

**INTERNATIONAL JOURNAL OF
MODERN EDUCATION
(IJMOE)**www.gaexcellence.com/ijmoe**PENILAIAN KUALITI PENGAJARAN GURU DENGAN
PENDEKATAN SIX SIGMA (T²QI-6 Σ) MELALUI
PEMERKASAAN KESAHAN MUKA DAN REFLEKSI
UNTUK TINDAKAN**

*EVALUATION OF TEACHER TEACHING QUALITY WITH SIX SIGMA
APPROACH (T²QI-6 Σ) THROUGH EMPOWERMENT OF FACE VALIDITY
AND REFLECTION FOR ACTION*

Siti Hannah Sabtu¹, Nor Azarif Sofia Hussain², Mohd Effendi Ewan Mohd Matore^{3*}, Siti Mistima Maat⁴

¹Fakulti Pendidikan, Universiti Kebangsaan Malaysia, Malaysia

 sitihannah81phd@gmail.com


 <https://orcid.org/0000-0001-7057-6484>

²Fakulti Pendidikan, Universiti Kebangsaan Malaysia, Malaysia

 azarifsofiaphd@gmail.com


 <https://orcid.org/0009-0004-3297-3621>

³Fakulti Pendidikan, Universiti Kebangsaan Malaysia, Malaysia

 effendi@ukm.edu.my

 <https://orcid.org/0000-0002-6369-8501>

⁴Fakulti Pendidikan, Universiti Kebangsaan Malaysia, Malaysia

 sitimistima@ukm.edu.my

 <https://orcid.org/0000-0002-5507-9081>

*Corresponding Author

Article Info:**Article history:**

Received date: 03.05.2026

Revised date: 21.05.2026

Accepted date: 22.06.2026

Published date: 29.06.2026

Abstrak:

Instrumen Kualiti Pengajaran Guru berasaskan pendekatan *Six Sigma* (T²QI-6 σ) dibangunkan sebagai inovasi penilaian kualiti pengajaran di sekolah rendah dan menengah. Pembangunan ini menyediakan bukti empirikal dalam menyokong Matlamat Pembangunan Lestari (SDG 4) menerusi pemeraksanaan penilaian kualiti pengajaran guru. Kajian ini mengintegrasikan elemen reflektif melalui pendekatan *reflection-for-action* yang meliputi aspek kesedaran, pengetahuan, sikap dan amalan guru. Kesahan muka dinilai menggunakan metodologi *Content Validity Ratio* (CVR) oleh Lawshe (1975). Meskipun CVR lazimnya digunakan untuk kesahan kandungan, kajian ini mengadaptasi pendekatan tersebut bagi menilai kesahan muka secara sistematik dan kuantitatif. Langkah ini membolehkan tahap persetujuan panel penilai terhadap kejelasan bahasa, struktur ayat dan kesesuaian item diukur secara empirikal bagi

To cite this document:

Sabtu, S. H., Hussain, N. A. S., Matore, M. E. E. M., & Maat, S. M. (2026). Penilaian Kualiti Pengajaran Guru Dengan Pendekatan Six Sigma ($T^2Q_i-6\sigma$) Melalui Pemeriksaan Kesahan Muka Dan Refleksi Untuk Tindakan. *International Journal of Modern Education*, 8(30), 741-757.

DOI: 10.35631/IJMOE.830047

mengukuhkan bukti kesahan instrumen. Melalui pensampelan bertujuan, sepuluh penilai (enam guru dan empat pakar bahasa) telah dilibatkan. Instrumen $T^2Q_i-6\sigma$ merangkumi 125 item yang berteraskan kerangka DMAIC Six Sigma (takrif, ukur, analisis, penambahbaikan dan kawalan). Analisis menunjukkan kesahan muka yang kukuh, di mana 113 item melepasi nilai minimum CVR ($N = 10$, $CVR_{kritikal} = 0.62$), manakala 12 item selebihnya telah ditambah baik. Hasilnya, instrumen ini mempunyai potensi tinggi sebagai alat pengukuran psikometrik bagi meningkatkan kualiti pengajaran. Kajian ini mencadangkan penilaian reflektif dengan mengambil kira pandangan pelajar sebagai sumber triangulasi untuk menyokong pembangunan profesional berterusan. $T^2Q_i-6\sigma$ dijangka mampu meningkatkan kompetensi guru serta menyumbang kepada literatur teori dan pelaksanaan Six Sigma dalam konteks pendidikan di Malaysia. Kajian lanjutan disarankan untuk menguji keberkesanan model DMAIC Six Sigma secara lebih meluas dalam konteks pendidikan di Malaysia.

Kata Kunci:

CVR, Instrumen, Kesahan Muka, Kualiti Pengajaran, SDG, Six Sigma

Abstract:

The Teacher Teaching Quality Instrument based on the Six Sigma approach ($T^2Q_i-6\sigma$) was developed as an innovation for evaluating pedagogical quality in primary and secondary schools. This development provides empirical evidence supporting Sustainable Development Goal 4 (SDG 4) through the empowerment of teacher teaching quality assessments. The study integrates reflective elements via a reflection-for-action approach, encompassing teacher awareness, knowledge, attitudes, and practices. Face validity was evaluated using the Content Validity Ratio (CVR) methodology by Lawshe (1975). Although CVR is conventionally utilized for content validity, this study adapts the approach to assess face validity systematically and quantitatively. This step allows the level of agreement among the evaluation panel regarding linguistic clarity, sentence structure and item suitability to be measured empirically to reinforce the instrument's validity evidence. Through purposive sampling, ten evaluators (comprising six teachers and four language experts) were involved. The $T^2Q_i-6\sigma$ instrument consists of 125 items aligned with the Six Sigma DMAIC framework (Define, Measure, Analyze, Improve and Control). Analysis demonstrates robust face validity, where 113 items exceeded the minimum CVR threshold ($N = 10$, $CVR_{critical} = 0.62$), while the remaining 12 items were subsequently refined. Consequently, the instrument exhibits high potential as a psychometric tool for enhancing teaching quality. The study advocates for reflective assessment by incorporating student perspectives as a source of triangulation to support continuous professional development. $T^2Q_i-6\sigma$ is anticipated to bolster teacher competency and contribute to the theoretical and practical literature of Six Sigma within the Malaysian educational context. Future research is recommended to test the broader effectiveness of the Six Sigma DMAIC model within the Malaysian educational landscape.



© The authors (2026). This is an Open Access article distributed under the terms of the Creative Commons Attribution (CC BY NC) (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact ijmoe@gaexcellence.com

Keyword:

CVR, Face Validity, Instrument, SDG, Six Sigma, Teaching Quality

Pengenalan

Proses pengesahan (*validation*) merupakan elemen penting dalam memastikan instrumen yang dibangun benar-benar menghasilkan dapatan kajian yang tepat apabila digunakan dalam penyelidikan sebenar. Ciri psikometrik sesuatu instrumen hanya dapat dijamin sekiranya prosedur kesahan dilaksanakan dengan betul, walaupun instrumen tersebut mempunyai tahap kebolehpercayaan yang tinggi. Hal ini bermakna, proses kesahan perlu dijalankan dengan teliti agar instrumen yang dibangun mempunyai ciri yang sah dipertahankan, bermakna, praktikal, tepat, dan sesuai digunakan (Furr, 2021). Oleh itu, pelaksanaan pengesahan secara sistematik amat penting dalam menghasilkan instrumen yang benar-benar sah untuk penyelidikan.

Kajian lepas menunjukkan bahawa penilaian kesahan muka perlu dilaksanakan bagi memastikan konstruk yang dibangun adalah jelas serta menepati tujuan pengukuran (Azwani Masuwai et al., 2024; Taherdoost, 2016). Pandangan ini turut disokong oleh Cohen et al. (2018) yang menekankan keperluan ujian kesahan muka untuk menilai ketepatan penggunaan bahasa, frasa, dan ayat dalam item yang dibina. Selaras dengan itu, Allen et al. (2023) turut menegaskan bahawa instrumen dengan kesahan muka yang tinggi adalah penting untuk menjamin ketepatan maklumat yang diperoleh serta menghasilkan instrumen yang berkualiti.

