

Kappa Fleiss Analysis: Evidence Of Content Validity for Formative Assessment Literacy Test for Teachers of Fundamental Subjects in Vocational Special School

Analisis Fleiss Kappa: Evidens Kesahan Kandungan Ujian Literasi Pentaksiran Formatif Guru-Guru Subjek Asas Sekolah Menengah Pendidikan Khas Vokasional

Ibnatul Jalilah Yusof*, Fathin Edora Abdul Rahim

School of Education in Measurement and Evaluation, Faculty of Social Sciences and Humanities, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

*Corresponding author: ijalilah@utm.my

Article history: Received: 19 November 2022 Received in revised form: 26 June 2023 Accepted: 27 March 2024 Published online: 30 April 2024

Abstract

The Formative Assessment Literacy Test (FAT) is a new instrument designed to identify and assess fundamental subject teachers' formative assessment literacy (FAL) in Vocational Special School. In order to determine the level of FAL, teachers require a precise measurement tool in terms of content validity as proven by expert validity. The purpose of this study is to validate the FAT instrument through expert validity and expert agreement standardization by using Kappa Fleiss analysis. Three experts with more than ten years of experience in the field of measurement and evaluation education, as well as an expert from the field of Special Education were chosen to verify the FAT content. Overall, the FAT content is rated as excellent with kappa (k) value of 0.84. While for each knowledge component's concept, principle, and skill component's method, $k = 0.81$, 0.81 , and 0.89 respectively. Any changes and alterations is performed in response to expert advice and suggestion. In conclusion, based on expert evaluation, FAT has good content validity (excellent) and may be used to detect and measure teachers' knowledge and abilities on formative assessment.

Keywords: Formative assessment tes, formative assessment, expert validity, Fleiss Kappa analysis

Abstrak

Ujian literasi pentaksiran formatif (ULPF) adalah merupakan instrumen baru yang dibina khusus untuk mengenalpasti dan mengukur tahap literasi pentaksiran formatif (LPF) guru-guru dalam Pendidikan Khas Vokasional. Untuk mengetahui tahap LPF guru memerlukan alat pengukuran yang jitu dari segi kesahan kandungan melalui kesahan pakar. Kajian ini adalah bertujuan untuk mengesahkan instrumen ULPF melalui kesahan pakar dan keseragaman persetujuan dari pakar dengan menggunakan analisis Kappa Fleiss. Tiga pakar telah dipilih untuk mengesahkan kandungan ULPF yang mempunyai pengalaman dalam bidang pengukuran dan penilaian dalam pendidikan melebihi 10 tahun dan seorang pakar daripada bidang Pendidikan Khas. Secara keseluruhan, kandungan item ULPF dikategorikan sebagai cemerlang dengan nilai kappa (k) secara keseluruhannya adalah 0.84. Manakala bagi setiap komponen pengetahuan iaitu konsep $k = 0.81$, prinsip $k = 0.81$ dan komponen kemahiran iaitu kaedah $k = 0.89$. Perubahan dan modifikasi juga dilakukan berdasarkan saranan dan ulasan yang diberikan oleh pakar. Kesimpulannya, berdasarkan semakan pakar dan nilai k , ULPF mempunyai kesahan kandungan yang tinggi (cemerlang) dan ianya boleh digunakan untuk mengenal pasti serta mengukur pengetahuan dan kemahiran guru terhadap pentaksiran formatif.

Kata kunci: Literasi pentaksiran formatif, pentaksiran formatif, kesahan pakar, analisis Kappa Fleiss

© 2024 Penerbit UTM Press. All rights reserved

■1.0 PENGENALAN

Sekolah Menengah Pendidikan Khas Vokasional (SMPKV) adalah salah sebuah sekolah pendidikan khas yang berada di bawah Program Pendidikan Khas Integrasi dan ditadbir oleh Jabatan Pendidikan Teknikal (Kementerian Pelajaran Malaysia, 2021). SMPKV dibina bertujuan menyediakan pendidikan vokasional kepada murid berkeperluan khas (MBK) atau sekarang dikenali sebagai murid berkeperluan pendidikan khas (MBPK) untuk menerima latihan kemahiran yang berkaitan dengan pekerjaan yang bersesuaian dengan kekurangan mereka (Bahagian Pendidikan Khas, 2019). Pendidikan KV ini dikatakan lebih fleksibel, berfokus, menarik serta bersesuaian dengan keupayaan dan kebolehan MBPK untuk menguasai ilmu kemahiran (BPK, 2019). Selain menyediakan MBPK dengan kemahiran vokasional untuk mendapatkan pekerjaan, penerapan kemahiran keboleherjaan juga diperlukan untuk memastikan MBPK benar-benar bersedia untuk menghadapi situasi sebenar dalam alam pekerjaan.

Terdapat tiga jenis kemahiran keboleherjaan yang penting boleh diterapkan kepada MBPK menurut Anizam *et al.*, (2020) iaitu (i) kemahiran asas (membaca, menulis, mengira), (ii) kemahiran kompetensi ditempat kerja dan (iii) kemahiran keusahawanan. Tambah Anizam *et al.*, (2020) lagi, penerapan asas seperti kemahiran membaca dan mengira (literasi dan numerasi) merupakan kemahiran yang menjadi cabaran utama untuk MBPK. Kajian mendapati kadar pengangguran dalam kalangan MBK masih berada pada tahap yang membimbangkan (Syed Chear, 2018; Halim, Muda & Izam, 2019). Antara faktor yang menyumbang kepada kadar penangguran dalam kalangan MBK adalah kerana mereka dikatakan masih lemah untuk menguasai kemahiran asas seperti kemahiran komunikasi, membaca, menulis dan mengira (Ab Halim & Wan Mohamed, 2019). Bahkan, perkara ini turut didapati dalam Mitchell dan Karchmer (2012) dan disokong melalui kajian Anizam *et al.* (2020) yang mendapati kelemahan MBPK menguasai kemahiran seperti membaca, menulis, mengira serta berkomunikasi menyebabkan mereka tidak berjaya mendapatkan pekerjaan.

