujuk taburan aras kesukaran item mengikut konstruk atau standard pembelajaran bahasa Arab Kolej Vokasional. Penetapan konstruk adalah berdasarkan andaian kesukaran, item huruf sepatutnya lebih mudah berbanding perkataan atau ayat pendek dan begitulah seterusnya. Konstruk ayat pendek yang sepatutnya adalah sukar menunjukkan dapatan yang bercanggah dengan andaian kesukaran apabila tiada item sukar konstruk ayat pendek pada ujian ELA4322. Ini menunjukkan pembinaan item bagi konstruk ayat pendek lebih menjurus item sederhana sahaja walaupun andaian awal sepatutnya item tersebut berfungsi sebagai item sukar. Penelitian pada konstruk perkataan adalah seimbang pada ujian ELA3312 namun tidak pada ujian ELA4322

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

apabila bilangan item adalah sama antara item sederhana dan item sukar. Perhatian seharusnya diberikan kepada konstruk huruf, kedua-dua ujian menunjukkan terdapat item sukar daripada konstruk huruf yang dijangka adalah mudah untuk dijawab betul bagi pelajar walaupun dari pelajar berkebolehan rendah. Pemerhatian lanjutan terhadap item boleh dilakukan bagi mengesan item dari sudut format sama ada memerlukan pengubahsuaian bagi menghasilkan item yang lebih baik.

Jadual 16: Taburan Kesukaran Item Mengikut Konstruk Pembelajaran

Ujian	**ELA3312**			**ELA4322**		
Konstruk/Kesukaran	**Mudah**	**Sederhana**	**Sukar**	**Mudah**	**Sederhana**	**Sukar**
Huruf	6	3	1	5	4	1
Perkataan	1	8	1	0	5	5
Ayat Pendek	0	4	6	0	10	0
Jumlah Item	7	15	8	5	19	6
			30			30

Berdasarkan analisis yang dijalankan terhadap kedua-dua ujian bahasa Arab. Dapatan menunjukkan bahawa item yang ditadbir adalah bersesuaian kecuali pada beberapa item yang tidak memenuhi andaian pengukuran seperti item12 daripada ujian ELA4322 yang boleh digugurkan atau diubahsuai semula kerana faktor bias item terhadap kumpulan jantina pelajar. Jadual 17 berikut merupakan rumusan statistik bagi kedua-dua ujian yang menunjukkan pembuktian empirikal yang baik bagi unidimensi ujian, kesesuaian item melalau julat *infit/oufit MNSQ* item yang sesuai serta faktor residual tidak tertakrif yang rendah. Indeks pemisahan dan kebolehpercayaan item juga menunjukkan kemampuan item untuk berfungsi secara tekal walaupun ditadbir kepada sampel yang lain.

Jadual 17: Rumusan Statistik Ujian Bahasa Arab Kolej Vokasional

Statistik	**ELA3312**	**ELA4322**
Andaian Model Rasch		
Min *infit MNSQ*	1.00	1.00
Min *outfit MNSQ*	.93	.96
Faktor kedua PCA	2.6	3.0
Min kesukaran item	0.00 *logits*	0.00 *logits*
Sisihan piawai kesukaran item	1.10 *logits*	1.46 *logits*
Min kebolehan pelajar	1.99 *logits*	1.08 *logits*
Sisihan piawai kebolehan pelajar	0.89 *logits*	0.99 *logits*
Indeks kebolehpercayaan item	.96	.98
Indeks kebolehpercayaan pelajar	.65	.75
Indeks pemisahan item	4.89	8.08
Indeks pemisahan pelajar	1.36	1.54

5. Kesimpulan

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

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

pengukuran Rasch menyediakan analisis psikometrik item yang baik dalam menentukan kualiti setiap item.