Kesahan muka berfungsi memastikan item yang dibina adalah tepat, disusun dengan teratur, dan menggunakan pilihan kata serta penyesuaian linguistik yang sesuai. Kajian oleh Yusoff (2019) menjelaskan bahawa kesahan muka perlu dilakukan bagi mengesahkan ketepatan penggunaan bahasa, tatabahasa, dan kesesuaian tempoh masa menjawab soal selidik. Tambahan pula, penyelidik terkini mencadangkan bahawa kesahan muka kebiasaannya dijalankan selepas kesahan kandungan dan sebelum kajian rintis serta kajian sebenar dilaksanakan (Artino et al., 2014; Kamaruddin & Matore, 2021; Sovey, Osman & Matore, 2022).

Patel dan Desai (2020) pula berpendapat bahawa pertimbangan pakar adalah aspek yang sangat penting dalam kesahan muka bagi memastikan item instrumen bersesuaian serta mudah difahami oleh responden. Allen et al. (2023) menjelaskan bahawa pengesahan kesahan muka boleh dijalankan oleh pakar terpilih melalui pendekatan kualitatif mahupun kuantitatif. Dalam pendekatan kualitatif, biasanya temu bual individu atau berkumpulan digunakan, di mana responden diberikan item untuk semakan sebelum diminta memberikan penilaian (Sadeghi et al., 2021). Data yang diperoleh kemudiannya dianalisis melalui analisis tematik bagi mengenal pasti kecenderungan responden dalam memberikan penilaian. Sebaliknya, pendekatan kuantitatif menilai kesahan muka dengan melibatkan pakar atau responden ujian melalui kejelasan dan tahap kefahaman terhadap ayat bagi setiap konstruk (Dalawi et al., 2023).

Kajian ini mempunyai beberapa elemen kebaruaran. Pertama, instrumen T²Qi-6σ dibangun berdasarkan model DMAIC Six Sigma yang masih jarang diaplikasikan dalam penilaian kualiti pengajaran guru di sekolah. Kedua, kajian ini mengadaptasikan analisis CVR bagi tujuan kesahan muka, sedangkan pendekatan tersebut lazimnya digunakan dalam penilaian kesahan kandungan. Ketiga, kajian ini mengintegrasikan perspektif reflection-for-action dalam penilaian kualiti pengajaran guru. Keempat, kajian ini mencadangkan dimensi kesedaran, pengetahuan, sikap dan amalan sebagai prospek baharu dalam penilaian kualiti pengajaran guru pada masa hadapan.

Kerangka Teori

Bahagian tinjauan literatur ini menghuraikan dan menekankan kajian-kajian utama yang berkaitan serta mempunyai perkaitan langsung dengan bidang penyelidikan semasa

Model Kesahan Lawshe

Kajian terkini banyak menggunakan analisis Content Validity Ratio (CVR) untuk menilai kesahan kandungan instrumen (Romero et al., 2023; Yaw & Matore, 2024). Walau bagaimanapun, aplikasi CVR dalam menilai kesahan muka masih sangat terhad. Meskipun terdapat kaedah kuantitatif lain seperti Cohen's Kappa dan Face Validity Index (FVI) (Azwani Masuwai et al., 2024; Dalawi et al., 2023), ramai penyelidik lebih memilih CVR bagi menilai kesahan muka item instrumen pengajaran guru menggunakan pendekatan Six Sigma (T²Qi-6σ). Pemilihan ini berpunca daripada kelebihan CVR yang bersifat telus, mudah digunakan, mempunyai pengiraan komputer yang ringkas, dilengkapi jadual nilai kritikal sebagai rujukan, serta memberi penekanan terhadap persetujuan pakar tentang kepentingan sesuatu item (Lawshe, 1975).

Lawshe (1975) memperkenalkan kaedah analisis CVR pada tahun 1974 dalam Content Validity Conference anjuran Society for Industrial and Organizational Psychology (SIOP). Kaedah ini kemudiannya menjadi rujukan penting untuk penyelidik membuat keputusan sama ada mengekalkan atau menggugurkan item instrumen melalui pendekatan kuantitatif. Malah, CVR turut menekankan keutamaan sesuatu item, iaitu sama ada wajar dikekalkan atau digugurkan daripada instrumen yang dibina (Gilbert & Prion, 2016). Kajian oleh Divayana et al. (2020) turut menegaskan bahawa penggunaan CVR berupaya menapis item secara empirikal dan kuantitatif bagi memastikan setiap item benar-benar mewakili domain konstruk yang dibangunkan.

Lawshe (1975) menegaskan bahawa langkah paling penting dalam penggunaan CVR adalah setiap panel penilai perlu menilai tahap kepentingan item berdasarkan tiga kategori: (i) penting, (ii) berguna tetapi tidak penting, dan (iii) tidak diperlukan. Seperti yang dijelaskan oleh Creswell dan Guetterman (2019), panel penilai memainkan peranan utama dalam menentukan sama ada sesuatu item pengukuran harus dikekalkan. Kesahan muka kemudiannya ditentukan berdasarkan nisbah CVR bagi sesuatu item, iaitu dengan membandingkan bilangan pakar yang menilai item sebagai penting dengan jumlah keseluruhan panel pakar yang terlibat.

Hasil pengiraan CVR menghasilkan nilai antara -1 hingga +1. Nilai CVR bersamaan nilai +1 menunjukkan semua panel menilai item tersebut sebagai penting untuk kesahan muka. Sebaliknya, nilai CVR kurang daripada 0 menunjukkan kurang daripada separuh panel menilai item sebagai penting, manakala nilai CVR bersamaan nilai 0 menggambarkan wujud perbezaan

dalam kalangan panel, iaitu sebahagian menganggap item penting dan sebahagian lagi tidak. Nilai CVR melebihi nilai 0-pula menunjukkan lebih daripada separuh panel menilai item sebagai penting. Lawshe (1975) menyatakan bahawa semakin tinggi nilai CVR, semakin tinggi tahap kesahan sesuatu item. Jika CVR bersamaan dengan 1, ini bermakna terdapat persetujuan penuh dalam kalangan panel pakar terhadap kepentingan item tersebut, sekali gus membuktikan kesahan muka yang tinggi.

Pendekatan Six Sigma Dalam Pendidikan

Dalam era globalisasi masa kini, usaha bersepadu dijalankan di seluruh dunia untuk meningkatkan mutu pendidikan di pelbagai negara. Salah satu cabaran utama ialah melahirkan pendidik yang benar-benar berkualiti, dan hal ini menjadi keutamaan kepada pembuat dasar pendidikan di setiap negara (Blomeke et al., 2022). Walaupun begitu, masih terdapat ramai guru yang berdepan kesukaran dari segi amalan pedagogi serta penyesuaian terhadap kaedah dan strategi pengajaran moden (Almutairi & Shraid, 2021; Blomeke & Olsen, 2019; Deti Rostini et al., 2022). Tambahan pula, kaedah penilaian kualiti pengajaran guru yang digunakan pada masa ini masih belum benar-benar berkesan dalam meningkatkan mutu pengajaran dan pembelajaran di bilik darjah (Darling-Hammond et al., 2015; Gerritsen et al., 2016). Secara ringkasnya, penilaian sedia ada lebih menekankan aspek pemarkahan terhadap amalan pengajaran guru, namun kurang memberi penekanan kepada usaha mengenal pasti kelemahan serta menggalakkan strategi PdP yang lebih berkesan (Callahan & Sadeghi, 2015).

Sehubungan itu, kebanyakan negara kini sedang berusaha memastikan sistem pendidikan mereka setanding dengan negara-negara terbaik di dunia. Keberkesanan pendekatan Six Sigma dalam sektor industri berorientasikan keuntungan telah menarik minat para penyelidik untuk menerokai potensi pelaksanaannya dalam bidang pendidikan (Jayamohan & Bhasi, 2021; LeMahieu et al., 2017). Kajian oleh Abdulla dan Kavilal (2022) melaporkan bahawa Six Sigma telah berjaya diintegrasikan ke dalam pelbagai sektor pendidikan sejak tahun 2000. Malahan, tinjauan literatur sistematik oleh Sabtu dan Matore (2024) merumuskan bahawa aplikasi Six Sigma dalam pendidikan berpotensi meningkatkan aspek kurikulum, infrastruktur, prestasi pembelajaran murid, pencapaian akademik, komuniti pembelajaran, serta keberkesanan proses pengajaran dan pembelajaran.