Namun begitu, menurut Omar dan Ali (2019) kajian spesifik di SMPKV seperti penerapan kemahiran asas masih lagi kurang di Malaysia. Malah, kebanyakan kajian lebih tertumpu kepada program transisi kerjaya secara teknikal seperti kajian Tahir, Mustafa dan Yasin (2009), Anizam *et al.* (2013), Haryanti (2017), dan Izam dan Wan Mohamed (2020) yang kurang mengambil kira kemahiran asas MBK. Kajian menunjukkan penerapan pembelajaran yang baik bagi MBK adalah melalui pentaksiran formatif (Butler & Schenellert, 2015; Brookhart & Lazarus, 2017). Walaubagaimanapun, masih terdapat jurang dalam kajian yang melibatkan pentaksiran formatif khususnya dalam merangka dan menambah baik amalan dan proses pentaksiran formatif bagi MBK di Malaysia sekaligus menjadi tanda tanya adakah guru-guru di SMPKV mempunyai literasi pentaksiran yang mencukupi. Kesannya, tiada pendekatan atau penambahbaikan dapat dilakukan kerana kajian yang khusus untuk mengetahui amalan dan tahap literasi pentaksiran formatif guru-guru pendidikan khas adalah sangat terhad.

Kebanyakan kajian literasi pentaksiran adalah tidak spesifik kepada literasi pentaksiran formatif serta hanya melibatkan guru-guru dalam aliran sekolah biasa (Lian, Yew & Meng, 2014; Khalid *et al.*, 2021; Talib, 2008; Saion, 2014), serta pensyarah-pensyarah di Institusi tinggi (Muhammad *et al.*, 2020; Sulaiman, Kotamjani, Rahim & Hakim; 2020). Kajian ini melihat keperluan untuk mengukur tahap literasi pentaksiran formatif guru-guru pendidikan khas terutamanya bagi guru-guru SMPKV. Hal ini kerana, pentaksiran formatif adalah merupakan pentaksiran yang paling banyak diaplikasikan melalui teknik pembelajaran yang biasa guru-guru gunakan seperti menyoal secara lisan, latihan, perbincangan, projek dan tugas. Kesemua teknik ini apabila dilaksanakan dengan betul dan berkesan, ia dapat membantu meningkatkan serta membantu dalam pembelajaran MBPK.

Oleh itu, pembinaan ujian literasi pentaksiran formatif ini adalah penting kerana melalui hasil ujian ini tahap pengetahuan dan kemahiran guru mengaplikasikan pentaksiran formatif dalam bilik darjah boleh diketahui. Hasil daripada kajian ini turut boleh digunakan untuk menambahbaik literasi pentaksiran formatif guru-guru subjek asas di SMPKV. Walaubagaimanapun, ujian literasi pentaksiran formatif ini dibina khusus hanya untuk guru-guru subjek khas di SMPKV yang memenuhi kesahan kandungan sebuah ujian. Selain penghasilan jadual spesifikasi ujian, bukti kesahan kandungan juga dikumpul melalui analisis Kappa Fleiss bagi mengukur keseragaman persetujuan pakar dalam menilai item-item ujian yang dibina sekaligus.

■2.0 KAJIAN LITERATUR

2.1 Literasi Pentaksiran Formatif

Tema asal literasi pentaksiran (LP) mula dibangunkan dan dikenalkan oleh Richard Stiggins (1995). LP membawa maksud sebagai pengetahuan asas terhadap pentaksiran dan kemahiran yang diperlukan untuk melaksanakan pengetahuan tersebut (Stiggins, 1995). Berdasarkan definisi yang dibangunkan itu, Stiggins (1995) berpendapat guru yang celik pentaksiran adalah mereka yang tahu apa yang dinilai, mengapa ianya dinilai, cara ianya dinilai dan bagaimanakah untuk mengelak serta mencegah dari masalah yang mungkin timbul semasa proses penilaian dilakukan (Stiggins, 1995). Tambahan pada itu, Stiggins (1995) juga percaya bahawa melalui LP yang baik, seorang guru juga mampu untuk mengesan akibat daripada penilaian yang tidak tepat.

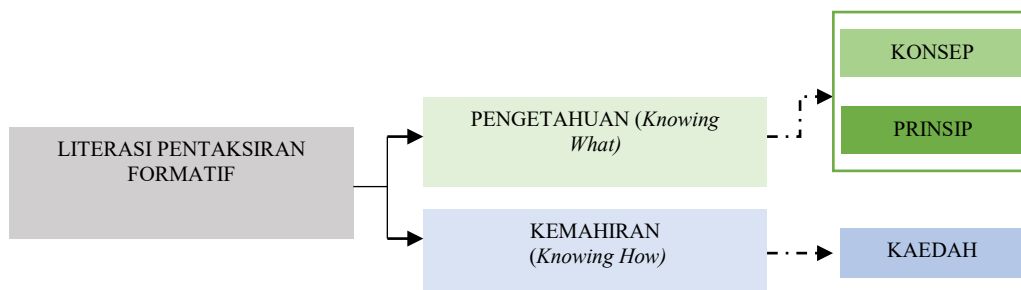
Definisi LP juga telah berkembang dan difafsirkan dengan mengikut keperluan pentaksiran dalam pendidikan. Sebagai contoh, Popham (2004) beranggapan LP adalah kefahaman individu tentang prinsip asas pentaksiran untuk mahir dalam membangunkan pentaksiran yang mengubah pembelajaran kepada aktiviti pentaksiran. Bagi Mertler dan Campbell (2005) pula, LP adalah kemahiran membentuk aktiviti penilaian kepada kefahaman dan pencapaian murid. Bahkan, LP juga merangkumi pelbagai pengetahuan, kemahiran dan sifat yang diperlukan untuk memahami tujuan dan proses pentaksiran (Smith *et al.*, 2013).

