

Konsepsualisasi Perancangan dan Pembinaan Instrumen Penilaian Dalam Era Kecerdasan Buatan

Conceptualizing The Planning and Development of Assessment Instruments in the Era of Artificial Intelligence

Rosidah Md Desa¹ & Mohd Effendi@Ewan bin Matore^{2*}

^{1,2}. Fakulti Pendidikan, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

Article progress

Accepted: 23 September 2025

Reviewed: 2 Oktober 2025

Published: 30 November 2025

*Mohd Effendi@ Ewan bin Matore, Fakulti Pendidikan, Universiti Kebangsaan Malaysia, Selangor
Email: effendi@ukm.edu.my

Abstrak: Kemajuan teknologi kecerdasan buatan (AI) dalam bidang pendidikan telah membawa perubahan ketara terhadap pendekatan penilaian, yang menuntut pembangunan instrumen lebih autentik, sahih dan selaras dengan prinsip kemajuan kecerdasan buatan. Dalam konteks ini, perancangan pembangunan instrumen sistematik menjadi teras penting bagi memastikan keupayaan menilai kompetensi pelajar dalam ekosistem pembelajaran berasaskan AI dapat dijalankan secara berkesan dan bertanggungjawab. Kertas konsep ini membincangkan cadangan perancangan dan pembinaan instrumen penilaian dalam era kecerdasan buatan. Metodologi yang digunakan dalam pembangunan instrumen menggunakan pendekatan seperti model ADDIE, serta pendekatan penilaian psikometrik seperti analisis faktor eksploratori (EFA), analisis faktor sahkan (CFA), Alpha Cronbach dan Composite Reliability untuk menguji kesahan dan kebolehpercayaan. Perhatian turut diberikan kepada panduan konseptual dan praktikal kepada penyelidik, pendidik dan pembuat dasar dalam mereka bentuk penilaian yang dapat mengukur literasi AI, kejujuran akademik dan etika teknologi secara efektif. Implikasinya, instrumen yang dibina bukan sahaja bertindak sebagai alat pengukuran, tetapi turut menjadi intervensi nilai dalam pendidikan digital. Instrumen ini boleh menyokong pelaksanaan kurikulum literasi AI dan membantu menilai keberkesanan program pembelajaran berasaskan teknologi. Kajian lanjutan disarankan agar melibatkan ujian terhadap populasi pelbagai, kajian longitudinal untuk menelusuri perubahan sikap pelajar terhadap AI, serta penggunaan pendekatan kualitatif seperti temu bual atau refleksi pelajar untuk memperkukuh pemahaman terhadap dinamika nilai dan keputusan etika pelajar. Perbincangan ini boleh memberi sumbangan bermakna dalam membina sistem penilaian yang relevan, sahih dan beretika dalam era kecerdasan buatan.

Kata kunci: pembinaan instrumen, kecerdasan buatan, penilaian, etika digital, penilaian abad ke-21

Abstract: The advancement of artificial intelligence (AI) technology in education has brought significant changes to assessment approaches, requiring the development of more authentic, valid, and AI-aligned instruments. In this context, systematic planning for instrument development serves as an essential foundation to ensure the effective and responsible evaluation of students' competencies within AI-based learning ecosystems. This conceptual paper discusses the proposed planning and development of assessment instruments in the era of artificial intelligence. The methodology

applied in the instrument development employs approaches such as the ADDIE model and psychometric evaluation techniques, including Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), Cronbach's Alpha, and Composite Reliability to examine validity and reliability. Attention is also given to providing conceptual and practical guidance for researchers, educators, and policymakers in designing assessments that can effectively measure AI literacy, academic honesty, and technological ethics. The implications suggest that the developed instrument not only functions as a measurement tool but also serves as a value-based intervention in digital education. This instrument can support the implementation of AI literacy curricula and assist in evaluating the effectiveness of technology-enhanced learning programs. Future studies are recommended to include diverse populations, conduct longitudinal research to trace changes in students' attitudes toward AI, and apply qualitative approaches such as interviews or student reflections to strengthen understanding of students' ethical reasoning and value dynamics. This discussion is expected to make a meaningful contribution toward developing a relevant, valid, and ethical assessment system in the era of artificial intelligence.