Walaupun begitu, himpunan kajian mengenai Six Sigma dalam pendidikan masih agak terhad. Beberapa penyelidikan menunjukkan bahawa pendekatan ini lebih banyak diaplikasikan dalam institusi pengajian tinggi (IPT) berbanding di peringkat sekolah (Al Kuwaiti & Subbarayalu, 2015; Wang, 2022; Yu & Ueng, 2012). Misalnya, Six Sigma digunakan untuk mempertingkatkan mutu pengajaran dan pembelajaran pensyarah di IPT (Wang, 2022). Namun, hasil kajian oleh Sabtu dan Matore (2024) mendapati bahawa penyelidikan tentang pendekatan Six Sigma dalam menilai dan mengukur kualiti pengajaran guru di sekolah masih kurang dijalankan. Malah, Sabtu dan Matore (2024) turut menegaskan bahawa sehingga kini, aplikasi Six Sigma dalam konteks pendidikan Malaysia belum pernah dikaji secara mendalam.

Model DMAIC Six Sigma

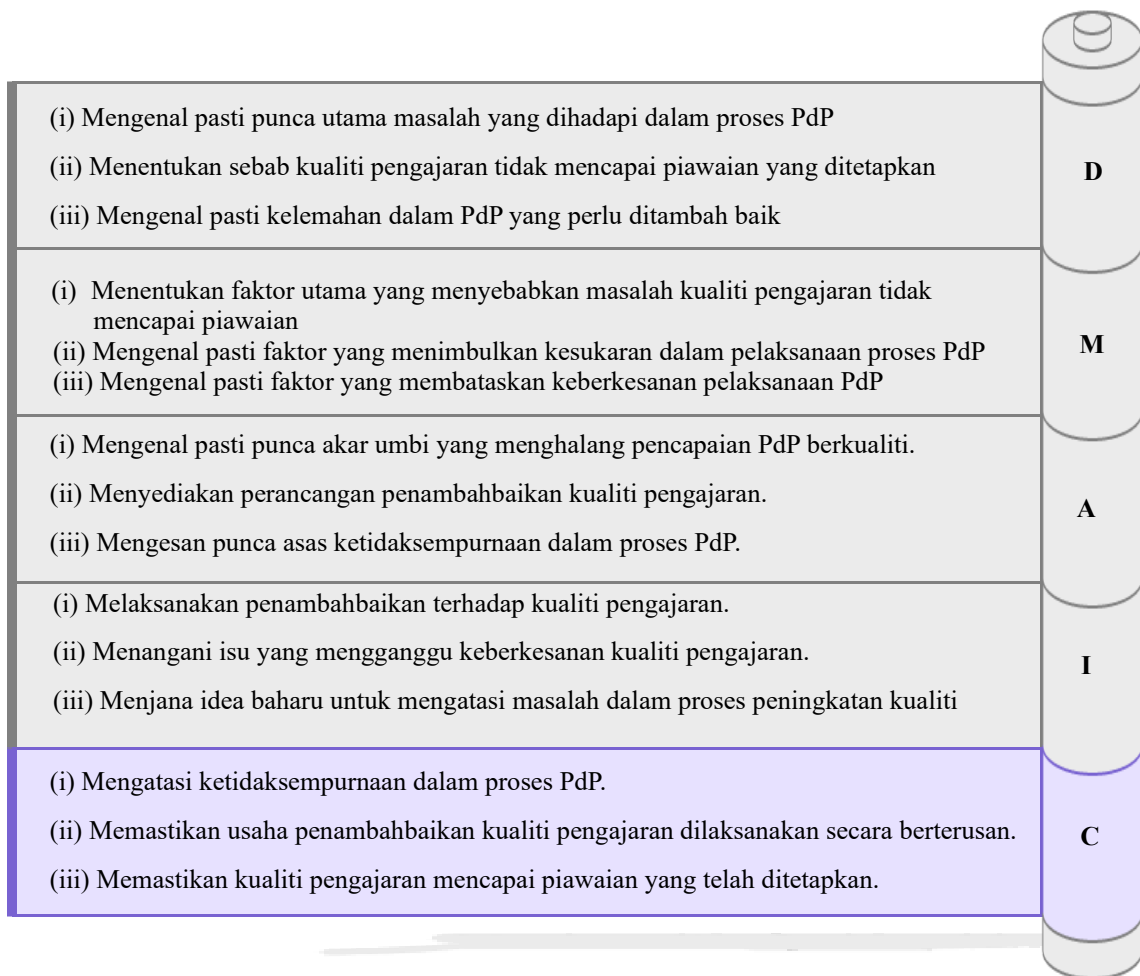
Model DMAIC Six Sigma merupakan satu kitaran penambahbaikan berterusan yang berpaksikan data, bertujuan mengesan kelemahan dan ketidakcekapan, khususnya yang boleh menjejaskan kualiti produk atau perkhidmatan (Costa et al., 2021; Mittal et al., 2023). Pyzdek dan Keller (2018) dalam karya mereka turut menjelaskan bahawa model DMAIC digunakan

bukan sahaja untuk memperbaiki proses sedia ada mengikut piawaian tertentu, tetapi juga bagi mengenal pasti peluang penambahbaikan tambahan. Terdapat bukti yang kukuh dalam literatur bahawa model DMAIC telah digunakan secara meluas oleh penyelidik luar negara sebagai strategi untuk mempertingkatkan kualiti dalam pelbagai bidang pendidikan (Jayamohan & Bhasi, 2021; Pal & Ghosh, 2022).

Model DMAIC Six Sigma dilihat sebagai pendekatan yang sistematik untuk menilai, menambah baik, serta meningkatkan mutu pengajaran guru demi mencapai tahap kecemerlangan yang memenuhi kepuasan pelanggan utama, iaitu para pelajar (Wang, 2022; Yu & Ueng, 2012). Model ini merupakan akronim kepada lima fasa penting iaitu: (i) takrif (define), (ii) ukur (measure), (iii) analisis (analyze), (iv) penambahbaikan (improve), dan (v) kawalan (control) (Bloom, 2018; Pyzdek & Keller, 2018). Setiap fasa dalam DMAIC perlu dilaksanakan secara berurutan dalam satu kitaran penambahbaikan bagi memastikan hasil yang optimum.

Kajian ini berasaskan aplikasi model DMAIC Six Sigma yang menjadi asas kepada pembangunan instrumen $T^2Q_i-6\sigma$. Melalui model ini, pendidik dapat menilai kualiti pengajaran dan seterusnya memperbaiki pengurusan proses PdP berdasarkan analisis data yang diperoleh. Rajah aliran proses model DMAIC turut digunakan untuk menggambarkan bagaimana instrumen $T^2Q_i-6\sigma$ dibangunkan, meliputi lima langkah iaitu takrif, ukur, analisis, penambahbaikan dan kawalan.

Pembinaan $T^2Q_i-6\sigma$ mewakili lima konstruk yang berpaksikan kitaran DMAIC Six Sigma. Secara keseluruhannya, 125 item telah dibangunkan dengan menggabungkan dua aspek utama: kualiti pengajaran guru serta indikator dalam model DMAIC Six Sigma. Memandangkan $T^2Q_i-6\sigma$ merupakan instrumen baharu, kerangka pembinaan instrumen oleh Miller dan Lovler (2020) digunakan sebagai panduan. Kerangka tersebut terdiri daripada sembilan langkah, bermula dengan pengumpulan maklumat daripada pelbagai sumber untuk membentuk kerangka konseptual dan atribut instrumen. Maklumat tersebut kemudiannya dijadikan asas konseptualisasi dan operasionalisasi sebelum pembinaan item dilakukan. Rajah 1 dalam artikel asal menunjukkan proses operasionalisasi lima konstruk yang diperoleh daripada model DMAIC Six Sigma.



