

Kesahan dan Kebolehpercayaan Ujian Kemahiran Proses Sains untuk Murid Sekolah Rendah Berdasarkan Model Pengukuran Rasch

(Validity and Reliability of Science Process Skills Test for Primary School Pupils Based on Rasch Measurement Model)

ELLYZA KARIM* & KAMISAH OSMAN

ABSTRAK

Kemahiran proses sains merangkumi kemahiran-kemahiran yang dilakukan saintis semasa mereka mengkaji dan menyiasat. Dalam tugas di bilik sains, murid boleh mengaplikasi kemahiran berkenaan untuk menyerupai langkah-langkah prosedur yang sistematik sebagaimana saintis uruskan. Bagi mengkaji tahap kemahiran proses sains, sebuah instrumen berskala Likert lima mata yang mengandungi 68 item dibina. Kajian menggunakan instrumen ini dijalankan secara kuantitatif berdasarkan reka bentuk tinjauan. Tujuan kajian ialah untuk menentukan sejauh mana kesahan dan kebolehpercayaan instrumen ini dapat memenuhi ciri psikometrik sesebuah penyelidikan. Analisis item membolehkan item yang sah, adil dan berkualiti terpilih manakala item sebaliknya ditambah baik atau dibuang. Kesahan item dalam kajian ini dikaji dalam aspek keserasian item, polariti item, kebebasan setempat dan analisis unidimensi menggunakan analisis model pengukuran Rasch. Kebolehpercayaan yang terlibat dalam kajian ini turut dianalisis berdasarkan model pengukuran Rasch daripada aspek indeks pengasingan serta indeks kebolehpercayaan item dan individu. Data dikaji menggunakan perisian Statistik IBM SPSS dan WINSTEPS. Pembinaan instrumen ini menyumbang kepada pertambahan kepelbagaian bentuk ujian terutama jenis pentaksiran masa kini yang mula cenderung ke arah bentuk penilaian sendiri. Instrumen-instrumen sebelum ini lebih mengkaji kepada aspek pengetahuan namun instrumen ini menumpukan kepada aspek pengetahuan dan kebolehan. Implikasi dapatan kajian ini memberi kelebihan kepada guru sains menilai tahap kemahiran proses sains murid berkaitan.

Kata kunci: Model pengukuran Rasch, kemahiran proses sains, kesahan, kebolehpercayaan, penilaian sendiri

ABSTRACT

Science process skills include the skills that scientists apply when they study and investigate. During scientific investigation, pupils may carry out those skills to resemble systematic procedural steps as the scientists organize. To study the level of science process skills, a five-point Likert scaled instrument containing 68 items was developed. The study was done quantitatively based on survey design. The purpose of the study was to determine the extent of the validity and reliability of this instrument to fulfill the psychometric characteristics of a research. Item analysis allows valid, fair and quality items to be selected while items otherwise enhanced or removed. The validity of the items in this study were analysed in terms of item fit, item polarity, local dependence and unidimensionality using the Rasch measurement model analysis. The reliability involved in this study was analysed based on Rasch measurement model from the aspect of item and individual separation index and reliability index. The data were processed using IBM SPSS Statistics and WINSTEPS. The construction of this instrument contributes to the increase in the variety of test forms, especially the types of current assessment that tend to be in the form of self-evaluation. The previous instruments studied the knowledge aspect but this instrument focused on the aspects of knowledge and ability. The implication of this study gives advantage to science teachers to evaluate the level of their student's science process skills.

Keywords: Rasch measurement model, science process skills, validity, reliability, self evaluation

PENGENALAN

Pengujian merupakan satu kaedah mengukur pengetahuan seseorang. Dalam konteks pengajaran dan pembelajaran sains, salah satu cara untuk menguji pengetahuan sains ialah melalui pengujian. Seiring dengan peredaran masa, pelbagai jenis ujian telah dibangunkan, tidak terhad sekadar menguji pengetahuan kemahiran proses sains

(KPS). Antaranya ialah ujian untuk menguji kebolehan KPS. Namun, bilangan murid yang ramai dalam satu kelas dan masa pengajaran dan pembelajaran sains yang terhad menimbulkan bebanan untuk guru menguji kebolehan setiap murid. Surat pekeliling daripada Majlis Peperiksaan Malaysia bertarikh 10 Januari 2011 yang mewajibkan semua guru sains untuk melaksanakan penilaian kerja amali daripada bulan Februari sehingga Julai pada tahun

tersebut, dan seterusnya memasukkan semua markah ke dalam pangkalan data peperiksaan pada bulan Ogos turut menambah beban aspek kekangan masa (Ong & Bibi Hazliana 2014). Maka, penyelesaian yang wajar ialah mewujudkan ujian penilaian sendiri. Ujian jenis ini mempunyai kekuatan dengan tidak membebaskan guru dalam hal pemerhatian. Guru cuma perlu memasukkan data maklum balas murid dan perisian membantu guru menganalisis data untuk menunjukkan tahap kebolehan tersebut. Guru kemudiannya membuat interpretasi terhadap analisis data. Jika ujian dibuat secara dalam talian, penilaian sendiri akan lebih memudahkan tugas guru setakat menganalisis dan menginterpretasi sahaja. Menurut Ross (2006), penilaian sendiri telah mempelbagaikan bentuk ujian maka ia dapat menyumbang ke arah peningkatan mutu penyelidikan dalam pendidikan sains mahupun dalam bidang psikometrik.

Oleh kerana penilaian sendiri melibatkan persepsi murid ke atas diri sendiri maka penghasilan instrumen untuk menguji kebolehan harus dibina secara terperinci. Ciri psikometrik ujian perlu diteliti untuk menghasilkan sebuah instrumen yang mempunyai kesahan dan kebolehpercayaan lebih-lebih lagi ke atas instrumen yang melibatkan penilaian sendiri. Teknik terkini dalam teori respons item seperti pengukuran berasaskan model pengukuran Rasch mempunyai kemampuan tersendiri dalam menganalisis kesahan dan kebolehpercayaan item malah sebagai nilai tambah model ini dapat menganalisis kesahan dan kebolehpercayaan murid. Lebih menarik lagi model pengukuran Rasch mampu menyusun item dan murid kepada aras kesukaran dan aras kebolehan mereka. Instrumen-instrumen sebelum ini dilihat lebih tertumpu kepada analisis kesukaran item dan indeks diskriminasi sahaja kerana bersifat menguji pengetahuan. Dalam artikel ini, model pengukuran Rasch diaplikasi untuk mengukur kesahan item dalam aspek keserasian item, polariti item, kebebasan setempat dan analisis unidimensi dalam menyokong isu psikometrik pembinaan instrumen. Selain itu, bukan sekadar aspek kebolehpercayaan item, malah kebolehpercayaan individu turut boleh dianalisis menggunakan model pengukuran Rasch bersama nilai tambah lain iaitu aspek pengasingan item dan individu.