Walaubagaimanapun, Talib (2008) merujuk LP sebagai kompetensi pengetahuan guru dalam bidang pentaksiran dimana pengetahuan itu adalah merangkumi kepada amalan prinsip, tatacara, strategi dan kaedah pentaksiran yang sesuai yang membantu kepada proses pembelajaran dan membantu guru untuk membuat keputusan yang adil serta tepat tentang murid. Namun begitu, bagi Chen Jiamu (2001) pula LP adalah meliputi kepada pengetahuan mengenai konsep, prinsip (*knowing what*) dan kaedah (*knowing how*) serta prosedural

aplikasi pengetahuan kepada amalan pentaksiran. Oleh kerana setiap definisi dibangunkan dengan konstruk yang berbeza, pengkaji telah melakukan semakan semula definisi LP. Melalui semakan itu, majoriti penyelidik bersetuju bahawa literasi pentaksiran adalah terdiri daripada dua konstruk yang utama iaitu (i) pengetahuan dan (ii) kemahiran pentaksiran, dimana pengetahuan itu adalah merangkumi kepada konsep dan prinsip pentaksiran manakala kemahiran pula dirujuk sebagai kaedah pentaksiran yang mengaplikasikan pengetahuan pentaksiran kepada amalan pentaksiran.

Dalam konteks pendidikan khas, pentaksiran memainkan peranan yang penting. Sebagai contoh, pentaksiran formatif berfungsi membantu pembelajaran murid berkeperluan pendidikan khas (MBPK) dengan mengesan kelemahan dan kekuatan, menggalakkan motivasi pembelajaran, membantu pembelajaran sendiri serta memberikan maklum balas kepada murid (Yan *et al.*, 2021). MBPK adalah murid yang kompleks dan mereka memerlukan pendekatan serta keperluan pembelajaran yang unik. Oleh kerana itu, guru-guru pendidikan khas perlu memahami sepenuhnya disebalik pendekatan dan keperluan tersebut. Untuk memahaminya, guru-guru memerlukan pengetahuan yang jitu berkenaan pentaksiran dan kemahiran mengaplikasikannya untuk mengintepretasikan maklumat tentang MBPK. Oleh yang demikian, literasi pentaksiran formatif adalah penting bagi setiap guru-guru pendidikan khas untuk memastikan pentaksiran formatif dapat digunakan dan berfungsi dengan baik dalam pembelajaran MBPK.

Justeru itu, dalam konteks kajian ini definisi literasi pentaksiran formatif dibangunkan dan diadaptasi daripada definisi-definisi LP dari pengkaji-pengkaji terdahulu dan dibentuk supaya ianya selari dengan keperluan dalam pentaksiran formatif. Definisi literasi pentaksiran formatif (LPF) dalam kajian ini adalah membawa maksud pengetahuan dan kemahiran guru dalam pentaksiran formatif seperti yang terdapat dalam Rajah 1. Secara spesifik, konstruk pengetahuan (*knowing what*) tersebut mengukur kepada konsep dan prinsip pentaksiran formatif manakala konstruk kemahiran (*knowing how*) itu mengukur kepada kaedah pentaksiran formatif. Oleh itu, Ujian Literasi Pentaksiran Formatif (ULPF) ini telah dibina berdasarkan definisi yang telah dibangunkan bertujuan mengenalpasti tahap literasi pentaksiran formatif guru-guru.



Rajah 1 Literasi Pentaksiran Formatif

2.2 Kesahan Ujian

Secara amnya, kesahan bermaksud sejauh manakah sesuatu instrumen itu mengukur apa yang perlu diukur (Field, 2005). Sebuah instrumen itu dikatakan sah dan relevan apabila skor yang diperolehi daripada alat ukur tersebut boleh dirumuskan apabila membuat keputusan terhadap kumpulan tertentu untuk tujuan yang tertentu. Menurut Taherdoost (2016), terdapat beberapa jenis kesahan seperti kesahan konstruk, kesahan kandungan, dan kesahan kriteria. Dalam konteks penulisan artikel kali ini, penulis menggunakan kesahan kandungan untuk mengesahkan isi kandungan instrumen kajian dalam Ujian Literasi Pentaksiran Formatif (ULPF)

Kesahan kandungan adalah merupakan jenis kesahan yang pertama sekali perlu dilakukan apabila membina instrumen yang baru (Ebel, 1967). Khususnya, kesahan kandungan berfungsi untuk melihat sejauhmana dimensi dan elemen sesuatu konsep itu berjaya didefinisikan (Sekaran & Boigie, 2011). Disamping itu, ianya juga merujuk kepada sejauhmana sesuatu ujian itu merangkumi kepada kandungan yang hendak dinilai (Ariffin, 2008) dan ia juga berfungsi untuk mengesahkan item-item pada ujian yang mewakili setiap konstruk yang hendak diukur (Miller, Lovler & McIntire, 2013). Secara umumnya, kesahan kandungan adalah proses yang melibatkan penilaian instrumen yang baru untuk memastikan ia merangkumi semua item yang penting dan menghilangkan item-item yang tidak perlu atau tidak diinginkan dalam pembinaan konstruk tertentu (Lewis *et al.*, 1995).