Keywords: instrument development, artificial intelligence, assessment, digital ethics, 21st-century assessment

Pengenalan

Kecerdasan buatan (AI) kini menjadi tonggak utama dalam merevolusikan landskap pendidikan abad ke-21. Alat seperti ChatGPT semakin kerap dimanfaatkan dalam proses pengajaran dan pembelajaran, memberikan akses pantas kepada maklumat serta mempercepatkan penyediaan tugas. Namun, kebergantungan kepada AI juga mendatangkan cabaran terhadap keaslian hasil kerja pelajar, termasuk risiko plagiarisme, pemalsuan, dan kebergantungan berlebihan kepada teknologi yang menjejaskan integriti akademik (Yeliena Prokhorova et al., 2024).

Sehubungan itu, terdapat keperluan mendesak untuk merangka instrumen penilaian yang tepat dan sahih bagi mengukur pencapaian pembelajaran, kemahiran, serta tahap integriti pelajar dalam penggunaan AI (Mohd Alif Aineh & Wirawati Ngui, 2024; Tang & Chaw, 2023). Instrumen ini harus meliputi bukan sahaja pengukuran kognitif tradisional, tetapi juga dimensi etika, sikap, dan literasi digital agar pelajar dapat memanfaatkan teknologi secara bertanggungjawab. Perancangan pembinaan instrumen yang tersusun menjadi teras penting bagi memastikan penilaian kekal relevan dan berdaya tahan dalam persekitaran digital.

Dengan mengambil kira kerangka teori terkini dan amalan terbaik, instrumen yang direka dengan rapi mampu menangani aspek metakognitif dan afektif, seterusnya membantu guru serta pentadbir pendidikan mengesan tahap literasi AI pelajar dan merangka intervensi yang lebih berfokus. Pendekatan ini bukan sahaja menyokong sistem pendidikan yang responsif dan berinovasi, tetapi juga membentuk generasi pelajar yang cekap teknologi, beretika, dan bertanggungjawab di era kecerdasan buatan (Knoth et al., 2024). Justeru, kajian ini menetengahkan cadangan perancangan dan pembinaan instrumen penilaian yang relevan dengan penggunaan AI.

Sorotan Literatur

Kepentingan Perancangan Instrumen Penilaian

Dalam era pendidikan digital masa kini, kecerdasan buatan (AI) telah memainkan peranan yang semakin penting dalam membentuk pendekatan baharu terhadap penilaian pembelajaran. AI bukan sahaja mempercepatkan proses analisis data pelajar, tetapi juga memperkenalkan mekanisme penilaian automatik, peribadi dan berskala besar. Namun, manfaat ini

tidak akan dapat direalisasikan secara berkesan tanpa perancangan yang teliti dalam pembinaan instrumen penilaian yang digunakan dalam sistem berasaskan AI.

Perancangan instrumen penilaian yang sistematik amat penting untuk memastikan bahawa apa yang diukur benar-benar mencerminkan kompetensi, pemikiran aras tinggi, dan perkembangan sahsiah pelajar. Dalam konteks AI, kekurangan perancangan akan menyebabkan penilaian yang berat sebelah, gagal mengambil kira konteks pembelajaran, serta menghasilkan keputusan yang tidak sah atau tidak sah. Seperti yang diketahui kesahihan instrumen merupakan elemen teras dalam penilaian berasaskan AI kerana ketidaktepatan dalam reka bentuk item boleh membawa kepada keputusan sistem yang menyesatkan dan menjejaskan keadilan pendidikan.