Objektif artikel ini ialah untuk menentukan kesahan item dan kebolehpercayaan item-individu untuk instrumen kemahiran proses sains, KePSa. Kesahan item dikaji dalam aspek keserasian item, polariti item, kebebasan setempat dan analisis unidimensi. Kebolehpercayaan pula bukan setakat dikaji ke atas item sahaja, malah melibatkan individu juga. Selain itu, indeks pengasingan item-individu turut dikaji untuk mengenal pasti hierarki kesukaran item manakala indeks pengasingan individu dapat membezakan antara individu yang berprestasi tinggi dan rendah. Model pengukuran Rasch turut menganalisis nilai ketekalan dalaman berdasarkan nilai alfa Cronbach. Terdapat 68 item untuk konstruk kemahiran proses sains (KPS) secara keseluruhan namun terdapat dua subkonstruk KPS iaitu KPS asas dan KPS bersepadu. KPS bersepadu merupakan gabungan beberapa KPS asas maka KPS bersepadu adalah

lebih kompleks daripada KPS asas. Domain kemahiran yang terkandung dalam KPS asas ialah memerhati, mengelas, mengukur dan menggunakan nombor, membuat inferens, meramal, berkomunikasi dan menggunakan perhubungan ruang dan masa. Domain KPS bersepadu pula terdiri daripada kemahiran mentafsir data, membuat definisi secara operasi, mengawal pemboleh ubah, membuat hipotesis dan mengeksperimen. Sehubungan dengan itu, KPS asas perlu dikuasai terlebih dahulu sebelum beralih ke KPS bersepadu (Punia et al. 2012). Daripada jumlah 68 item, bilangan item KPS asas ialah sebanyak 42 manakala KPS bersepadu mempunyai 26 item.

KPS ialah amalan yang menggalakkan pembentukan pelbagai kemahiran untuk memperolehi pengetahuan yang kemudiannya disebar untuk meningkatkan penggunaan kemahiran mental dan psikomotor seoptimum mungkin (Punia et al. 2012). Edy Hafizan dan Lilia (2010), Fatin Aliah dan Nor Athirah (2011), Ong et al. (2012), Kamisah dan Rian (2013), Ong dan Mohd Al-Junaidi (2013) serta Ong et al. (2015) telah membina instrumen untuk pentaksiran KPS. Namun instrumen tersebut menggunakan item berbentuk soalan aneka pilihan bertujuan menguji pengetahuan fakta sains murid-murid seperti dalam peperiksaan umum. Analisis skor instrumen itu menjadi alat pengukur pencapaian murid dalam bentuk markah, gred atau peratus. Berbeza dengan instrumen KePSa, perkara yang diukur ialah kebolehan. Kebolehan murid diukur melalui skala Likert lima mata dengan label kurang boleh, agak boleh, sederhana boleh, boleh dan sangat boleh. Analisis min logit berdasarkan persentil digunakan dalam penentuan aras kebolehan KPS murid di mana majoriti murid lepasan UPSR kohort 2017 didapati berada pada aras sederhana boleh. Instrumen KePSa merupakan ujian rujukan norma kerana min yang diperoleh dibandingkan dengan min kumpulan norma yang dirujuk berdasarkan persentil. Ini berbeza dengan ujian pengetahuan kerana ia merupakan ujian rujukan kriteria di mana skor yang diperoleh dibandingkan dengan kriteria tertentu seperti penguasaan konsep sebelum bergerak kepada konsep lain yang lebih mendalam.

Analisis kuantitatif dijadikan sebagai bukti empirikal untuk kesahan item dan kebolehpercayaan item-individu bagi menentukan tahap kebolehan KPS murid-murid ini. Nilai keserasian item menganggarkan darjah keserasian item yang mengukur suatu pemboleh ubah terpendam. Nilai keserasian item iaitu nilai *infit* dan *outfit* purata kuasa dua (MNSQ) berada dalam julat penetapan bernilai 0.5 hingga 1.5 logit (Boone et al. 2014; Bambang & Wahyu 2014). Statistik keserasian item boleh menunjukkan sejauh mana ketepatan atau kebolehamalan data berpadanan dengan model. Ini bermakna, pengkaji boleh mengetahui fenomena meneka jawapan dalam kalangan responden jika berlaku. *Infit* merupakan nilai kepekaan corak jawapan item yang disasarkan ke atas individu manakala *outfit* adalah nilai kepekaan terhadap kesukaran item. Polariti item pula dapat menunjukkan sejauh manakah item-item mengukur konstruk yang sama dan semua item mengukur

subkonstruk tunggal (Bond & Fox 2015). Analisis ini juga menunjukkan terdapatnya item atau individu yang memberi maklum balas bercanggah dengan pemboleh ubah (Linacre 2005). Seterusnya, analisis kebebasan setempat menonjolkan sejauh manakah respons kebolehan individu kepada mana-mana item adalah tidak berkait dengan respons item lain dalam konstruk yang sama. Akhir sekali, analisis unidimensi berupaya menentukan item-item dalam instrumen mengukur hanya satu keupayaan tunggal (Wright & Master 2002).

Item dengan indeks kebolehpercayaan yang tinggi menunjukkan item-item ini mempunyai ketekalan apabila ditadbir berulang kali kepada kumpulan responden lain tetapi mempunyai kebolehan yang hampir sama. Kebolehpercayaan item yang tinggi juga menunjukkan kecukupan item untuk mengukur apa yang sepatutnya diukur (Azrilah et al. 2013). Kebolehpercayaan individu yang tinggi pula menunjukkan bahawa jawapan murid adalah konsisten (Bambang & Wahyu 2015). Indeks pengasingan item pula menjelaskan bahawa melebihi nilai dua menunjukkan bilangan sampel individu cukup besar untuk mengesahkan hierarki kesukaran item instrumen (Jones & Fox 1998). Indeks pengasingan individu pula menunjukkan bahawa instrumen cukup sensitif untuk membezakan antara individu yang berprestasi tinggi dan rendah (Bambang & Wahyu 2014).

Analisis kesahan item merupakan kesahan konstruk, namun aspek kesahan muka, kesahan ramalan dan kesahan serentak turut dikaji dalam kajian sebenar bagi pembinaan instrumen KePSa. Walau bagaimanapun, artikel ini hanya membincangkan tentang kesahan item dan kebolehpercayaan item dan individu. Ini kerana isu tentang kesahan dan kebolehpercayaan item amat penting maka harus diberi penekanan dalam pembinaan instrumen. Justeru, instrumen atau alat ukur yang dibina itu perlulah mempunyai ciri-ciri psikometrik yang berkualiti supaya interpretasi skor yang diberikan tepat dan betul. Dalam hal ini, model pengukuran Rasch sebagai salah satu teknik dalam Teori Respons Item (TRI) dapat menyokong Teori Pengukuran Klasikal (TPK) dalam penentuan analisis ciri psikometrik sesebuah instrumen.