Pengesahan kandungan biasanya diperolehi daripada tiga sumber yang utama seperti kajian literatur, wakil dari populasi yang relevan dan pakar (Burns & Grove, 1993). Bagi sumber penilaian pakar, kesahan kandungan perlu dilakukan dengan merujuk kepada semakan pakar bidang subjek atau bidang pengukuran dan penilaian (*psychometrician*). Melalui semakan pakar, setiap kandungan ujian akan dinilai sama ada ianya memadai untuk mewakili sesuatu dimensi atau konstruk (Johnson & Christensen, 2012). Tambahan pada itu, melalui semakan pakar juga persetujuan bahawa item-item yang terkandung dalam skala itu mewakili item-item yang berkaitan dengan konsep yang diukur dapat dicapai dan dipersetujui (Marican, 2009). Penilaian pakar ini adalah penting kerana menurut (Dempsey & Dempsey, 1986; Polit & Hungler, 1991) tiada kaedah objektif yang lengkap untuk menjelaskan kesahan kandungan sesuatu instrumen. Justeru itu, kesahan kandungan bagi kajian ini akan menggunakan pendekatan penilaian pakar seperti dicadangkan oleh Creswell (2012), Popham (2000) dan AERA, APA dan NCME (1999) untuk menilai item yang dibina dalam ULPF agar berupaya untuk mencapai objektif dalam kajian.

2.3 Penilaian Keseragaman Persetujuan Pakar

Terdapat pelbagai kaedah yang boleh digunakan untuk mengumpul bukti kesahaan kandungan. Salah satunya adalah berdasarkan semakan pakar dimana semakan akan dianalisis bagi melihat keseragaman penilaian persetujuan oleh pakar. Kesemua pakar dalam kajian ini adalah dipilih berdasarkan kepakaran mereka dalam bidang pengukuran dan penilaian dan pendidikan khas. Dalam konteks kajian ini, penilaian keseragaman persetujuan pakar dianalisis dan diukur dengan menggunakan teknik Kappa Fleiss.

2.3.1 Kappa Fleiss

Kappa Fleiss (KF) adalah pengukuran persetujuan yang digunakan untuk mengukur tahap persetujuan di antara dua atau lebih penilai (juga dikenali sebagai hakim atau pemerhati) (Fleiss, 1971). Khususnya, KF digunakan apabila, (i) sasaran yang dinilai dipilih secara rawak dari populasi dan bukannya dipilih secara khusus dan (ii) penilai yang menilai sasaran ini tidak unik dan dipilih secara rawak daripada populasi penilai. Pembangunan KF ini adalah bertujuan mengatasi limitasi dan kekurangan kaedah Kappa Cohen (KC). Kaedah KF menyediakan tafsiran perbandingan statistik yang lebih mudah difahami berbanding KC dalam menentukan persetujuan antara penilai (Allen, 2017). Tambahan lagi Fleiss boleh digunakan dengan jumlah penilai yang lebih ramai berbanding Cohen yang hanya boleh dilaksanakan dengan dua penilai sahaja.

Nilai Kappa, *k* boleh ditafsirkan sebagai sejauh mana jumlah persetujuan yang diperhatikan dalam kalangan penilai melebihi apa yang diharapkan jika semua penilai membuat penilaian mereka secara rawak. Landis dan Koch (1977) mengkategorikan skala nilai yang berbeza bagi nilai Kappa (*k*) berdasarkan pandangan persetujuan pakar. Untuk kesemua tujuan, nilai persetujuan lebih daripada 0.75 boleh dikategorikan sebagai “cemerlang” manakala nilai dibawah 0.40 adalah dikategorikan sebagai persetujuan yang “lemah”. Bagi nilai yang mencapai kepada 0.4 sehingga 0.75 pula dikategorikan sebagai “baik” dan boleh dianggap masih boleh diterima (Landis & Koch, 1977). Jadual 1 menunjukkan nilai skala persetujuan KF.

Jadual 1 Nilai Skala Persetujuan Kappa Fleiss

Nilai Kappa	Skala Persetujuan
Bawah 0.4	Persetujuan Lemah
Di antara 0.4 hingga 0.75	Persetujuan Baik
Lebih 0.75	Persetujuan Cemerlang

Meskipun penilaian dilakukan oleh pakar telah ditetapkan (contoh: 3 pakar), namun begitu, penilaian item boleh dilakukan oleh individu berlainan (Fleiss, 1971). Sebagai contoh, item 1 dinilai oleh Pakar 1 dan 2 manakala item 2 boleh dinilai oleh Pakar 1 dan 3. KF juga boleh digunakan untuk menilai data-data berbentuk skala likert, skala ordinal ataupun skala nominal (Glen, 2022). Oleh kerana penilaian dilakukan dengan menggunakan subjek dan skala yang sama, maka nilai pekali kappa biasanya boleh mengendalikan lebih banyak variasi dan pengiraan kebolehppercayaan yang setanding dalam pelbagai bilangan penilai, nilai metrik yang berbeza dan saiz sampel yang tidak sama rata (Arguello *et al.*, 2011). Hal ini adalah jelas menunjukkan KF mempunyai kadar fleksibiliti yang lebih tinggi (Zapf, Castell, Morawietz, 2016). Bahkan, KF juga boleh didapati sebagai amalan yang sering digunakan dalam kajian (Arguello, Diza, Callan & Carterette, 2011). Oleh itu, kajian ini akan menggunakan analisis KF sebagai menentukan nilai persetujuan pakar terhadap instrument Ujian Literasi Pentaksiran Formatif (ULPF).