Selain itu, proses perancangan turut melibatkan pembangunan konstruk yang jelas, pembinaan item yang sesuai dengan persekitaran digital, dan pemilihan penunjuk yang sepadan dengan tahap kognitif serta hasil pembelajaran yang ingin diukur. Pembangunan instrumen penilaian AI perlu melalui langkah semakan pakar, ujian rintis, dan pengesahan statistik seperti analisis faktor pengesahan dan alpha kebolehpercayaan bagi memastikan kebolehpercayaan dan keabsahan instrumen. Tambahan pula, dalam konteks penggunaan AI secara meluas, aspek etika dan privasi turut menjadi faktor kritikal yang perlu dirancang sejak awal.

Dalam konteks pendidikan Malaysia, perkembangan kurikulum yang menekankan Kemahiran Berfikir Aras Tinggi (KBAT) dan pembelajaran sendiri menjadikan AI sebagai pelengkap kepada sistem pendidikan sedia ada. Oleh itu, penilaian yang dibantu AI perlu disokong dengan instrumen yang direka bentuk secara konstruktif, mengambil kira keperluan pedagogi tempatan serta selari dengan matlamat pendidikan nasional. Secara keseluruhannya, pembinaan instrumen penilaian untuk sistem berasaskan kecerdasan buatan menuntut perancangan yang menyeluruh, holistik dan berpaksikan prinsip-prinsip pedagogi, metrik dan etika. Tanpa perancangan yang rapi, AI bukan sahaja gagal menambah baik sistem penilaian, bahkan boleh memperluaskan ketidakseimbangan pendidikan. Justeru, kerjasama antara pakar teknologi, ahli pedagogi dan pembuat dasar adalah penting dalam membentuk asas kukuh bagi perancangan dan pelaksanaan instrumen penilaian AI dalam pendidikan.

Teori Dan Konseptual Dalam Pembangunan Instrumen

Pembinaan instrumen penilaian dalam konteks pendidikan memerlukan asas teori dan kerangka konseptual yang mantap untuk menjamin kesahan dan kebolehpercayaan. Dalam era kecerdasan buatan (AI), keperluan ini menjadi lebih mendesak memandangkan perubahan landskap pembelajaran dan penggunaan teknologi generatif seperti ChatGPT. Penilaian bukan sahaja harus bersifat autentik dan relevan, malah perlu mengambil kira interaksi pelajar dengan AI serta nilai etika dan tanggungjawab digital. Bahagian ini membincangkan pelbagai teori yang menjadi asas kukuh dalam pembangunan instrumen penilaian, dengan menekankan integrasi pendekatan pedagogi, teknologi, dan nilai.

Model Pembangunan Instrumen

Dalam proses pembangunan instrumen penilaian, pemilihan dan penggunaan model pembangunan yang sistematik adalah suatu keperluan asas bagi memastikan reka bentuk yang dilaksanakan bersifat sah, boleh dipercayai dan relevan dengan konteks semasa pendidikan. Instrumen yang dibangunkan tanpa asas model yang jelas berisiko menghasilkan item yang kabur, tidak berfokus dan sukar ditafsir. Oleh itu, model pembangunan instrumen bertindak sebagai panduan menyeluruh yang mengatur langkah demi langkah dalam mengenal pasti keperluan, membina konstruk, merangka item, melaksanakan ujian rintis, dan membuat semakan berdasarkan data empirik.

Dalam konteks pendidikan abad ke-21 yang dipacu oleh teknologi, terutamanya dengan kemunculan sistem kecerdasan buatan seperti ChatGPT dan platform AI lain, pembangunan instrumen memerlukan pendekatan yang berstruktur dan responsif terhadap perubahan teknologi dan pedagogi. Model pembangunan berfungsi bukan sekadar kerangka kerja, tetapi sebagai landasan yang menyepadukan teori, amalan dan konteks sebenar pengguna. Dengan berpandukan model pembangunan yang tepat, para penyelidik dan pendidik dapat memastikan instrumen yang dibina mampu mengukur konstruk secara tepat, mengelakkan bias, dan menyokong keputusan berasaskan data yang sah dan bermakna. Kajian oleh turut menekankan keperluan untuk menghubungkan struktur domain, model konstruk, dan kerangka pentaksiran secara jelas bagi menghasilkan instrumen yang menyeluruh dan bermakna dalam penilaian pendidikan era digital.