Rajah 1: Pengoperasian Konstruk DMAIC Six Sigma

Metodologi

Bahagian metodologi sesebuah artikel menerangkan prosedur yang digunakan dalam pelaksanaan kajian, merangkumi jenis kajian, pemilihan sampel, kaedah pengumpulan serta analisis data, pertimbangan etika, dan batasan kajian. Huraian yang jelas dan teliti amat penting bagi menjamin kebolehpercayaan serta kebolehulangan (replicability) dapatan kajian, selain menyediakan asas yang kukuh untuk pentafsiran dan pengeneralisasian hasil kajian.

Pemilihan Panel Pakar

Bougie dan Sekaran (2020) menegaskan bahawa bahasa yang digunakan dalam sesuatu instrumen yang dibina perlu hampir dengan kefahaman responden. Sehubungan itu, seramai enam orang calon ujian serta empat orang pakar bahasa telah dilibatkan dalam kajian ini bagi menilai kesahan muka item $T^2Q_i-6\sigma$. Jumlah keseluruhan panel penilai bagi kesahan muka $T^2Q_i-6\sigma$ adalah seramai sepuluh orang. Calon ujian terdiri daripada tiga orang guru sekolah rendah dan tiga orang guru sekolah menengah yang mewakili kaum Melayu, Cina dan India. Mereka dilantik untuk meneliti kejelasan ayat bagi setiap item $T^2Q_i-6\sigma$, termasuk dari aspek susunan, pemilihan perkataan, serta penyesuaian bahasa.

Empat orang pakar dilantik sebagai panel kesahan muka, kesemuanya berkelulusan PhD dan berpengalaman lebih 15 tahun dalam bidang Bahasa Melayu. Dua daripadanya berstatus Munsyi Dewan (Bahasa) dan Munsyi Muda Bahasa di Institut Pendidikan Guru (IPG), manakala dua lagi merupakan pakar Pendidikan Bahasa Melayu dari Universiti Utara Malaysia (UUM) dan Universiti Pendidikan Sultan Idris (UPSI). Pemilihan mereka dibuat berdasarkan kepakaran akademik serta pengalaman luas yang membolehkan penilaian instrumen dilakukan dengan lebih tepat dan berautoriti.

Penyediaan Soal Selidik

Instrumen T²Qi-6σ terdiri daripada tiga bahagian, iaitu Bahagian A, B, dan C. Bahagian A mengandungi soal selidik kesahan pakar yang merangkumi lima konstruk utama: (i) takrif, (ii) ukur, (iii) analisis, (iv) penambahbaikan dan (v) kawalan. Setiap konstruk mempunyai 25 item, menjadikan keseluruhan sebanyak 125 item yang perlu dinilai oleh peserta ujian serta panel pakar yang dilantik. Bahagian B pula menumpukan kepada profil demografi panel, yang meliputi jawatan, alamat organisasi, tempoh perkhidmatan, dan bidang kepakaran. Seterusnya, Bahagian C mengandungi pernyataan pengesahan oleh pakar yang wajib dilengkapkan bagi mengesahkan penilaian terhadap kesemua 125 item T²Qi-6σ. Peserta ujian dan panel pakar turut dikehendaki menyatakan sama ada kriteria instrumen ini memenuhi piawaian yang ditetapkan. Selain itu, disediakan satu ruangan khas untuk membolehkan peserta ujian dan pakar memberikan kesimpulan mereka mengenai tahap kesesuaian instrumen T²Qi-6σ.

Pengumpulan Data

Patel dan Desai (2020) mencadangkan penggunaan kaedah pengumpulan data secara bersemuka dan dalam talian. Dalam kajian ini, penilai yang dikenal pasti bagi menilai kesahan muka T²Qi-6σ telah dihubungi melalui e-mel untuk membentuk panel penilai. Setelah panel tersebut bersetuju menerima pelantikan, borang semakan kesahan muka diedarkan secara dalam talian, dengan tempoh masa tiga minggu diberikan untuk melaksanakan proses semakan kesahan muka bagi item-item T²Qi-6σ.

Analisis Keputusan

Kesemua 125 item T²Qi-6σ dibina dan dinilai oleh panel penilai berdasarkan kriteria yang digariskan oleh Oluwatayo (2012). Kriteria tersebut merangkumi penggunaan tatabahasa yang tepat, kesesuaian bahasa, ejaan yang betul, serta struktur ayat yang teratur. Seterusnya, setiap pakar meneliti semua item dan menilai kepentingannya dengan menggunakan skala penilaian Lawshe (1975). Skala ini terdiri daripada tiga tahap utama, iaitu skala 1 mewakili “sangat penting”, skala 2 merujuk kepada “berguna tetapi tidak penting”, manakala skala 3 menunjukkan “tidak perlu”. Skala ini digunakan oleh panel penilai bagi menentukan tahap kepentingan setiap item dalam instrumen yang dibangunkan. Selain itu, turut disediakan ruangan khas untuk komen, bagi membolehkan panel penilai memberikan cadangan serta ulasan yang dapat memperkukuh penggunaan bahasa dan memperjelas maksud bagi setiap item yang dinilai. Selepas memperoleh persetujuan daripada panel penilai, kesahan muka bagi item T²Qi-6σ ditentukan dengan menggunakan Persamaan (1) berikut:

$$CVR = \frac{n e^{-\left(\frac{N}{2}\right)}}{(N/2)} \dots\dots\dots \text{Persamaan (1)}$$

* Rujukan:

CVR	Nilai nisbah kesahan muka item
n_e	Bilangan panel pakar yang menilai item sebagai penting
N	Jumlah keseluruhan panel pakar terlibat

Dapatan dan Perbincangan

Kajian ini menunjukkan bahawa kesahan muka telah dinilai oleh panel pakar yang terdiri daripada enam orang calon ujian dan empat orang pakar bahasa yang dipilih berdasarkan kepakaran masing-masing. Enam orang calon ujian tersebut merangkumi tiga orang guru sekolah rendah dan tiga orang guru sekolah menengah daripada pelbagai kaum. Mereka dipilih secara rawak dan sukarela untuk menilai instrumen $T^2Q_i-6\sigma$ berdasarkan beberapa kriteria seperti susun atur item, format instrumen, kejelasan ayat, serta saiz dan jenis fon yang digunakan.

Dapatan Kesahan Muka Instrumen $T^2Q_i-6\sigma$

Hasil semakan mendapati bahasa dan perkataan dalam setiap item difahami dengan baik oleh guru-guru yang terlibat. Mereka juga dapat memahami arahan dengan jelas dan sesuai dengan tahap keupayaan masing-masing. Justeru, dapat disimpulkan bahawa keseluruhan item serta arahan dalam instrumen $T^2Q_i-6\sigma$ difahami dengan baik oleh guru. Selain itu, empat orang pakar bahasa berpengalaman luas dalam bidang bahasa Melayu turut melaksanakan semakan yang menumpukan kepada aspek struktur ayat, kesesuaian kosa kata dan penggunaan bahasa, serta ketepatan ejaan.

Dapatan persetujuan daripada enam orang calon ujian dan empat orang pakar bahasa yang menilai kesahan muka ke atas 125 item $T^2Q_i-6\sigma$ dianalisis menggunakan Microsoft Excel 2021. Persetujuan daripada sepuluh orang ahli panel ini diukur menggunakan indeks CVR. Berdasarkan penilaian panel, majoriti pakar menegaskan bahawa kesemua 125 item $T^2Q_i-6\sigma$ mempunyai tahap kepentingan yang tinggi dari segi kesahan muka. Analisis turut mengesahkan bahawa tiada item mencatatkan nilai CVR kurang daripada sifar, sekali gus menunjukkan kesemua panel penilai bersetuju bahawa 125 item $T^2Q_i-6\sigma$ memenuhi keperluan kesahan muka. Nilai kritikal CVR yang ditetapkan ialah $[N=10; CVR \text{ critical} = 0.62]$ Lawshe, (1975). Bagi memastikan setiap item $T^2Q_i-6\sigma$ memiliki kesahan muka, setiap item perlu mencapai sekurang-kurangnya nilai minimum 0.62 selaras dengan nilai kritikal tersebut.