TINJAUAN LITERATUR

Pelopop teori pembelajaran, Gagne (1965) melabelkan sains sebagai suatu proses. Prasyarat untuk mendapatkan konsep dan prinsip saintifik adalah melalui operasi proses sains. Kesemua proses ini adalah kemahiran yang digunakan oleh saintis, dipraktikkan dalam penyiasatan sains, boleh dipelajari oleh murid dan dipindahkan merentas domain isi kandungan sains. Memerhati, mengelas, mengukur dan menggunakan nombor, meramal, membuat inferens, menggunakan perhubungan ruang dan masa, mentafsirkan data, mendefinisikan secara operasi, mengawal pemboleh ubah, membuat hipotesis, berkomunikasi dan mengeksperimen adalah dua belas domain kemahiran

proses sains yang diamalkan dalam konteks pendidikan sains di Malaysia (BPK 2015). Penguasaan KPS yang baik dalam kalangan murid membolehkan mereka berfikir secara kritis dan proaktif untuk menyelesaikan sesuatu masalah dalam kehidupan harian dengan berkesan (Abu Hassan & Rohana 2003). Kajian Freitag & Pfeffer (2013) telah membuktikan bahawa masyarakat berliterasi sains dapat dibentuk menerusi penekanan proses sains. Maka, amat penting KPS dimulakan dari alam persekolahan bagi melahirkan generasi berliterasi sains.

Ross (2006) menekankan bahawa murid berhak mentaksir hasil kerja mereka kerana hanya mereka yang tahu sebanyak mana usaha dan masa yang telah mereka habiskan dalam menyiapkan tugas. Maka penilaian sendiri merupakan satu alat ukur yang wajar kerana murid sendiri yang mendapat pengalaman sepanjang pembelajaran itu. Secara tidak langsung penilaian sendiri telah membentuk generasi berfikiran kritis sebagaimana definisi kritis oleh Dewan Bahasa dan Pustaka (2016) iaitu tidak dengan begitu sahaja menerima atau mempersetujui sesuatu malahan menimbangakan buruk dan baiknya terlebih dahulu. Dalam konteks KPS, amalan saintifik yang diamalkan oleh murid dalam kelas sains berasal daripada amalan saintis semasa menyelesaikan masalah. Silibus sains menekankan KPS kerana dalam kemahiran saintifik, kesemua amalan KPS dialami oleh murid sendiri semasa penyiasatan sains maka mereka layak untuk menilai secara sendiri sejauh mana amalan tersebut berjaya dikuasai mereka.

Pengajaran dan pembelajaran hendaklah mengutamakan kaedah yang berpusatkan pelajar yang memberi peluang semaksimum mungkin untuk mereka belajar secara sendiri dan secara inkuiri (Siti Rahayah et al. 2010). Penilaian sendiri mempunyai banyak kelebihan seperti mana yang disenaraikan dalam tinjauan literatur pada tesis Adibah (2013). Antaranya penilaian sendiri membolehkan pelajar secara individu mengukur dan membuat refleksi sendiri dalam penguasaan sesuatu kemahiran, memberi keyakinan kepada pelajar tentang peranannya dalam menilai perkembangan diri dan pencapaiannya, mengurangkan kebergantungan pelajar kepada guru dan membolehkan pelajar bersikap kritis terhadap kerjanya.

Kajian ini penting untuk memberi pendedahan kepada pengkaji atau pembina instrumen agar dapat memberi lebih perhatian mengenai aspek kesahan dan kebolehpercayaan sesuatu instrumen sebelum melaksanakan pengukuran. Kajian ini juga turut menjelaskan kepada umum bahawa analisis item dengan menggunakan TRI berdasarkan model pengukuran Rasch dapat memperlihatkan kebolehpercayaan dan kesahan item yang lebih terperinci. Model pengukuran ini juga dapat melengkapi TPK dalam mengesahkan sesuatu instrumen agar ianya sah dan boleh dipercayai untuk diguna pakai dalam pengukuran. Kebiasaannya instrumen KPS yang menguji pengetahuan menggunakan analisis indeks fasiliti, indeks diskriminasi dan analisis pengganggu dalam lingkungan TPK. Indeks jenis ini mempunyai kelemahan iaitu nilai-nilainya bergantung kepada populasi (Lim & Ong

2009). Jurang yang dapat dilengkapi dalam pembinaan ujian KePSa ialah dengan menggunakan analisis TRI berdasarkan model pengukuran Rasch dengan penentuan aspek keserasian, polariti, kebebasan setempat dan unidimensi.

Pengesahan instrumen dilakukan menerusi lima tahap iaitu tahap semak pakar, ubah suai atau baiki item, kajian rintis, analisis kesahan dan pekali kebolehppercayaan sebagaimana yang dilaksanakan Abdul et al. (2010). Semakan pakar dianalisis secara kuantitatif menerusi kaedah *Fuzzy Delphi*. Kemudian item diubah suai dan dibaiki selepas mendapat komen dan cadangan para pakar seramai tiga belas orang. Kajian rintis untuk instrumen KePSa telah dijalankan ke atas 300 orang sampel di tiga buah negeri dalam Malaysia yang dipilih secara rawak mudah. Seterusnya analisis kesahan dan kebolehppercayaan dijalankan berdasarkan model pengukuran Rasch. Pemurnian item dilakukan buat kali kedua sebelum kajian sebenar dilaksanakan ke atas 514 orang sampel lain dari seluruh Malaysia.

Model pengukuran Rasch mempunyai kelebihan kerana menepati keperluan pengukuran yang mana ia mampu mengubah skor mentah yang dipengaruhi oleh item dan ciri-ciri sampel kepada ukuran linear yang mempunyai interval yang sama (Wright & Stone 2004). Ini bererti, model pengukuran Rasch boleh menukar skala ordinal ke bentuk skala interval berdasarkan nilai logit. Skala interval lebih baik daripada skala ordinal kerana sela antara nombor-nombor adalah sama. Bagi skala interval, saiz atau lebar sela di antara angka-angka adalah diketahui dan perbezaan di antara nombor-nombor turutan adalah sama sepanjang skala ini. Berbeza dengan skala ordinal, sela di antara angka-angka itu tidak semestinya sama (Siti Rahayah 2008).