3.0 METODOLOGI KAJIAN

Kajian ini menggunakan kaedah deskriptif bagi menganalisis semakan pakar. Instrumen yang menggunakan borang pengesahan item. Sebelum semakan pakar dan analisis persetujuan dijalankan, beberapa prosedur pembinaan ujian telah dilakukan termasuk penentuan komponen literasi pentaksiran melalui sorotan kajian (lihat seksyen 2.1.1) dan temubual awal guru pendidikan khas, dan pembinaan jadual spesifikasi ujian seperti dalam Jadual 2

Jadual 2 Jadual Spesifikasi Ujian

	Literasi Pentaksiran Formatif			Bil. Item
	Pengetahuan		Kemahiran	
	Konsep	Prinsip	Kaedah	
Mengingati	6 (1,2,4,7,9,11)	4 (13,19,20,22)		11
Memahami	5 (3,5,6,8,10)	7 (12,14,15,16,17,18,21)		11
Mengaplikasi			13 (23,24,25,26,27,28,29,30,31,32,33,34,35)	13
Bil. Item	11	11	13	35

2.3.2 Persampelan

Prosedur bagi mendapatkan pengesahan pakar dimulakan dengan pemilihan pakar secara persampelan bertujuan. Jumlah pakar yang disarankan oleh pengkaji adalah di antara tiga hingga sepuluh orang (Lynn, 1986), namun begitu jumlah sepuluh orang dilihat tidak relevan (Lynn, 1986) dan menyebabkan persetujuan sukar dicapai. Oleh yang demikian, kajian ini menetapkan seramai tiga orang pakar untuk melakukan pengesahan kandungan item dalam ULPF.

Jadual 3 menunjukkan tiga pengesahan pakar yang telah dipilih berdasarkan pengalaman dan kelayakan mereka. Ketiga-tiga orang pakar ini telah dipilih berdasarkan kepakaran mereka dalam bidang pengukuran dan penilaian dalam Pendidikan dan juga salah seorang pakar juga adalah merupakan pensyarah kanan dalam Pendidikan khas yang mempunyai pengalaman melebihi 10 tahun dalam bidang masing-masing. Surat kebenaran pengesahan yang diperakui daripada pihak fakulti diserahkan kepada setiap bagi tujuan etika penyelidikan.

Jadual 3 Pengesahan Pakar

Pakar	Pengalaman	Subjek	Universiti
A	Melebihi 10 tahun	-Pengukuran dan penilaian -Statistik dalam kajian Pendidikan -Pensyarah kanan Pendidikan khas	Universiti Kebangsaan Malaysia
B	Melebihi 10 tahun	-Pengukuran dan penilaian dalam Pendidikan -Pendidikan Teknik dan Vokasional	Universiti Kebangsaan Malaysia
C	Melebihi 10 tahun	-Pengukuran dan penilaian pendidikan	Universiti Pendidikan Sultan Idris

2.3.3 Instrumen

Borang pengesahan item telah dibina bagi tujuan semakan pakar. Borang pengesahan item mengandungi kriteria seperti;

- i. Kejelasan
- ii. Perkataan
- iii. Hubungan dengan masalah
- iv. Ukuran konstruk literasi pentaksiran formatif (pengetahuan dan kemahiran)

Bagi penilaian kriteria pula, penggunaan skala mata persetujuan 3 hingga 5 telah disarankan oleh Lynn (1986). Kajian ini menggunakan skala 3 mata persetujuan seperti di bawah;

- 1= Tidak diterima (pembetulan keseluruhan diperlukan)
- 2= Bawah Jangkaan (sedikit pembetulan diperlukan)
- 3= Diterima (tiada pembetulan diperlukan)

Selain borang pengesahan instrumen, kesemua pakar juga turut disertakan penerangan tentang objektif kajian, kerangka konsep kajian, definisi operasi dan jadual spesifikasi ujian.

2.3.4 Analisis Data

Melalui persetujuan pakar yang telah dikumpulkan satu analisis perkiraan persetujuan dilakukan dengan menggunakan *Kappa Fleiss Equation* menggunakan *Microsoft Excel*. Jadual 4.2 menjelaskan tiga julat nilai persetujuan yang disarankan oleh Fleiss (1981). Julat nilai persetujuan kurang dari 0.4 bermaksud tidak diterima (tiada persetujuan), julat nilai di antara 0.4 sehingga 0.75 bermaksud baik (persetujuan lemah) manakala julat nilai melebihi 0.75 bermaksud cemerlang (persetujuan baik).

4.0 DAPATAN KAJIAN

Nilai Kappa (k) untuk keseluruhan ULPF adalah dilampirkan dalam Jadual 4.1 di bawah. ULPF mengandungi dua konstruk yang utama pengetahuan dan kemahiran yang merangkumi kepada tiga komponen utama iaitu konsep, prinsip dan kaedah. Setiap komponen juga dianalisis setiap satu untuk mendapatkan tafsiran persetujuan bagi setiap item. Secara keseluruhan Ujian Literasi Pentaksiran Formatif (ULPF) menunjukkan kadar tafsiran persetujuan yang cemerlang dengan memperolehi k sebanyak 0.84 untuk keseluruhan jumlah item. Manakala, setiap komponen pula memberikan nilai k yang berbeza seperti konsep dan prinsip (0.81) dan kaedah (0.89). Kesimpulannya, kesahan kandungan bagi ULPF adalah cemerlang dan item-item yang terkandung dalam ULPF adalah relevan serta sesuai untuk mengukur apa yang sepatutnya diukur.