Oleh itu, mana-mana item yang memperoleh nilai CVR lebih rendah daripada nilai ini dianggap mempunyai masalah yang signifikan daripada perspektif pakar. Penyelidik perlu mengambil tindakan dengan meneliti semula item bermasalah tersebut sebelum memutuskan sama ada ia perlu ditambah baik atau digugurkan daripada instrumen $T^2Q_i-6\sigma$. Analisis statistik menunjukkan bahawa terdapat 12 item $T^2Q_i-6\sigma$ yang berada di bawah nilai yang ditetapkan. Jadual 1 memperlihatkan taburan nilai CVR yang diperoleh bagi setiap daripada 125 item $T^2Q_i-6\sigma$. Secara keseluruhannya, sebanyak 113 item $T^2Q_i-6\sigma$ telah mencapai nilai maksimum CVR iaitu 1.000, yang menunjukkan bahawa semua panel penilai sebulat suara menganggap item-item tersebut sangat penting.

Jadual 1: Taburan Nilai CVR

Konstruk	Bilangan Item Mencapai Nilai Kritikal (CVRkritikal ≥ 0.62)	Bilangan Item Tidak Mencapai Nilai Kritikal (CVRkritikal < 0.62)	Item Tidak Mencapai Nilai Kritikal
Takrif	23	2	D21, D24
Ukur	23	2	M9, M12
Analisis	23	2	A6, A24
Penambahbaikan	21	4	I15, I17, I18, 122
Kawalan	23	2	C2, C19

Jadual 2 menyenaraikan item-item yang tidak mencapai nilai kritikal CVR iaitu 0.62. Hasil analisis menunjukkan bahawa 750mpirica item T²Qi-6 σ memperoleh nilai CVR sebanyak 0.6, yang menggambarkan terdapat seorang penilai tidak bersetuju dengan kepentingan item tersebut dalam T²Qi-6 σ . Sementara itu, tiga item yang lain dengan nilai CVR 0.4 menunjukkan bahawa terdapat dua orang penilai tidak bersetuju terhadap item yang dibangunkan dalam T²Qi-6 σ .

Dapatan kajian ini membuktikan bahawa item-item T²Qi-6 σ telah disahkan dari aspek kesahan muka dengan kualiti dan keteguhan yang tinggi, selaras dengan prosedur yang telah ditetapkan. Menariknya, variasi dapatan kajian serta maklum balas daripada panel penilai menunjukkan bahawa proses kesahan muka terhadap item T²Qi-6 σ dilaksanakan dengan teliti dan menyeluruh. Bukti kukuh yang diperoleh dalam kajian ini juga menegaskan bahawa proses kesahan muka yang dilaksanakan secara rapi mampu menghasilkan nilai kesahan muka yang lebih konkrit berbanding kaedah yang digunakan oleh penyelidik terdahulu. Keputusan positif ini turut menyokong pandangan bahawa penggunaan kaedah analisis CVR memberikan tafsiran yang lebih meyakinkan dengan disokong oleh bukti 750mpirical yang kukuh.

Jadual 2: Senarai Item yang Tidak Memenuhi Nilai Kritikal [CVR = 0.62]

Konstruk	Item	Nilai CVR	Isu	Cadangan Penambahbaikan
<i>Define</i>	D21	0.4	Pembetulan ayat	Aktiviti di dalam bilik darjah berpusatkan murid
	D24	0.6	Penggantian perkataan	Gantikan " <i>persuade</i> " dengan " <i>encourage</i> "
<i>Measure</i>	M9	0.6	Penggantian perkataan	Gantikan " <i>in</i> " dengan " <i>related to</i> "
	M12	0.6	Pembetulan ayat	Menghadapi masalah disiplin murid di dalam bilik darjah
<i>Analyze</i>	A6	0.6	Penggantian perkataan	Gantikan " <i>issue</i> " dengan " <i>declare</i> "
	A24	0.4	Penggantian perkataan	Gantikan " <i>fixes</i> " dengan " <i>fix</i> "
<i>Improve</i>	I15	0.6	Penggantian perkataan	Gantikan " <i>improve</i> " dengan " <i>improvement</i> "
	I17	0.6	Pembetulan ayat	Komunikasi dalam kalangan murid

	I18	0.4	Penggantian perkataan	Gantikan " <i>issue</i> " dengan " <i>declare</i> "
	I22*	0.6	Penggantian perkataan	Gantikan " <i>issue</i> " dengan " <i>declare</i> "
Control	C2	0.6	Penggantian perkataan	Gantikan " <i>increase</i> " dengan "enhance"
	C19	0.6	Gugur perkataan	Gugurkan perkataan ' <i>for</i> '

Kaedah analisis CVR yang digunakan dalam kajian ini turut menyediakan jaminan kualiti dalam proses kesahan muka melalui kepelbagaian panel penilai yang terlibat (Romero Jeldres et al., 2023). Penglibatan enam orang calon ujian telah mengukuhkan lagi kebolegunaan item-item T²Qi-6 σ . Semua maklum balas yang diterima telah diteliti dan diberikan perhatian oleh penyelidik. Secara keseluruhan, ulasan yang dikemukakan oleh empat orang pakar lebih bersifat kritikan membina, dengan tumpuan khusus kepada aspek struktur ayat dan penggunaan perkataan dalam item T²Qi-6 σ . Tiada kelemahan utama dikesan seperti item yang dibina tidak berkaitan atau gagal mencapai maksud yang diinginkan. Oleh itu, kepelbagaian panel penilai yang dilibatkan dalam proses kesahan muka ini secara tidak langsung telah menjamin kualiti item-item T²Qi-6 σ serta memberikan keyakinan bahawa instrumen ini adalah sah, jelas, dan mencerminkan pengukuran yang dimaksudkan.

Analisis CVR juga telah memperluas pengetahuan penyelidik tentang prosedur untuk menilai kesahan item-item T²Qi-6 σ . Penekanan terhadap kepentingan setiap item melalui kaedah CVR membantu penyelidik menapis item secara empirikal bagi memastikan bahawa item benar-benar mewakili konstruk yang diukur. Divayana et al. (2020) menjelaskan bahawa skor CVR yang tinggi menunjukkan kestabilan dan kekukuhan instrumen kerana ia mengukur aspek yang sepatutnya diukur dari perspektif pakar. Dapatan analisis menunjukkan bahawa 113 item T²Qi-6 σ memperoleh nilai maksimum CVR iaitu 1.00, yang membuktikan kesepakatan penuh panel penilai bahawa 113 item tersebut adalah sangat penting.

Walau bagaimanapun, terdapat sebanyak 12 item yang tidak mencapai nilai minimum (CVR_{kritikal} = 0.62). Item D24, M9, M12, A6, I15, I17, I22, C3, dan C19 mencatatkan nilai CVR 0.6, yang bermaksud seorang daripada panel penilai tidak bersetuju dengan kepentingan item-item tersebut. Manakala, item D21, A24, dan I18 memperoleh nilai CVR 0.4, menandakan terdapat dua orang panel penilai yang tidak bersetuju. Dapatan daripada maklum balas panel pakar menunjukkan bahawa kesemua 12 item ini berkaitan dengan penggunaan perkataan yang kurang tepat serta struktur ayat yang memerlukan penambahbaikan. Lawshe (1975) menyarankan dua pendekatan untuk item yang tidak mencapai nilai minimum, iaitu sama ada item tersebut ditambah baik atau digugurkan terus daripada instrumen. Selaras dengan saranan Lawshe (1975), penyelidik telah mengambil langkah memperhalusi dan menambah baik kesemua 12 item tersebut. Dengan kata lain, jumlah keseluruhan 125 item T²Qi-6 σ yang dinilai kekal tanpa sebarang item digugurkan dalam fasa penilaian kesahan muka.