Menurut Bambang dan Wahyu (2015), skor data mentah bukanlah hasil pengukuran kerana ia merupakan maklum balas yang ditanda responden daripada instrumen yang ditadbir. Skor mentah sekadar maklumat awal yang tidak lain hanyalah ringkasan data berupa angka tetapi tidak memberikan data dari sudut pengukuran. Maka ia tidak bersifat linear. Contohnya jika seorang murid mendapat markah 80/100 tetapi menjawab dengan betul majoriti soalan beraras rendah iaitu aras mengingat dan memahami, manakah lebih baik jika dibandingkan dengan murid yang mendapat markah 70/100 tetapi betul banyak soalan pada aras tinggi iaitu menganalisis dan menilai? Oleh kerana itu, pendekatan selain penggunaan analisis skor mentah sangat diperlukan dalam konteks penilaian pendidikan. Selain itu, hal ini juga diperlukan bagi mendapatkan maklumat tambahan iaitu dari segi kebolehan yang dimiliki responden dan pada masa yang sama untuk menentukan kualiti item yang diberi. Maka, dengan mengaplikasi model pengukuran Rasch, data mentah hasil ujian ditukar ke bentuk linear berdasarkan nilai logit dengan mempunyai interval yang sama. Selepas proses ini, data siap untuk ditafsir kerana telah memiliki maklumat yang wajar terhadap kebolehan responden dan kualiti item.

Pelbagai kaedah boleh digunakan dalam analisis item. Analisis item menggunakan model pengukuran Rasch

menyenaraikan beberapa langkah penentuan psikometrik untuk mengesahkan kualiti item yang dikaji. Antaranya keserasian (*fit*) item yang ditentukan berdasarkan nilai *infit* MNSQ dan *outfit* MNSQ. MNSQ (mean square) adalah statistik kesesuaian min kuasa dua. Statistik keserasian item menunjukkan sejauh mana sesuatu data adalah berpadanan dengan model pengukuran Rasch. Item yang tidak *fit* menunjukkan ia tidak mengukur konstruk yang serupa dan merupakan item yang lemah (Siti Rahayah 2008). Menurut Bond & Fox (2015) nilai *outfit* item yang rendah mencadangkan respons sampel sangat boleh diramalkan apabila kebolehan sampel tersebut jauh daripada item. Nilai *infit* item yang rendah pula mencadangkan respons item sangat mudah diramal apabila kebolehan sampel dekat dengan item. Manakala, nilai *outfit* item yang tinggi mencadangkan respons luar biasa apabila kebolehan sampel jauh daripada item. Nilai *infit* item yang tinggi pula mencadangkan terdapat respons luar biasa apabila kebolehan sampel mendekati item.

Langkah seterusnya dalam penentuan kesahan item ialah analisis polariti item atau keselarian item. Ukuran keselarian bernilai positif menunjukkan semua item dalam instrumen itu berfungsi ke arah selari untuk mengukur konstruk dalam instrumen berkenaan. Sekiranya terdapat indeks yang bernilai negatif maka, pengkaji perlu menyemak semula data sama ada perlu diubah suai atau digugurkan. Indikator negatif menunjukkan terdapat item yang bercanggah dengan pemboleh ubah yang dikaji (Linacre 2012). Korelasi titik pusat ukuran (PTMEA) adalah juga merupakan pengesanan awal kepada kesahan konstruk (Bond & Fox 2015).

Item yang dianalisis oleh model pengukuran Rasch tidak boleh bersandar antara satu sama lain. Ini bermakna, maklum balas boleh atau kurang boleh pada satu item membawa maksud yang bermakna dalam aspek kebolehan iaitu *trait* terpendam yang dikaji, bukan sebaliknya. Sehubungan dengan itu, tidak boleh ada korelasi antara dua item. Jika terdapat korelasi maka wujud ketidaktepatan pada kualiti item. Item yang bersandar boleh mengganggu kebolehppercayaan instrumen. Interkorelasi antara item adalah bertujuan mengukur homogeniti item-item melalui pengujian instrumen bagi membantu mengukuhkan kesahan konstruk. Nilai korelasi antara ukuran item-item sepatutnya rendah iaitu kurang daripada 0.70 (Linacre 2012) menunjukkan kebebasan setempat. Ini bermakna tiada perkaitan wujud antara item.

Hakikatnya kelakuan manusia sangat kompleks dan boleh dikatakan bersifat multidimensi. Tapi kajian pelakuan manusia tidak boleh bermula dengan ciri multidimensi semata-mata. Setiap pemboleh ubah dalam penyelidikan perlu bermula dengan ciri unidimensi yang mantap. Apabila beberapa dimensi telah dibina, barulah kajian untuk menggabungkan dimensi tersebut dilakukan untuk mewujudkan suatu perhubungan. Contohnya dalam kajian konstruk KPS, subkonstruk yang terlibat ialah KPS asas dan KPS bersepadu. Untuk mengetahui sejauh mana subkonstruk tersebut benar mengukur apa yang sepatutnya

diukur, maka analisis unidimensi dibuat. Unidimensi menganggap item-item dalam instrumen mengukur satu keupayaan tunggal (Bambang & Wahyu 2014).

REKABENTUK DAN INSTRUMEN KAJIAN

Kajian ini berbentuk kuantitatif dan menggunakan reka bentuk tinjauan. Instrumen KePSa ditadbir dalam bentuk ujian berskala Likert lima mata bagi mendapatkan maklum balas murid terhadap tahap kebolehan KPS mereka. Sebanyak 68 item terdapat dalam ujian ini dan dibahagikan kepada 12 domain KPS. Teknik pungutan data bagi instrumen ini ialah secara bersemuka dalam kumpulan besar dijalankan di sekolah dengan izin pentadbir. Ujian KePSa dijawab berdasarkan maklum balas skala 1-kurang boleh, 2-agak boleh, 3-sederhana boleh, 4-boleh dan 5-sangat boleh dalam jangka masa satu jam lima belas minit. Contohnya untuk domain memerhati, terdapat 10 indikator. Antaranya indikator menyusun urutan peristiwa dengan betul. Item bagi indikator ini ialah murid diminta memberi maklum balas tentang tahap kebolehan mereka menyusun semula urutan fasa bulan jika diberikan kad gambar yang tidak mengikut susunan. Item ini perlu ditentukan oleh murid sendiri tahap kebolehannya setelah melalui pengalaman mempelajari topik fasa bulan semasa tahun lima. Kesemua item yang dibina dalam ujian KePSa merupakan item dalam silibus sains KSSR. Dua orang guru sains sekolah rendah yang berpengalaman luas dilantik sebagai konsultan untuk memastikan item-item ini adil untuk ditadbir kepada sampel kajian.