Jadual 4 Nilai Kappa Bagi Setiap Komponen ULPF

Komponen Ujian Literasi Pentaksiran	Bilangan Item	Nilai Kappa	Tafsiran Persetujuan
Formatif			
PENGETAHUAN			
Konsep	11	0.81	Cemerlang
Prinsip	11	0.81	Cemerlang
KEMAHIRAN			

Kaedah	13	0.89	Cemerlang
Keseluruhan	35	0.84	Cemerlang

Manakala, Jadual 5 pula menunjukkan sekiranya ada berlaku pindaan dan modifikasi yang dilakukan pada mana-mana item yang terdapat didalam ULPF setelah proses kesahan pakar selesai. Jadual 4.2 menunjukkan lapan item yang di nilaikan berbeza-beza oleh pakar. Terdapat sedikit modifikasi yang dilakukan bagi item 18, 28 dan 29.

Jadual 5 Pindaan Item ULPF

Komponen Literasi Pentaksiran Formatif	Item	Persetujuan dan Komen Pakar	Pindaan				
Konsep	1	Seorang pakar menilai dengan nilai “2” (diterima dengan sedikit pembetulan), dengan ulasan <i>“risau redundant”</i> Dua pakar menilai dengan nilai “3” (Diterima tanpa pembetulan)	Tiada pindaan dan perubahan dilakukan				
	2	Seorang pakar menilai dengan nilai “2” (diterima dengan sedikit pembetulan), dengan ulasan <i>“risau redundant”</i> Dua pakar menilai dengan nilai “3” (Diterima tanpa pembetulan)	Tiada pindaan dan perubahan dilakukan				
	4	Seorang pakar menilai dengan nilai “2” (diterima dengan sedikit pembetulan), dengan ulasan <i>“mungkin turutan no 1”</i> Dua pakar menilai dengan nilai “3” (Diterima tanpa pembetulan)	Tiada pindaan dan perubahan dilakukan				
Prinsip	13	Seorang pakar menilai dengan nilai “2” (diterima dengan sedikit pembetulan), dengan ulasan <i>“mungkin turutan”</i> Dua pakar menilai dengan nilai “3” (Diterima tanpa pembetulan)	Tiada pindaan dan perubahan dilakukan				
	18	Dua orang pakar menilai dengan nilai “2” (diterima dengan sedikit pembetulan) dengan ulasan <i>“risau redundant”</i> dan <i>“tiada istilah maklum balas formatif”</i> Seorang pakar menilai dengan nilai “3” (diterima dan tanpa pembetulan)	<i>“Maklum balas yang berkesan”</i> telah ditukar ganti dengan <i>“Maklum balas adalah bertujuan”</i>				
	19	Seorang pakar menilai dengan nilai “2” (diterima dengan sedikit pembetulan) dengan ulasan <i>“tiada istilah maklum balas formatif”</i> Dua orang pakar menilai dengan nilai “3” (diterima tanpa pembetulan)	Tiada pindaan dan perubahan dilakukan				
Kaedah	28	Seorang pakar menilai dengan nilai “2” (diterima dengan sedikit pembetulan) dengan ulasan <i>“ujian prestasi digunakan untuk menilai konstruk- rujuk semula prestasi ini melibatkan apa bagi skop Pendidikan khas”</i> Dua pakar menilai dengan nilai “3” (diterima tanpa pembetulan)	Perubahan telah dilakukan keseluruhan dengan menukar item 28: <i>“Ujian prestasi digunakan untuk menilai konstruk A Sosial B Afektif C Kemahiran”</i> Kepada item 28 yang baru: ” <table border="1" data-bbox="1075 1570 1453 1936"> <thead> <tr> <th><i>Ujian Prestasi</i></th> <th><i>Elemen yang diuji</i></th> </tr> </thead> <tbody> <tr> <td></td> <td> <ul style="list-style-type: none"> <i>Kualiti hasil kerja</i> <i>Kemahiran menggunakan peralatan dan bahan</i> <i>Kepantasan membaca dan menterjemahkan</i> </td> </tr> </tbody> </table>	<i>Ujian Prestasi</i>	<i>Elemen yang diuji</i>		<ul style="list-style-type: none"> <i>Kualiti hasil kerja</i> <i>Kemahiran menggunakan peralatan dan bahan</i> <i>Kepantasan membaca dan menterjemahkan</i>
<i>Ujian Prestasi</i>	<i>Elemen yang diuji</i>						
	<ul style="list-style-type: none"> <i>Kualiti hasil kerja</i> <i>Kemahiran menggunakan peralatan dan bahan</i> <i>Kepantasan membaca dan menterjemahkan</i> 						

			<table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td style="text-align: center;">arahan</td> </tr> <tr> <td style="text-align: center;">• S</td> </tr> </table> <p>28 “Berikut adalah menunjukkan elemen yang diuji dalam Ujian Prestasi, S boleh diwakili sebagai?”</p> <p>A Kebolehan merancang prosedur kerja</p> <p>B Kebolehan dan perbezaan antara individu</p> <p>C Kebolehan untuk menunjukkan perkara yang belum dikuasai”</p>	arahan	• S
	arahan				
• S					
29	Seorang pakar menilai dengan nilai “2” (diterima dengan sedikit pembedaan) dengan ulasan “ <i>mohon semakan tahap kritis dalam kalangan pelajar khas kerana kebanyakan pelajar khas di kv hanya membuat repetitive work dalam survival skills</i> ”	Dua pakar menilai dengan nilai “3” (diterima tanpa pembedaan)	<p>Pernyataan dalam item 29:</p> <p>“-Pertukaran idea secara kritis -Murid berfikir secara kritis -Membantu meningkatkan keyakinan murid ”</p> <p>Ditukar kepada pernyataan yang baru:</p> <p>“-Pertukaran idea secara kreatif -Murid berfikir dengan kreatif -Membantu meningkatkan komunikasi dua hala murid”</p>		