Secara keseluruhan, dapatan menunjukkan bahawa panel penilai bersetuju bahawa instrumen T²Qi-6 σ adalah sesuai serta relevan untuk digunakan dalam sistem pendidikan di Malaysia. Analisis CVR turut membuktikan bahawa instrumen ini memiliki tahap kesahan muka yang tinggi dan sesuai digunakan. Persetujuan panel penilai terhadap kesahan muka jelas menunjukkan bahawa T²Qi-6 σ merupakan instrumen yang sesuai dan signifikan dalam usaha meningkatkan kualiti pengajaran guru berasaskan model Six Sigma DMAIC. Oleh itu, bagi

mempertingkatkan kualiti pengajaran guru, adalah sangat penting untuk mengenal pasti kekuatan dan kelemahan kualiti pengajaran melalui penggunaan instrumen T²Qi-6σ, agar penambahbaikan dapat dilakukan demi mencapai kecemerlangan pengajaran. Kajian lanjutan juga perlu dijalankan bagi menentukan sama ada instrumen T²Qi-6σ benar-benar sesuai dan relevan digunakan dalam konteks peningkatan kualiti pengajaran guru dengan aplikasi Six Sigma di Malaysia.

Reflection-for-Action dalam Penilaian Kualiti Pengajaran Guru

Salah satu objektif kajian ini juga ialah meneroka pandangan tentang penilaian kualiti pengajaran guru secara reflektif bagi refleksi untuk tindakan (*reflection-for-action*). Sebagai pengkaji yang meneliti isu ini, terdapat dua refleksi untuk tindakan yang boleh diusulkan. Pertama, saya menyedari bahawa keberkesanan penilaian kualiti pengajaran tidak hanya terletak pada reka bentuk instrumen semata-mata, tetapi turut menuntut penghayatan guru terhadap makna penilaian itu sendiri. Hal ini selari dengan pandangan Burhan-Horasanlı dan Hart (2024) yang menegaskan bahawa *reflection-for-action* mampu memperkasa guru sebagai agen transformasi kolaboratif, kerana ia menuntut guru melihat penilaian bukan sekadar ukuran prestasi, sebaliknya sebagai dorongan untuk membina budaya pembelajaran yang kolektif. Dalam konteks ini, penilaian instrumen seperti T²Qi-6σ dapat digunakan bukan sahaja untuk mengenal pasti kelemahan, tetapi juga sebagai wacana reflektif bagi guru memperkaya amalan pedagogi mereka. Kedua, saya mendapati bahawa penilaian kualiti pengajaran yang sahih akan lebih bermakna sekiranya ia mengambil kira perspektif pelajar sebagai penerima langsung pengajaran. Kajian oleh Bas Senden, Nilsen dan Teig (2023) menunjukkan bahawa maklum balas pelajar boleh memberikan gambaran yang sahih tentang kualiti pengajaran, khususnya dalam aspek pengurusan bilik darjah, pengaktifan kognitif, dan iklim sokongan. Refleksi untuk tindakan di sini ialah kepentingan menggabungkan data penilaian guru dengan suara pelajar sebagai sumber triangulasi. Hal ini bukan sahaja meningkatkan kebolehpercayaan penilaian, tetapi turut menanam kesedaran dalam kalangan guru bahawa keberkesanan PdP diukur melalui pengalaman pembelajaran yang dihayati murid.

Prospek Penilaian Kualiti Pengajaran Guru

Hasil kajian ini turut memberikan implikasi penting terhadap prospek penilaian kualiti pengajaran guru, khususnya dari aspek kesedaran, pengetahuan, sikap dan amalan. Analisis menunjukkan bahawa penilaian kualiti pengajaran tidak seharusnya hanya bergantung kepada keberkesanan instrumen semata-mata, tetapi juga perlu dilihat sebagai satu usaha membina budaya profesional yang menekankan kesedaran guru terhadap peranan mereka dalam memastikan keberhasilan pembelajaran. Hal ini seiring dengan dapatan Al Masaeid et al. (2023) yang mendapati bahawa pelaksanaan metodologi Six Sigma dalam kalangan pengetua sekolah di Jordan masih berada pada tahap sederhana, dengan kelemahan khusus pada domain kawalan pasca-proses. Kekurangan ini menunjukkan bahawa kesedaran dan sikap terhadap kesinambungan kawalan kualiti masih perlu diperkukuh, sekaligus memberi gambaran bahawa aspek refleksi berterusan dalam amalan pengajaran adalah prospek yang wajar diberikan perhatian dalam konteks Malaysia. Kajian Jin et al. (2024) turut menegaskan bahawa keberkesanan Lean Six Sigma dalam pelbagai sektor banyak bergantung kepada keupayaan membina budaya organisasi yang sistematik, integratif dan berterusan, yang juga relevan diaplikasikan dalam konteks pendidikan.

Selain itu, pengetahuan dan amalan guru dalam mengaplikasikan prinsip Six Sigma dalam PdP juga berpotensi memberi impak langsung terhadap keberkesanan pengajaran. Farooqi, Khalid dan Shafiq (2023) menekankan bahawa amalan Six Sigma di institusi pengajian tinggi mampu meningkatkan keberkesanan pengajaran-pembelajaran melalui proses pengukuran, analisis serta penambahbaikan berterusan. Implikasi daripada dapatan ini mengisyaratkan bahawa prospek penilaian kualiti pengajaran di sekolah boleh diperkaya dengan pendekatan yang sama, di mana guru bukan sahaja menilai kualiti pengajaran berdasarkan instrumen, tetapi juga mengembangkan pengetahuan dan kemahiran mereka untuk mempraktikkan penambahbaikan secara sistematik.

Secara keseluruhannya, dapatan ini mencadangkan bahawa penilaian kualiti pengajaran guru pada masa hadapan perlu mengambil kira dimensi kesedaran, pengetahuan, sikap dan amalan secara lebih menyeluruh. Kesedaran membolehkan guru melihat penilaian sebagai peluang pembangunan profesional; pengetahuan membekalkan kerangka teori dan teknikal yang diperlukan; sikap menentukan keterbukaan mereka terhadap perubahan; manakala amalan pula memastikan amalan penilaian dan penambahbaikan dapat diterapkan secara konsisten dalam bilik darjah. Justeru, prospek penilaian kualiti pengajaran bukan sahaja menumpukan kepada pengukuran keberkesanan PdP, tetapi juga membina guru yang reflektif, inovatif, dan berorientasikan penambahbaikan berterusan.

Implikasi Terhadap Pentaksiran Masa Hadapan

Impak jangka panjang dalam pentaksiran sempena kurikulum baharu masa hadapan ialah lahirnya sistem penilaian yang lebih reflektif, sistematik dan berorientasikan penambahbaikan berterusan. Penggunaan instrumen seperti $T^2Q_i-6\sigma$ yang berasaskan model Six Sigma DMAIC berpotensi memastikan pentaksiran bukan sekadar mengukur prestasi guru, tetapi turut menanam budaya profesionalisme, kesedaran sendiri, dan keterbukaan terhadap perubahan. Hal ini akan membina guru yang lebih inovatif, reflektif serta mampu menghubungkan maklum balas pelajar, data empirikal dan amalan pedagogi untuk meningkatkan kualiti PdP. Dalam jangka panjang, pentaksiran sebegini dapat menyokong pembentukan kurikulum adaptif yang lebih sejajar dengan keperluan abad ke-21 serta Matlamat Pembangunan Lestari (SDG).