Pada usia 12 tahun, murid-murid mungkin terlebih membuat anggaran ketika menanda skala pemeringkatan tahap kebolehan diri sendiri kerana kelemahan kemahiran kognitif mereka untuk mengintegrasikan maklumat. Malah pada usia yang belum matang, mereka lebih terdedah kepada pemikiran berbentuk harapan (Ross 2006). Maka cara supaya instrumen ini boleh dipercayai perlu dikenal pasti. Antaranya ialah kehadiran pengkaji atau guru sewaktu pentadbiran instrumen di sisi sampel amat penting untuk mengawal dan membimbing mereka dalam hal ini. Murid juga perlu dikawal daripada berbincang sesama sendiri supaya keputusan kawan tidak mempengaruhi maklum balas diri sendiri. Sehubungan dengan itu, kawalan yang rapi dan penerangan terperinci daripada pengkaji telah meningkatkan keyakinan pengkaji dengan kejujuran murid sewaktu memberi maklum balas secara bertulis untuk ujian KePSa. Malah dapatan data empirik yang konsisten meningkatkan kesahan dan kebolehpercayaan pengukuran instrumen. Inilah kepentingan sesebuah instrumen mempunyai ciri psikometrik yang baik.

SAMPel KAJIAN

Produk Kurikulum Standard Sekolah Rendah, KSSR terbentuk sepenuhnya pada akhir tahun 2016 selepas enam

tahun silibus baru ini dilaksanakan setelah diperkenalkan mulai tahun 2011. Kohort pertama merujuk kepada murid tahun enam iaitu lepasan UPSR tahun 2016 yang kini berada di tingkatan dua. Pada tahun 2017, murid ini mengalami peralihan dari alam sekolah rendah ke sekolah menengah. Kohort kedua pula merupakan murid lepasan UPSR tahun 2017 yang kini berada di tingkatan satu.

Sampel kajian merupakan 514 orang murid tahun enam yang telah selesai menduduki peperiksaan UPSR. Mereka merupakan murid lepasan UPSR kohort kedua. Pemilihan sampel adalah secara rawak berstrata dari seluruh Malaysia melibatkan murid daripada 13 negeri dan tiga wilayah persekutuan dengan mengambil kira faktor lokasi dan jantina. Daripada 16 lokasi ini, 252 murid tinggal di luar bandar manakala 262 tinggal di bandar. Kesemua sampel terdiri daripada 226 orang murid lelaki dan 288 orang murid perempuan. Strata pertama ialah negeri, manakala strata seterusnya ialah lokasi diikuti dengan jantina. Sebelum ujian sebenar ditadbir, kajian rintis telah dilaksanakan berdasarkan teknik persampelan rawak mudah untuk mengenal pasti kekuatan dan kelemahan instrumen baru ini. Bagi tujuan ini, seramai 300 orang murid tahun enam kohort pertama yang tidak terlibat dengan kajian sebenar telah dipilih untuk menjawab ujian KePSa.

TEKNIK PENGANALISISAN DATA

Dapatan kajian dianalisis secara kuantitatif menggunakan model pengukuran Rasch. Analisis kajian rintis menunjukkan kesemua aspek keserasian item, polariti item, kebebasan setempat dan analisis unidimensi memenuhi julat panduan. Indeks kebolehpercayaan item berada pada tahap 0.96 manakala kebolehpercayaan individu ialah 0.99. Indeks pengasingan item direkodkan pada 9.47 dan pengasingan individu ialah 5.51. Oleh kerana dapatan analisis kesahan item dan kebolehpercayaan item dan individu pada kajian rintis telah mematuhi kehendak model pengukuran Rasch maka, instrumen telah diteruskan untuk ditadbir ke atas sampel sebenar.

Selain kesahan item, instrumen KePSa turut telah dianalisis dalam aspek kesahan kandungan menggunakan kaedah *Fuzzy Delphi* (FDM). Seramai tiga belas orang pakar dalam kalangan pendidik bidang sains pendidikan di sekolah rendah, sekolah menengah, institut perguruan, universiti awam, universiti swasta dan bahagian pembangunan kurikulum dipilih untuk memeriksa kandungan instrumen. Kaedah persampelan pakar adalah secara bertujuan dengan melihat ciri kepakaran dan memberi sumbangan dalam bidang sains pendidikan. Antara sumbangan pakar ialah penulis buku teks, penulis modul, fasilitator lantikan RECSAM (Regional Centre for Education in Science and Mathematics) dan lain-lain lagi. Kesemua tiga prasyarat FDM dipatuhi di mana nilai ambang d bagi setiap item yang diukur adalah kurang atau sama dengan 0.2 dan peratusan kesepakatan kumpulan pakar adalah melebihi 75%. Malah nilai *defuzzification* juga iaitu nilai skor *Fuzzy*

(A) didapati melebihi atau sama dengan nilai potongan- α iaitu 0.5 (Mohd Ridhuan et al. 2014; Ramlan & Ghazali 2018). Sebahagian analisis FDM kajian ini boleh dirujuk pada artikel oleh Ellyza et al. (2017).

Kesahan item dalam artikel ini melibatkan penentuan keserasian item, polariti item, kebebasan setempat dan analisis unidimensi yang dianalisis menggunakan model pengukuran Rasch melalui perisian Statistik IBM SPSS versi 22 dan *Winsteps* versi 3.73. Kebolehpercayaan turut dianalisis berdasarkan model pengukuran Rasch dalam dua aspek iaitu indeks kebolehpercayaan serta indeks pengasingan. Bukan setakat analisis item malah kedua-dua indeks ini turut mampu menganalisis individu juga.

DAPATAN DAN PERBINCANGAN

Jadual 1 menunjukkan dapatan kesahan dalam model pengukuran Rasch yang dikaji daripada nilai *fit* atau nilai keserasian bagi setiap item. Item yang tidak *fit* merupakan item yang perlu diubah suai atau dibuang. Item-item yang *fit* menunjukkan item yang boleh mengukur satu konstruk yang sama. Pengkaji lepas menetapkan julat yang perlu dipatuhi bagi menguji andaian keserasian item iaitu nilai *MNSQ* item harus berada pada julat 0.5 – 1.5 logit (Boone et al. 2014; Bambang & Wahyu 2014). Keserasian item melalui *infit MNSQ* dan *oufit MNSQ* yang berada dalam julat tersebut yang mana bersesuaian bagi data politomus memberi makna bahawa kesemua item adalah sepadan dengan model Rasch dan kesemua item adalah produktif dalam mengukur tahap kebolehan KPS responden serta tidak menimbulkan kekeliruan kepada responden untuk memberi respons kepada item-item dalam instrumen KePSa. Nilai

MNSQ yang rendah menandakan berlaku pertindihan item manakala nilai *MNSQ* yang tinggi menandakan item berada di luar konsep (Khan et al. 2013). Dapatan kajian ini menunjukkan nilai *fit* berada di antara 0.77-1.49 yang mana mematuhi nilai panduan. Menurut Bambang dan Wahyu (2014), nilai *MNSQ* yang lebih rendah daripada 0.5 menandakan item itu mudah diramal manakala nilai *MNSQ* yang lebih daripada 1.5 menandakan item sukar diramal. Selain itu, nilai *oufit ZSTD* dan *infit ZSTD* perlu juga perlu berada dalam lingkungan -2 hingga +2 (Bond & Fox 2015), namun jika nilai *oufit* dan *infit MNSQ* diterima, indeks ZSTD boleh diabaikan (Linacre 2012). Maka dalam kajian ini indeks ZSTD tidak diberi perhatian.