5.0 RUMUSAN

Berdasarkan kesahan pakar dan analisis kandungan yang dijalankan bagi Ujian Literasi Pentaksiran Formatif (ULPF) dapatlah dirumuskan bahawa kesahan kandungan dan item-item yang terkandung di dalam instrumen ini adalah sah, relevan dan membolehkan pengukuran literasi pentaksiran formatif guru diukur dengan kejituan yang diharapkan. Kesahan kandungan pakar adalah penting apabila membina instrumen yang baru. Bahkan, instrumen yang telah berjaya dimurnikan didapati menunjukkan ciri-ciri kesahan yang lebih tepat (Asbullah *et al.*, 2018).

Kesimpulannya, hasil analisis dengan menggunakan pengiraan seperti dalam saranan Fleiss (1981) boleh disimpulkan ULPF mempunyai kadar kesahan yang cemerlang mengikut setiap komponen yang merangkumi kepada konsep, prinsip dan kaedah literasi pentaksiran formatif. (Tambah sedikit kalau ada kajian terdahulu guna Fleiss juga; apa pendapat dorang utk kesahan item dorang mengenai Fleiss) Justeru itu, instrumen ULPF ini boleh dikatakan bersesuaian dengan mempunyai kadar kebolehpercayaan yang dikategorikan sebagai cemerlang serta ianya adalah wajar untuk digunakan untuk para guru yang telah dipilih untuk menjadi responden kajian ini.

Penghargaan

Kami ingin mengucapkan ribuan terima kasih di atas peluang dan kerjasama kepada yang terlibat dalam penerbitan artikel ini. Tidak lupa juga kepada pembaca semua yang sudi meluangkan masa untuk membaca artikel ini. Kami juga merakamkan setinggi-tinggi penghargaan secara khususnya kepada barisan Panel Penyemakkan Artikel Jurnal daripada Universiti Teknologi Malaysia yang profesional dalam memberi bimbingan dan input yang berharga bagi menjayakan penerbitan artikel ini.

Rujukan

- Ab Halim, F., Wan Muda, W.H., & Izam, S. (2019). The Relationship between Employability Skills and Self-Efficacy of Students With Learning Disabilities In Vocational Stream.
- Anizam, M., Y, Mohd Ali., M, Mohd Noor., N. (2020). Penerapan Kemahiran Kebolehgajian Terhadap Murid Berkeperluan Pendidikan Khas. *Online journal for TVET practitioners* 5(1), 36-42
- Arguello, J., Diaz, F., Callan, J., Carterette, B. (2011). A methodology for evaluating aggregated search results. Proceedings of the 33rd European conference on Advances in information retrieval. 141–152 Springer-Verlag, Berlin, Heidelberg.
- Ariffin, S., R. (2008). *Inovasi dalam pengukuran dan penilaian pendidikan*. Bangi: Universiti Kebangsaan Malaysia.