Dapatan akhir menunjukkan bahawa sebahagian besar item daripada dua ujian penilaian akhir bahasa Arab Kolej Vokasional adalah memenuhi andaian pengukuran Rasch. Indeks pemisahan dan kebolehpercayaan item yang tinggi membuktikan item dapat dibahagikan kepada hierarki kesukaran item yang pelbagai. Item-item yang telah ditentu ukur boleh dilibatkan semula sebagai item ujian yang baik atau digunakan dalam kajian lanjutan sekiranya diperlukan. Pemerhatian akhir kepada taburan kesukaran item mengikut konstruk pada Jadual 16 juga menunjukkan nisbah taburan yang sesuai berdasarkan bilangan item mengikut aras kesukaran (mudah-sederhana-sukar).

Pengkaji merasakan keperluan kajian lanjutan dijalankan terhadap item-item ujian bahasa Arab Kolej Vokasional secara berterusan terhadap set-set ujian daripada semester lalu dan akan datang bagi mengetahui kualiti item yang baik untuk disimpan sebagai bank item bahasa Arab yang berkualiti tinggi untuk dilibatkan semula sebagai item ujian yang baharu. Tambahan lagi, pertambahan tentu ukuran item yang lebih banyak dijangka mampu merapatkan jurang kesukaran antara item seterusnya memenuhi andaian kesahan konstruk selaras dengan keperluan pengukuran model Rasch yang lebih baik. 30 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.

Kelulusan Etika dan Persetujuan untuk Menyertai Kajian (*Ethics Approval and Consent to Participate*)

Penyelidik menggunakan garis panduan etika penyelidikan daripada Jawatankuasa Etika Penyelidikan USM (JEPeM). Semua prosedur yang melibatkan sampel manusia telah dijalankan mengikut piawaian jawatankuasa JEPeM. Kebenaran dan persetujuan mengikuti kajian turut diperolehi daripada peserta kajian. Kajian ini juga adalah berdasarkan kelulusan Bahagian Perancangan dan Penyelidikan Dasar Pendidikan, Kementerian Pendidikan Malaysia (eRAS 2.0) serta Sektor Perancangan, Penyelidikan dan Pembangunan TVET, Bahagian Pendidikan dan Latihan Teknikal Vokasional (BPLTV).

Penghargaan

Penghargaan kepada semua pihak yang terlibat menjayakan penyelidikan, penulisan dan penerbitan kajian ini.

Kewangan

Penyelidikan, penulisan dan penerbitan kajian ini adalah tidak melibatkan penajaan daripada pihak lain. Sebarang kos berkaitan adalah ditanggung oleh penulis sendiri.

Konflik Kepentingan

Tiada sebarang konflik kepentingan berkaitan penyelidikan, penulisan dan penerbitan kajian ini.

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Rujukan

Agensi Kelayakan Malaysia. (2014). Garis Panduan Amalan baik : Penilaian Pelajar. Agensi Kelayakan Malaysia. *Agensi Kelayakan Malaysia*. www.mqa.gov.my

Akta Pendidikan, (1996).

Azrilah Abdul Aziz, Mohd Saidfudin Masodi, Azami Zaharim. (2017). *Asas Model Pengukuran: Pembentukan Skala & Struktur Pengukuran (Ketiga).* Penerbit UKM.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences: Third edition*. Routledge Taylor & Francis. https://doi.org/10.4324/9781410614575

Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch Model Fundamental Measurement in the Human Sciences (Fourth Edi)*. Routledge.

Boone, W. J. (2016). Rasch analysis for instrument development: Why,when,and how? *CBE Life Sciences Education, 15*(4), 1–7. https://doi.org/10.1187/cbe.16-04-0148

Boone, W. J. (2020). Rasch Basics for the Novice. In M. S. Khine (Ed.), *Rasch Measurement Applications in Quantitative Educational Research* (pp. 9–30). Springer Nature Singapore Pte Ltd. https://doi.org/https://doi.org/10.1007/978-981-15-1800-3

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer Science+Business Media. https://doi.org/10.1007/978-94-007-6857-4

Damit, M. A. A., & Omar, M. K. (2018). Pelaksanaan Outcome-Based Education di Kolej Vokasional, Malaysia. *Graduate Research In Education Seminar* (GREduc) 2018, December 2018, 138–146.