Nilai polariti item yang diperolehi sebaiknya memiliki nilai positif dan melebihi 0.3 logit (Bond & Fox 2015; Wu & Adams 2007). Nilai polariti yang baik membuktikan item mengukur konstruk yang ingin diukur dan semua item bergerak ke arah pengukuran sub konstruk tunggal. Dapatan kajian ini menunjukkan kesemua nilai PTMEA ialah positif bagi kesemua 68 item dengan nilai logit antara 0.36 hingga 0.70 maka kesemua item dilihat telah mengukur konstruk yang ingin diukur. Item dengan nilai logit PTMEA 0.36 merupakan item termudah dalam ujian KePSa, yang berada dalam domain pemerhatian di bawah indikator menggunakan deria rasa. Bagi murid, amat mudah untuk membezakan rasa masam, manis, masin dan pahit. Manakala item paling sukar dengan nilai logit PTMEA 0.70 ialah item dalam domain mengeksperimen di bawah indikator menstruktur perancangan prosedur. Murid merasakan amalan mereka bentuk langkah-langkah menjalankan eksperimen (prosedur) secara saintifik merupakan item yang sangat menguji tahap kebolehan KPS mereka.

JADUAL 1. Analisis kesahan dan kebolehpercayaan berdasarkan model pengukuran Rasch

Kriteria	Perincian	Nilai/ Julat
Keserasian item (<i>fit</i>)	<i>Infit MNSQ</i>	0.77-1.41
	<i>Oufit MNSQ</i>	0.79-1.45
Polariti item	Korelasi ukuran titik pusat (PTMEA)	0.36-0.70
Kebebasan setempat	Korelasi residual terpiawai	0.32-0.52
Unidimensionaliti	PCA	49.1
KPS keseluruhan	Model yang dijangka	50.0
	Tahap gangguan	3.1
	Nilai eigen	4.1
KPS asas	PCA	47.3
	Model yang dijangka	47.9
	Tahap gangguan	3.4
KPS bersepadu	Nilai eigen	2.7
	PCA	52.8
	Model yang dijangka	53.6
Kebolehpercayaan individu	Tahap gangguan	4.7
	Nilai eigen	2.6
	Kebolehpercayaan item	0.99
Pengasingan individu		5.41
Pengasingan item		12.71

Dapatan kebebasan setempat iaitu padanan item dengan nilai piawai korelasi residual menunjukkan nilai antara 0.32 hingga 0.52. Julat ini menepati syarat kebebasan setempat iaitu nilai korelasi kurang daripada 0.70 (Linacre 2012). Padanan item pada julat ini memiliki nilai korelasi yang lemah. Ini bermaksud item-item itu tidak bersandar dengan item lain dalam subkonstruk yang sama. Berbeza dengan kajian rintis yang dijalankan Faizal Amin et al. (2014) yang menunjukkan beberapa itemnya yang mengalami masalah kebebasan setempat kerana melebihi 0.7. Dapatan sebegini memberi makna bahawa item-item ini mempunyai maksud pengukuran yang sama atau menggabungkan beberapa dimensi lain yang dikongsi bersama. Oleh itu, item yang melebihi nilai 0.7 perlu diberi perhatian dan perlu digugurkan salah satu item bagi setiap pasang item yang terlibat. Namun, dalam kajian ini perkara tersebut tidak berlaku.

Unidimensi menganggap item-item dalam instrumen mengukur satu keupayaan tunggal (Wright & Master 2002). Tiga ciri yang dipandang untuk mengkaji unidimensi ialah analisis komponen prinsipal (PCA) iaitu nilai varians yang dijelaskan oleh ukuran, tahap gangguan item dalam kontras pertama dan nilai Eigen. Dapatan KPS secara keseluruhan memenuhi syarat pertama iaitu melebihi 20 peratus (Conrad et al. 2012) dengan nilai PCA sebanyak 49.1 peratus, sangat hampir dengan nilai model iaitu 50.0 peratus. Bagi KPS asas nilai PCA ialah 47.3 menghampiri nilai model iaitu 47.9 peratus manakala bagi KPS bersepadu nilai PCA ialah 52.8 menghampiri nilai model iaitu 53.6 peratus.

Bagi syarat kedua iaitu tahap gangguan item yang seharusnya berada di bawah nilai 15 peratus (Fisher 2007) juga dipatuhi oleh kedua-dua subkonstruk. KPS asas mencatat tahap gangguan pada 3.4 peratus manakala KPS bersepadu 4.7 peratus. Secara keseluruhan tahap gangguan adalah baik iaitu hanya 3.1 peratus. Syarat ketiga pula iaitu nilai Eigen menunjukkan nilai Eigen keseluruhan 4.1. Nilai dapatan kurang daripada lima menunjukkan tidak wujud dengan jelas dimensi kedua (Linacre 2005). Kedua-dua nilai Eigen subkonstruk juga mematuhi syarat ini iaitu 2.7 untuk KPS asas dan 2.6 untuk KPS bersepadu. Maka konstruk KPS dalam kajian ini mencapai unidimensi, begitu juga dengan subkonstruk KPS asas dan KPS bersepadu.

Bambang dan Wahyu (2014) menentukan nilai di bawah 0.67 sebagai kebolehpercayaan yang lemah, 0.67-0.80 sebagai mencukupi, 0.81-0.90 sebagai bagus, 0.91-0.94 sebagai sangat bagus dan di atas 0.94 sebagai istimewa. Kebolehpercayaan menjelaskan sejauh mana pengukuran yang dilakukan berulang kali tetapi tetap konsisten menghasilkan maklumat yang sama. Maklumatnya bukan tidak ada perbezaan langsung, masih terdapat perbezaan tetapi nilainya kecil dan masih dalam had yang diterima. Nilai kebolehpercayaan kajian ini bagi individu ialah 0.97 manakala bagi item ialah 0.99. Nilai menghampiri 1.00 ini dapat merumuskan bahawa konsistensi maklum balas daripada murid kuat dan kualiti item juga bagus.