Kesimpulan

Objektif utama kajian ini adalah untuk menilai kesahan muka serta kesesuaian instrumen $T^2Q_i-6\sigma$ bagi kegunaan dalam sistem pendidikan di Malaysia. Selain itu, kajian ini juga meneroka pandangan tentang penilaian kualiti pengajaran guru secara reflektif melalui pendekatan reflection for-action serta mencadangkan prospek penilaian kualiti pengajaran guru yang merangkumi aspek kesedaran, pengetahuan, sikap dan amalan. Penilaian kesahan muka instrumen ini memberi tumpuan kepada aspek adaptasi linguistik serta kesesuaian secara keseluruhan. Dapatan kajian jelas menunjukkan bahawa $T^2Q_i-6\sigma$ mempunyai tahap kesahan muka yang tinggi dan sesuai digunakan dalam kalangan sampel guru di Malaysia. Secara umum, aspek linguistik bagi semua item yang dianalisis didapati sesuai untuk kefahaman serta penguasaan guru. Kritikan kecil daripada pakar hanya melibatkan aspek penggunaan perkataan dan struktur ayat, yang telah diperhalusi oleh penyelidik bagi memastikan keteguhan dan konsistensi instrumen $T^2Q_i-6\sigma$. Oleh itu, dalam usaha meningkatkan dan memperkukuh standard pendidikan di Malaysia, pembangunan instrumen $T^2Q_i-6\sigma$ adalah penting untuk mengenal pasti kekuatan serta kelemahan guru dari segi kualiti pengajaran. Pelaksanaan pendekatan Six Sigma dalam sektor pendidikan melalui model DMAIC berpotensi membantu

Malaysia mencapai matlamat pendidikannya. Penyelidik berkeyakinan bahawa T²Qi-6σ berupaya menjadi satu instrumen yang sah dan boleh dipercayai dalam menilai serta meningkatkan kualiti pengajaran guru berdasarkan kerangka Six Sigma DMAIC. Kajian ini menyumbang kepada bidang psikometrik pendidikan melalui pembangunan dan penilaian kesahan muka instrumen T²Qi-6σ yang berasaskan pendekatan Six Sigma. Dari sudut metodologi, kajian ini memperluas aplikasi CVR dalam konteks kesahan muka. Dari sudut teori, kajian ini menghubungkan model DMAIC Six Sigma dengan penilaian kualiti pengajaran guru. Dari sudut praktikal pula, instrumen yang dibangunkan berpotensi membantu sekolah mengenal pasti kekuatan dan kelemahan pengajaran secara lebih sistematik bagi menyokong penambahbaikan berterusan. Namun demikian, T²Qi-6σ masih berada pada tahap pembangunan awal dan memerlukan penyelidikan lanjut, khususnya dalam aspek penilaian ciri psikometrik sebelum diaplikasikan secara praktikal. Dalam konteks ini, adalah disyorkan agar kajian lanjutan dijalankan dengan menilai kesemua 125 item yang telah diperhalusi melalui kajian rintis menggunakan analisis Model Pengukuran Rasch. Penggunaan analisis Rasch bagi menyelidik secara menyeluruh aspek kesahan dan kebolehpercayaan instrumen serta menambah baik item yang dibangunkan untuk menghasilkan pengukuran konstruk laten yang sah, tepat, dan konsisten merupakan satu peluang penyelidikan yang signifikan.

Penghargaan:	Penulis merakamkan setinggi-tinggi penghargaan kepada Kumpulan Penyelidikan Universiti (KPU) Penilaian Pendidikan, Universiti Kebangsaan Malaysia (UKM) atas sokongan dan kemudahan yang diberikan sepanjang kajian ini dijalankan. Penghargaan juga diberikan kepada pihak yang turut membantu secara langsung atau tidak langsung dalam kajian ini.
Penyataan Pembiayaan:	Kajian ini menerima sokongan kewangan daripada Fakulti Pendidikan, Universiti Kebangsaan Malaysia melalui Dana Penyelidikan FPEND di bawah Nombor Geran (GG-2024-044).
Pernyataan Konflik Kepentingan:	Penulis mengisytiharkan bahawa tiada konflik kepentingan berkaitan penerbitan kertas kerja ini. Semua penulis telah menyumbang kepada kajian ini dan telah meluluskan versi akhir manuskrip untuk penyerahan kepada International Journal of Modern Education (IJMOE)
Pernyataan Etika:	Kajian ini tidak melibatkan sebarang responden manusia, haiwan, atau data sensitif yang memerlukan kelulusan etika. Penulis mengesahkan bahawa penyelidikan ini telah dijalankan selaras dengan prinsip integriti akademik dan piawaian etika penerbitan yang diterima umum.
Pernyataan Sumbangan Penulis:	Semua penulis telah menyumbang secara signifikan terhadap pembangunan manuskrip ini. Kesemuanya bertanggungjawab terhadap pengkonsepian, metodologi dan penyeliaan keseluruhan kajian. Semua penulis turut mengambil bahagian dalam penulisan, semakan dan penambahbaikan manuskrip. Semua penulis telah membaca dan meluluskan versi akhir manuskrip sebelum penyerahan.

Rujukan

- Abdulla, A., & Kavilal. (2022). Analytical Investigation of Higher Education Quality Improvement by Using Six Sigma Approach. *HighTech and Innovation Journal*, 3(2), 196–206. <https://doi.org/10.28991/hij-2022-03-02-07>
- Al Kuwaiti, A., & Subbarayalu, A. V. (2015). Appraisal of Students Experience Survey (SES) as a Measure to Manage The Quality of Higher Education in the Kingdom of Saudi Arabia: An Institutional Study Using Six Sigma Model. *Educational Studies*, 41(4), 430–443. <https://doi.org/10.1080/03055698.2015.1043977>
- Allen, M. S., Robson, D. A., & Ilescu, D. (2023). Face Validity: A Critical but Ignored Component of Scale Construction in Psychological Assessment. *European Journal of Psychological Assessment*, 39(3), 153–156. <https://doi.org/10.1027/1015-5759/a000777>
- Al Masaeid, L. M., Al Sharafat, M. Q., Al Hawamed, M. H., Al-Zoubi, R. M., Hamadin, K. M., & Al-Nemrat, A. A. (2023). *Empowering Educational Management: Teachers' Insights into School Principals' Use of Six-Sigma Methodology*. Universitas Park Ltd. Sti.
- Almutairi, T. S., & Shraid, N. S. (2021). Teacher Evaluation by Different Internal Evaluators: Head of Departments, Teachers Themselves, Peers and Students. *International Journal of Evaluation and Research in Education*, 10(2), 588–596. <https://doi.org/10.11591/ijere.v10i2.20838>
- Artino, A. R., La Rochelle, J. S., Dezee, K. J., & Gehlbach, H. (2014). Developing Questionnaires for Educational Research: AMEE Guide No. 87. *Medical Teacher*, 36(6), 463–474. <https://doi.org/10.3109/0142159X.2014.889814>
- Azwani Masuwai, Hafizhah Zulkifli, & Mohd Isa Hamzah. (2024). Evaluation of Content Validity and Face Validity of Secondary School Islamic Education Teacher Self-Assessment Instrument. *Cogent Education*, 11(1), 1–12. <https://doi.org/10.1080/2331186X.2024.2308410>
- Bas Senden, T., Nilsen, N., & Teig, N. (2023). The validity of student ratings of teaching quality: Factorial structure, comparability, and the relation to achievement. *Studies in Educational Evaluation*, 78, 101274. <https://doi.org/10.1016/j.stueduc.2023.101274>
- Blomeke, S., Jentsch, A., Ross, N., Kaiser, G., & König, J. (2022). Opening up the Black Box: Teacher Competence, Instructional Quality and Students' Learning Progress. *Learning and Instruction*, 79, 1–20. <https://doi.org/10.1016/j.learninstruc.2022.101600>
- Blomeke, S., & Olsen, R. V. (2019). Consistency of Results Regarding Teacher Effects Across Subjects, School Levels, Outcomes and Countries. *Teaching and Teacher Education*, 77, 170–182. <https://doi.org/10.1016/j.tate.2018.09.018>
- Bloom, D. (2018). The Excellent Education System Using Six Sigma To Transform Schools. In *Routledge Taylor & Francis Group*. Routledge Taylor & Francis Group. <http://www.taylorandfrancis.com>
- Bougie, R., & Sekaran, U. (2020). *Research Methods For Business A Skill Building Approach* (8th Editions) John Wiley & Sons, In.
- Burhan-Horasanlı, E., & Hart, D.O. (2024). Conceptualizing reflection-for-action: Empowering teachers as agents of collaborative and collective transformation. *Teaching and Teacher Education*, 152, 104802. <https://doi.org/10.1016/j.tate.2024.104802>
- Callahan, K., & Sadeghi, L. (2015). Teacher Perceptions of The Value of Teacher Evaluations: New Jersey's ACHIEVE NJ. *International Journal of Educational Leadership Preparation*, 10(1), 46–59.