- Asbullah, L. H., Lubis, M. A., & Aladdin, A. (2018). Kesahan Dan Kebolehpercayaan Instrumen Pengetahuan Kolokasi Bahasa Arab IPT (I-KAC IPT) Menggunakan Model Pengukuran Rasch. *Asean Comparative Education Research Journal On Islam And Civilization* (Acer-J). 2(1), 97-106. eISSN2600-769X, Bahagian Pendidikan Khas. (2019). Pentaksiran Bilik Darjah (PBD) Murid Berkeperluan Khas (MBK) Masalah Pembelajaran tahap 1
- Brennan, R. L., & Prediger, D. J. (1981). Coefficient Kappa: some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41 (3), 687-699.
- Brookhart, S. & Lazarus, S. (2017) Formative Assessment for Students with Disabilities. Commissioned by the Council of Chief State School Officers State Collaboratives on Assessing Special Education Students and Formative Assessment, Washington, DC.
- Burns, N., & Grove, S. K. (1993). The practice of nursing research conduct, critique, and utilization (2nd ed.). Philadelphia: WB Saunders Company.
- Butler, D. & Schnellert, L. (2015). "Success for students with learning disabilities: What does self-regulation have to do with it?" In T. Cleary (Ed.), Self-regulated learning interventions with at-risk populations: Academic, mental health, and contextual considerations. Washington, DC: American Psychological Association Press.
- Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence, and kappa. *Journal of Clinical Epidemiology*, 46 (5), 423-429.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 20 (1), 37-46.
- Davis, L.L. (1992). Instrument review: Getting the most from your panel of experts. *Applied Nursing Research*, 5, 194-197.
- Dempsey, P. A., & Dempsey, A. D. (1986). The research process in nursing (2nd ed.). Boston: Jones and Bartlett Publishers. tggu almeer
- Ebel, R. L. (1967). Evaluating content validity. In D. Payne & R. McMorris (Eds.), *Educational and Psychological Measurement: Contributions to theory and Practice*. 85-94. Waltham Blaisdell.
- Feinstein, A. R. & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of the two paradoxes. *Journal of Clinical Epidemiology*, 43 (6), 543- 549.
- Fleiss, J. L., Levin, B., & Paik, M. C. (1981). The measurement of interrater agreement. *Statistical methods for rates and proportions*, 2(212-236), 22-23.
- FIELD, A. P. (2005). *Discovering Statistics Using SPSS*, Sage Publications Inc.
- Friese, S (2020). Measuring Inter-coder Agreement – Why Cohen’s Kappa is not a good choice. Retrieved from <https://atlasti.com/2020/07/12/measuring-intercoder-agreement/> Retrieved date on Retrieved on March 8, 2021
- Glen, S. (2022). "Fleiss' Kappa" From StatisticsHowTo.com: Elementary Statistics for the rest of us! <https://www.statisticshowto.com/fleiss-kappa/> Retrieved date on October 07, 2022
- Jiamu, C. 2001. The great importance of the distinction between declarative and procedural knowledge. *Análise Psicológica* 4(XIX), 559-566. <https://doi.org/10.14417/ap.387>
- Johnston, P., & Wilkinson, K. (2009). Enhancing validity of critical tasks selected for college and university program portfolios. *National Forum of Teacher Education Journal*, 19(3), 1-6.
- Mitchell, R.E. & Karchmer, M. A. (2012). Demographic and achievement characteristics of deaf and hard-of-hearing students. Oxford Handbooks Online. doi: 10.1093/oxfordhb/9780199750986.013.0003
- Kementerian Pelajaran Malaysia. (2013; 2021). Laman web rasmi Kementerian Pelajaran Malaysia.
- Kraemer, H. C. (1979). Ramifications of a population model for k as a coefficient of reliability. *Psychometrika*, 44 (4), 461-472
- Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363-374.
- Lian, L. H., Yew, W. T., & Meng, C. M. (2014). Enhancing Malaysian Teachers' Assessment Literacy. *International Education Studies*; 7(10). ISSN 1913-9020, E-ISSN 1913-9039.
- Lindell, M.K., & Brandt, C.J. (1999). Assessing interrater agreement on the job relevance of a test: A comparison of the CVI, T, rWG(J), and rWG(J) indexes. *Journal of Applied Psychology*, 84, 640-647.
- Lynn, M.R. (1986). Determination and quantification of content validity. *Nursing Research*, 35, 382-385.
- Omar, M, & Ali, D. F. (2019). A Review of vocational education for special needs learners. *Jurnal Persatuan Teknik dan Vokasional Malaysia [Journal of Technic and Vocational Association Malaysia]*, 8, 58-65.
- McHugh, Mary L. (2012). "Interrater reliability: The kappa statistic". *Biochemia Medica*. 22 (3), 276-282. doi:10.11613/bm.2012.031. PMC 3900052. PMID 23092060.
- Mertler, C. A., & Campbell, C. (2005). *Measuring teachers' knowledge & application of Classroom Assessment Concepts: Development of the "Assessment Literacy Inventory"*. American Educational Research Association
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research In Nursing & Health*, 30(4), 459-467.
- Popham, W. J. (2000). *Modern educational measurement: practical guidelines for educational leaders*. Los Angeles: Allyn & Bacon
- Popham, W. J. (2004). Why Assessment Literacy Is Professional Suicide. *Educational Leadership*, 62(1), 82-83.
- Tahir, M, Yassin, Mustafa. (2009) Pendidikan Teknik Dan Vokasional Untuk Pelajar Berkeperluan Khas (Vocational Dan Technical Education For Special Needs Students). *The Asia Pacific Journal of Educators and Education* (formerly known as Journal of Educators and Education), 24 (1), 1-15. ISSN 2289-9057
- Taherdoost, H. (2016). Validity and Reliability of the Research Instrument; How to test the validation of a Questionnaire/Survey in research. *International Journal of Academic Research in Management (IJARM)*. 5(3), 28-36
- Talib, R., Kamsah, M. Z., Ghafar, M. N. A., Zakaria, M. A. Z. M., & Naim, H. A. (2013). T-assess: Assessment literacy for Malaysian teachers. Paper presented at the *International Conference on Assessment for Higher Education Across Domains and Skills, Kuala Lumpur*.
- Sekaran, U., & Bougie, R. (2011). *Research methods for business: a skill building approach* (5th ed.). New Delhi: John Wiley & Sons
- Smith, K., Finney, S., & Fulcher, K. (2019). Connecting assessment practices with curricula and pedagogy via implementation fidelity data. *Assessment and Evaluation in Higher Education*, 44(2), 263-282. <https://doi.org/10.1080/02602938.2018.1496321>
- Smith, P., Cheema, J., Kumi-Yeboah, A., Warrican, S.J., & Alleyne, M. (in press). Language-based literacy differences in the literacy performance of bidialectal youth. *Teachers CollegeRecord*, 120(1)
- Stiggins, R. (1995). Assessment literacy for the 21st Century. *Phi Delta Kappan*, 77, 238
- Sulaiman, T., Kotamjani, S. S., Rahim, S. S. A., & Hakim, M. N. (2020). Malaysian Public University Lecturers' Perceptions and Practices of Formative and Alternative Assessments. *International Journal of Learning, Teaching and Educational Research*, 19(5), 379-394. <https://doi.org/10.26803/IJLTER.19.5.23>
- Waltz, C.F., & Bausell, R.B. (1981). *Nursing research: Design, statistics, and computer analysis*. Philadelphia: F. A. Davis
- Xie, Q. (2013). Agree or disagree? A demonstration of an alternative statistic to Cohens kappa for measuring the extent and reliability of agreement between observer. In *Proceedings of the Federal Committee on Statistical Methodology Research Conference, The Council of Professional Associations on Federal Statistics, Washington, DC, USA, 2013*.
- Zapf, A., Castell, S., Morawietz, L. (2016). Measuring inter-rater reliability for nominal data – which coefficients and confidence intervals are appropriate? *BMC Med Res Methodol* 16, 93. <https://doi.org/10.1186/s12874-016-0200-9>