Indeks pengasingan individu digunakan untuk mengelaskan sampel. Nilai pengasingan individu yang

rendah dengan sampel yang relevan menunjukkan bahawa instrumen itu mungkin tidak cukup sensitif untuk membezakan antara sampel beraras kognitif tinggi dengan yang rendah. Lebih banyak item diperlukan. Indeks pengasingan item pula digunakan untuk mengesahkan hierarki item. Nilai pengasingan item yang rendah iaitu nilai kurang daripada tiga, untuk aras kesukaran item peringkat tinggi, sederhana dan rendah menunjukkan bahawa sampel tidak cukup besar untuk mengesahkan hierarki aras kesukaran item pada instrumen itu (Boone et al. 2014). Jones dan Fox (1998) menetapkan bahawa jika melebihi nilai dua, bilangan sampel individu cukup besar untuk mengesahkan hierarki kesukaran item instrumen. Bagi indeks pengasingan individu juga, ini menunjukkan bahawa instrumen cukup sensitif untuk membezakan antara individu yang berprestasi tinggi dan rendah. Dalam kajian ini kedua-dua indeks pengasingan individu dan item melepasi nilai rujukan ini iaitu 5.41 dan 12.71 masing-masing.

Sebagai rumusan, analisis kesahan item dan kebolehpercayaan item-individu pada instrumen KePSa telah mematuhi syarat yang dirujuk daripada para pelopor dan pengkaji model pengukuran Rasch terdahulu. Instrumen KePSa didapati relevan digunakan untuk mengkaji tahap kebolehan KPS murid lepasan UPSR malah boleh dikaitkan dengan pencapaian akademik. Bukan setakat guru sekolah rendah malah guru sains di tingkatan satu boleh menggunakan instrumen ini untuk menguji tahap kebolehan KPS murid yang baru mereka kenali sekaligus meramal aras pencapaian akademik murid tersebut sebelum meneruskan sesi pengajaran dan pembelajaran. Kepentingan mengetahui tahap kebolehan KPS murid atau tahap pencapaian sains terdahulu mereka memudahkan guru merancang pengajaran dan memilih strategi atau pendekatan pengajaran yang sesuai untuk dilaksanakan. Kajian oleh Shahlan et al. (2017) menunjukkan KPS mempunyai hubungan yang signifikan dengan pencapaian akademik pelajar. Dapatan ini bermakna apabila pelajar memiliki kemahiran proses sains yang tinggi, mereka turut akan mempunyai pencapaian akademik yang baik dalam peperiksaan.

KESIMPULAN

Secara keseluruhannya, pengujian ciri psikometrik bagi instrumen KePSa mendapati item-item KePSa mempunyai kesahan dan kebolehpercayaan yang mematuhi syarat menggunakan analisis pengujian model pengukuran Rasch. Kajian ini mampu menyumbang kepada suatu instrumen pengukuran KPS berciri teori respons item yang sesuai digunakan dalam konteks murid sekolah rendah di Malaysia. Usaha ini dapat menyokong pengujian teori pengukuran klasik dalam pembinaan instrumen dan menawarkan reka bentuk instrumen pengukuran alternatif terhadap pengukuran KPS. Analisis sebegini membantu para penyelidik berfikir dalam cara yang lebih

canggih berkenaan dengan pemboleh ubah yang ingin diukur. Banyak manfaat lain boleh dikaji menggunakan model pengukuran Rasch. Kajian lanjutan menggunakan peta Wright boleh diteruskan untuk mengkaji kebolehan murid dan aras kesukaran item. Malah melalui pengukuran perbezaan kefungsiian item (DIF), kajian mendalam boleh dijalankan berdasarkan faktor tertentu seperti jantina atau lokasi khususnya apabila mempertimbangkan pemboleh ubah tersebut merupakan antara prediktor utama pencapaian KPS dalam kalangan pelajar. Justeru, kajian analisis item ini tidak seharusnya terhenti setakat ini, malah harus diteruskan dari semasa ke semasa dengan menambah baik ujian KePSa seperti yang telah dibincangkan. Tidak terbatas di situ, dapatan ujian KePSa turut boleh direplikasi untuk mengkaji perhubungannya dengan pemboleh ubah lain seperti meramal pencapaian sains murid atau motivasi pelajar dalam pembelajaran sains.