- Cohen, L., Manion, L., & Morrison, K. (2018). *Research Methods in Education* (Edisi ke-8). Routledge.
- Costa, L. B. M., Godinho Filho, M., Fredendall, L. D., & Devós Ganga, G. M. (2021). Lean Six Sigma in The Food Industry: Construct Development and Measurement Validation. *International Journal of Production Economics*, 1(2), 1–27. <https://doi.org/10.1016/j.ijpe.2020.107843>
- Creswell, J. W., & C. Guetterman, T. (2019). *Educational Research Planning, Conducting And Evaluating Quantitative And Qualitative Research* (6th Edition). Pearson Education, Inc.
- Dalawi, I., Isa, M. R., Chen, X. W., Azhar, Z. I., & Aimran, N. (2023). Development of the Malay Language of Understanding, Attitude, Practice and Health Literacy Questionnaire on COVID-19 (MUAPHQ C-19): Content Validity & Face Validity Analysis. *BMC Public Health*, 23(1131), 1–14.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2015). Evaluating Teacher Evaluation. *Kappanmagazine.Org*, Jun, 8–15. <https://doi.org/10.4324/9780203053874-32>
- Deti Rostini, Rifqi Zaeni Achmad Syam, & Willya Achmad. (2022). The Significance of Principal Management on Teacher Performance and Quality of Learning. *Jurnal Pendidikan*, 14(2), 2513–2520. <https://doi.org/10.35445/alishlah.v14i2.1721>
- Divayana, D. G. H., Suyasa, P. W. A., & Adiarta, A. (2020). Content Validity Determination of the Countenance-Tri Kaya Parisudha Model Evaluation Instruments using Lawshe's CVR formula. *Journal of Physics: Conference Series*, 1516, 1–7. <https://doi.org/10.1088/1742-6596/1516/1/012047>
- Farooqi, M. T. K., Khalid, M. N., & Shafiq, A. (2023). Impact of Six Sigma Practices on Teaching-Learning Effectiveness: A Case Study of the University of Okara. *Global Social Sciences Review*, VIII(II), 657–670. [https://doi.org/10.31703/gssr.2023\(VIII-II\).58](https://doi.org/10.31703/gssr.2023(VIII-II).58)
- Furr, R. M. (2021). *Psychometrics: An Introduction* (4th Editions). SAGE Publications.
- Gerritsen, S., Plug, E., & Webbink, D. (2016). Teacher Quality and Student Achievement: Evidence From a Sample of Dutch Twins. *Journal of Applied Econometrics*, 32(3), 643–660. <https://doi.org/10.1002/jae>
- Gilbert, G. E., & Prion, S. (2016). Making Sense of Methods and Measurement: Lawshe's Content Validity Index. *Clinical Simulation in Nursing*, 12(12), 530–531. <https://doi.org/10.1016/j.ecns.2016.08.002>
- Jayamohan K.G, & Bhasi. A.B. (2021). Development of a Tool for Measuring Performance of Higher Educational Institutions by Applying Six Sigma Methodology in Teaching-Learning Process. *International Journal of Mechanical Engineering*, 6(3), 3525–3531.
- Jin, H., Qu, P., Zhou, Y., Qu, Q., & Li, H. (2024). Green Lean Six Sigma practice in the industrial and service sectors: from the systematic literature review to an integration diagram. *Production Planning & Control*, 36(3), 394–419. <https://doi.org/10.1080/09537287.2024.2369779>
- Lawshe, C. H. (1975). A Quantitative Approach to Content Validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- LeMahieu, P. G., Nordstrum, L. E., & Cudney, E. A. (2017). Six Sigma In Education. *Quality Assurance in Education*, 25(1), 91–108. <https://doi.org/10.1108/QAE-12-2016-0082>
- Kamaruddin, M., & Matore, M. E. E. M. (2021). Development and validation of psychometric properties of the 10 IB learner profile instrument (10IBLP-I): A combination of the rasch and classical measurement model. *International Journal of Environmental Research and Public Health*, 18(12), 6455. <https://doi.org/10.3390/ijerph18126455>

- Miller, L. A., & Lovler, R. L. (2020). *Foundations Of Psychological Testing A Practical Approach* (6th Editions). Sage Publications Ltd.
- Mittal, A., Gupta, P., Kumar, V., Al Owad, A., Mahalwat, S., & Singh, S. (2023). The Performance Improvement Analysis Using Six Sigma DMAIC Methodology: A Case Study on Indian Manufacturing Company. *Heliyon*, 9(3). <https://doi.org/10.1016/j.heliyon.2023.e14625>
- Oluwatayo, J. A. (2012). Validity and Reliability Issues in Educational Research. *Journal of Educational and Social Research*, 2(May), 391–400. <https://doi.org/10.5901/jesr.2012.v2n2.391>
- Pal, M., & Ghosh, S. (2022). Applied Six Sigma Theory to Reduce the Drop-Out Rate Among the Student: Systematic Review. *Journal of Social Sciences and Management Studies*, 3(234), 104–108. <https://doi.org/10.56556/jssms.v1i3.234>
- Patel, N., & Desai, S. (2020). ABC of Face Validity for Questionnaire. *International Journal of Pharmaceutical Sciences Review and Research*, 65(1), 164–168. <https://doi.org/10.47583/ijpsrr.2020.v65i01.025>
- Pyzdek, T., & Keller, P. (2018). *The Six Sigma Handbook A Complete Guide for Green Belts, Black Belts and Managers at All Levels* (5th Editions). McGraw Hill Education.
- Romero Jeldres, M., Díaz Costa, E., & Faouzi Nadim, T. (2023). A Review of Lawshe’s Method for Calculating Content Validity in the Social Sciences. *Frontiers in Education*, 8(November). <https://doi.org/10.3389/educ.2023.1271335>
- Sabtu, S. H., & Motare, M. E. M. (2024). Systematic Literature Review on the Implementation of the Six Sigma Approach in Education. *International Journal of Evaluation and Research in Education (IJERE)*, 13(1), 262. <https://doi.org/10.11591/ijere.v13i1.26196>
- Sadeghi, N., Akbarfahimi, M., & Taghizadeh Hajlouee, G. (2021). The Persian Version of “Meaningful Activity Participation Assessment” Questionnaire: Evaluation of Its Face and Content Validity in Elderly. *Function and Disability Journal*, 4(1), E44. <https://doi.org/10.32598/fdj.4.44>
- Sovey, S., Osman, K., Matore, M. E. E. M. (2022). Exploratory and Confirmatory Factor Analysis for Disposition Levels of Computational Thinking Instrument Among Secondary School Students. *European Journal of Educational Research*, 11(2), 639–652. <https://doi.org/10.12973/eu-jer.11.2.639>
- Taherdoost, H. (2016). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *International Journal Of Academic Research in Management (IJARM)*, 5(3), 28–36. <https://doi.org/10.2139/ssrn.3205040>
- Wang, Q. (2022). Application of Six Sigma Management-based Teaching Method in Financial Management Course Online Teaching. *International Journal of Emerging Technologies in Learning*, 17(1), 60–73. <https://doi.org/10.3991/ijet.v17i01.28269>
- Yaw, P. W., & Matore, M. E. M. (2024). Development and Validation of an Instrument To Measure Science Teachers ’ Instructional Preparedness in Stem. *Educational Administration: Theory and Practice*, 30(4), 8937–8949. <https://doi.org/10.53555/kuey.v30i4.2932>
- Yu, K. T., & Ueng, R. G. (2012). Enhancing Teaching Effectiveness by Using The Six-Sigma DMAIC Model. *Assessment and Evaluation in Higher Education*, 37(8), 949–961. <https://doi.org/10.1080/02602938.2011.592933>
- Yusoff, M. S. B. (2019). ABC of Response Process Validation and Face Validity Index Calculation. *Education in Medicine Journal*, 11(3), 55–61. <https://doi.org/10.21315/eimj2019.11.3.6>