Isnani, I., Utami, W. B., Susongko, P., & Lestiani, H. T. (2019). Estimation of college students' ability on real analysis course using Rasch model. *Research and Evaluation in Education, 5*(2), 95–102. https://doi.org/10.21831/reid.v5i2.20924

Kolen, M. J., & Brennan, R. L. (2014). *Test Equating, Scaling, and Linking*. Springer Science+Business Media.

KPM. (2019). Maklumat umum TVET. *KPM*. https://www.moe.gov.my/en/maklumat-umum-tvet

Linacre, J. M. (2012). Winsteps® Rasch Measurement Computer Program User's Guide (Version 3.74.0). *Winsteps*. https://www.winsteps.com/

Mahamud, S., & Othman, N. (2019). Item Analysis of Technology Challenges Among Adolescent Using Rasch Model. *International Journal of Academic Research in Business and Social Sciences, 9*(6). https://doi.org/10.6007/ijarbss/v9-i6/6229

Marina Mukhtar, & Jamil Ahmad. (2015). Assessment for Learning: Practice in TVET. *Procedia - Social and Behavioral Sciences*, 204, 119–126.

Sbeglia, G. C., & Nehm, R. H. (2018). Measuring evolution acceptance using the GAENE: influences of gender, race, degree-plan, and instruction. *Evolution: Education and Outreach*, *11*(1), 18. https://doi.org/10.1186/s12052-018-0091-9

Samsudin, M. A., Chut, T. S., Ismail, M. E., & Ahmad, N. J. (2020a). A Calibrated Item Bank for Computerized Adaptive Testing in Measuring Science TIMSS Performance. *Eurasia Journal of Mathematics, Science and Technology Education, 16*(7), em1863. https://doi.org/10.29333/ejmste/8259

Samsudin, M. A., Norfarah, Nordin & Chut, Thodsaphorn. (2020b). Psychometric Assessment of Mathematics Algebra Item Banks for Computerized Adaptive Test. *Technology Reports of Kansai University*, 62. 2061-2076.

Hasni Shamsuddin (2014). *Pengkalibrasian Item Untuk Penggabungan Dua Ujian Matematik Tingkatan 2.* [Tesis sarjana yang tidak diterbitkan]. Universiti Sains Malaysia.

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Hasni Shamsuddin (2020). *Pembinaan Bank Item Matematik Tingkatan 1 Melalui Prosedur Penggabungan Ujian Menggunakan Model Rasch.* [Tesis PhD yang tidak diterbitkan]. Universiti Sains Malaysia.

Sick, J. (2010). Rasch Measurement in Language Education Part 5 : Assumptions and requirements of Rasch measurement. *Shiken: JALT Testing & Evaluation SIG Newsletter, 14*(October), 23–29.

Sumintono, B. (2018). Rasch Model Measurements as Tools in Assesment for Learning. *Proceedings of the 1st International Conference on Education Innovation* (ICEI 2017) https://doi.org/10.2991/icei-17.2018.11

Yasin, R. M., Yunus, F. A. N., Rus, R. C., Ahmad, A., & Rahim, M. B. (2015). Validity and Reliability Learning Transfer Item Using Rasch Measurement Model. *Procedia - Social and Behavioral Sciences, 204* (November 2014), 212–217. https://doi.org/10.1016/j.sbspro.2015.08.143

Yu, C. H. (2020). Objective Measurement: How Rasch Modeling Can Simplify and Enhance Your Assessment. In M. S. Khine (Ed.), *Rasch Measurement Aplication in Quantitative Educational Research* (pp. 47–73). Springer Nature Singapore Pte Ltd. https://doi.org/https://doi.org/10.1007/978-981-15-1800-3_4

© 2022 by the authors. Published by Secholian Publication. This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).