RUJUKAN

- Abdul Halim, Lilia Halim, Thamby Subahan Mohd Meerah & Kamisah Osman. 2010. Pembangunan instrumen penyelesaian masalah sains. *Jurnal Pendidikan Malaysia* 35(1).
- Abu Hassan Kassim & Rohana Hussin. 2003. Tahap penguasaan kemahiran proses sains dan hubungannya dengan pencapaian kimia di kalangan pelajar tingkatan empat Daerah Johor Bahru. Kertas kerja *Seminar Kebangsaan Pendidikan*. Anjuran Fakulti Pendidikan Universiti Teknologi Malaysia. Skudai, 19-21 Oktober.
- Adibah Abdul Latif. 2013. Pembangunan instrumen penilaian akhlak pelajar di institusi pengajian tinggi awam. Tesis Dr. Fal, Fakulti Pendidikan, Universiti Teknologi Malaysia, Skudai.
- Azrilah Abdul Aziz, Mohd Saidfudin Masodi & Azami Zaharim. 2013. *Asas Model Pengukuran Rasch: Pembentukan Skala & Struktur Pengukuran*. Bangi: Fakulti Pendidikan UKM.
- Bahagian Pembangunan Kurikulum (BPK). 2015. *Dokumen Standard Kurikulum dan Pentaksiran Sains Tingkatan Satu*. Putrajaya: Kementerian Pendidikan Malaysia.
- Bambang Sumintono & Wahyu Widhiarso. 2014. *Aplikasi Model Rasch untuk Penelitian Ilmu-ilmu Sosial*. Cimahi: Trim Komunikata Publishing House.
- Bambang Sumintono & Wahyu Widhiarso. 2015. *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Cimahi: Trim Komunikata Publishing House.
- Bond, T. G. & Fox, C. M. 2015. *Applying the Rasch Model Fundamental Measurement in the Human Sciences*. 3rd ed. New Jersey: Lawrence Erlbaum Associates.
- Boone, W. J. 2016. Rasch analysis for instrument development: Why, when, and how? *CBE Life Sciences Education Journal* 15(4). doi:10.1187/cbe.16-04-0148
- Boone, W. J., Staver, J. R. & Melissa, S. Y. 2014. *Rasch Analysis in the Human Sciences*. Dordrecht: Springer.
- Conrad, K.M., Conrad, K.J., Dennis, M.L. & Funk, R. 2012. *Validation of the self help improvement scale to the Rasch measurement model GAIN methods report 1.0*. http://gaincc.org/_data/files/Psychometrics_and_Publications/Working_Papers/Conrad_et_al_2011_SPS_Report.pdf.
- Edy Hafizan Mohd Shahali & Lilia Halim. 2010. Development and validation of a test of integrated science process skills. *Procedia Social and Behavioral Sciences* 9: 142-146. doi:10.1016/j.sbspro.2010.12.127.
- Ellyza Karim, Jamil Ahmad & Kamisah Osman. 2017. Fuzzy delphi method for content validation of integrated science process skills instrument. *International Journal of Academic Research in Business and Social Sciences* 7(6): 773. doi: 10.6007/IJARBS/v7-i6/3037.
- Faizal Amin Nur Yunus, Ruhizan Mohammed Yasin, Mohd Bekri Rahim & Ridzwan Che Rus. 2014. Kesahan dan kebolehppercayaan instrumen pemindahan pembelajaran berdasarkan pendekatan model Rasch. *First Technical and Vocational Education International Seminar*. doi:10.13140/2.1.2655.9686.
- Fatin Aliah Phang & Nor Athirah Tahir. 2011. Kemahiran proses sains dalam kalangan mahasiswa pendidikan sains di UTM. *Journal of Science and Mathematics Educational* 3: 123-133.
- Fisher, J.W.P. 2007. Rating scale instrument quality criteria. *Rasch Measurement Transactions* 21(1): 1095.
- Freitag, A. & Pfeffer, M.J. 2013. Process, not product: Investigating recommendations for improving citizen science "Success". *PLoS ONE* 8(5). <https://doi.org/10.1371/journal.pone.0064079>.
- Gagne, R.M. 1965. The psychological basis of science - a process approach. *AAAS Miscellaneous Publication*: 65-68.
- Jones, J.A. & Fox, C.M. 1998. Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology* 45(1): 30-45.
- Kamus Dewan Edisi Keempat. 2016. Kuala Lumpur: Percetakan Rina.
- Khan, A., Chien, C.-W. & Brauer, S.G. 2013. Rasch-based scoring offered more precision in differentiating patient groups in measuring upper limb function. *Journal of Clinical Epidemiology* 66(6): 681-687. doi:10.1016/j.jclinepi.2012.12.014.
- Lim, P.H. & Ong, S.L. 2009. Perbandingan pencapaian murid dalam ujian sains berdasarkan versi Bahasa Inggeris dan Bahasa Melayu. *Jurnal Pendidik dan Pendidikan* 24: 143-164.
- Linacre, J.M. 2005. Standard errors: means, measures, origins and anchor values. *Rasch Measurement Transactions* 19(3): 1030.
- Linacre, J.M. 2012. *A user's guide to WINSTEPS: Rasch model computer programs*. Chicago: MESA Press.
- Ong, E.T. & Bibi Hazliana Mohd Hassan. 2014. Penguasaan kemahiran proses sains asas dalam kalangan murid sekolah rendah di selangor berdasarkan jantina, lokasi sekolah dan aras tahun. *Jurnal Pendidikan Sains & Matematik Malaysia* 3(2).
- Ong, E.T. & Mohd. Al-Junaidi Mohamad. 2013. Pembinaan dan penentuan instrumen kemahiran proses sains untuk sekolah menengah. *Jurnal Teknologi (Social Sciences)* 66(1): 7-20.
- Ong, E.T., Norjuhana Mesmen, Siti Eshah Mokshein, Sabri Mohd Salleh, Nik Azmah Nik Yusuff dan Koon, P.Y. 2015. Basic science process skills test for primary schools: Item development and validation. *Jurnal Pendidikan Sains dan Matematik Malaysia* 5(1): 70-88.

- Ong, E.T., Wong, Y.T., Sopia Mad Yassin, Sadiyah Baharom, Asmayati Yahya & Zahid Md. Said. 2012. Malaysian based science process skills inventory: Development, validation and utilisation. *CREAM – Current Research in Malaysia* 1(1): 125-149.
- Punia Turiman, Jizah Omar, Adzliana Mohd Daud & Kamisah Osman. 2012. Fostering the 21st century skills through scientific literacy and science process skills. *Procedia Social and Behavioral Sciences* 59: 110-116. doi: 10.1016/j.sbspro.2012.09.253.
- Ramlan Mustapha & Ghazali Darusalam. 2018. *Aplikasi Kaedah Fuzzy Delphi dalam Penyelidikan Sains Sosial*. Kuala Lumpur: Penerbit Universiti Malaya.
- Rian Vebrianto & Kamisah Osman. 2012. Keberkesanan penggunaan pelbagai media pengajaran dalam meningkatkan kemahiran proses sains dalam kalangan pelajar. *Jurnal Pendidikan Malaysia* 37(1): 1-11.
- Ross, J.A. 2006. The reliability, validity and utility of self-assessment. *Practical Assessment Research & Evaluation* 11(10). <http://pareonline.net/getvn.asp?v=11&n=10>.
- Shahlan Surat, Natasha, Saadiah Kummin & Shalinawati Ramli. 2017. Kreativiti sains, kemahiran proses sains dan pencapaian akademik pelajar sekolah menengah. *Prosiding Seminar Pendidikan Serantau ke-8 2017*: 896.
- Siti Rahayah Ariffin. 2008. *Inovasi dalam Pengukuran dan Penilaian Pendidikan*. Bangi: Fakulti Pendidikan UKM.
- Siti Rahayah Ariffin, Jamil Ahmad & Nur 'Ashiqin Najmuddin. 2010. Pembangunan instrumen kemahiran generik pelajar berasaskan penilaian pensyarah dengan menggunakan model pengukuran Rasch pelbagai faset. *Jurnal Pendidikan Malaysia* 35(2): 43-50.
- Siti Rahayah Ariffin, Suriana Mohd Asari, Shafiza Mohamed, Shah Nazim Shahar, Noriah Mohd Ishak, Rosseni Din, Abdul Ghafur Ahmad, Hamidah Yamat, Rosadah Majid & Siti Fatimah Mohd Yasin. 2010. Validity of UKM1 intelligence test using Rasch analysis. *Procedia Social and Behavioral Sciences* 7: 205-209. doi: 10.1016/j.sbspro.2010.10.029.
- Wright, B. D. & Masters, G. N. 2002. Number of person or item strata. *Rasch Measurement Transactions* 16(3): 888.
- Wright, B. D. & Stone, M. H. 2004. *Making Measures*. Chicago: The Pheneron Press.
- Wu, M. & Adams, R. 2007. *Applying the Rasch Model to Psychosocial Measurement: A Practical Approach*. Melbourne: Educational Measurement Solutions.
- Ellyza Karim*
Jabatan Sains
Institut Pendidikan Guru Kampus Pendidikan Teknik
71760 Bandar Enstek
Negeri Sembilan
Emel: ellyza@pendidikguru.edu.my
- Kamisah Osman
Fakulti Pendidikan
Universiti Kebangsaan Malaysia
43600 Bangi
Selangor
Emel: kamisah@ukm.edu.my
- Pengarang untuk surat-menyurat, e-mel: ellyza@pendidikguru.edu.my
- Diserahkan: 15 Ogos 2018
Diterima: 1 Oktober 2018
Diterbitkan: 31 Disember 2018