Salah satu kerangka yang paling kerap diaplikasikan dalam pembangunan instrumen ialah Model ADDIE yang terdiri daripada lima fasa utama iaitu *Analysis, Design, Development, Implementation, dan Evaluation*. Model ini menyediakan kerangka pembangunan instrumen yang sistematik, membolehkan pengkaji mengenal pasti keperluan, merancang reka bentuk item, membangunkan alat ukur, melaksanakannya, dan membuat penilaian semula berdasarkan data yang diperolehi (Samuel Boateng et al., 2024). Dalam konteks penilaian penggunaan AI dalam pendidikan, fasa analisis misalnya, akan melibatkan pengenalpastian konstruk seperti keupayaan menilai hasil AI, kesedaran etika dan kejujuran akademik.

Teori Pembelajaran dan Kognitif

Teori pembelajaran memainkan peranan penting dalam pembangunan instrumen penilaian kerana ia memberi asas kepada pemahaman tentang bagaimana pelajar memperoleh, memproses dan menggunakan maklumat. Teori ini membantu memastikan bahawa konstruk yang ingin diukur dalam instrumen adalah sejajar dengan cara pelajar belajar dan membina pengetahuan. Selain itu, teori pembelajaran juga menyokong penyusunan item yang sesuai dengan tahap kognitif pelajar, keperluan perkembangan, dan konteks pembelajaran sebenar.

Teori Konstruktivisme menyarankan bahawa pelajar membina pengetahuan melalui pengalaman dan interaksi. Dalam pembangunan instrumen, pendekatan ini membantu memastikan item mencerminkan konteks pembelajaran sebenar serta menyokong pembinaan makna secara aktif (Sigit Wibowo et al., 2025). Teori sosio konstruktivisme pula menekankan peranan interaksi sosial dan budaya dalam pembelajaran, justeru item yang dibina perlu mengambil kira nilai kebersamaan, kolaborasi dan komunikasi pelajar dalam suasana pembelajaran digital.

Teori pembelajaran kognitif memberi asas penting dalam pembinaan instrumen penilaian yang bersifat diagnostik dan berorientasikan pemikiran aras tinggi. Di bawah kerangka ini, Teori Pemprosesan Maklumat menerangkan bagaimana pelajar menerima, memproses dan menyimpan maklumat, yang boleh dijadikan asas dalam menentukan struktur item penilaian, tahap kesukaran soalan, serta format yang sesuai. Selain itu, Model Pemprosesan Berasaskan Waktu (Time-Based Resource-Sharing, TBRS) memberikan pandangan lanjut tentang pengurusan beban kognitif dan penggunaan sumber mental dalam tugas yang melibatkan interaksi manusia dan komputer (Natalia Sevchenko et al., 2023). Dalam konteks kecerdasan buatan, model ini dapat membantu merangka item yang menilai kecekapan pelajar dalam menangani tugas penilaian sendiri berdasarkan maklumat kecerdasan buatan dalam situasi masa nyata dan tekanan kognitif.

Teori Etika Penggunaan Teknologi

Teori-teori etika dalam penggunaan teknologi seharusnya turut diberi perhatian dan dimasukkan dalam kerangka pembinaan instrumen penilaian, khususnya dalam konteks era kecerdasan buatan (AI) yang semakin berkembang dalam bidang pendidikan. Hal ini penting bagi memastikan instrumen yang dibina bukan sahaja mengukur aspek kognitif dan kemahiran pelajar, tetapi juga mempertimbangkan nilai, sikap, dan pertimbangan etika dalam penggunaan teknologi. Sebagai contoh, kajian oleh Nguyen et al. (2023) telah mengetengahkan prinsip-prinsip etika seperti keadilan (fairness), ketelusan (transparency), privasi (privacy), dan akauntabiliti (accountability) sebagai asas penting dalam memastikan teknologi digunakan secara bertanggungjawab. Prinsip-prinsip ini boleh dijadikan rujukan dalam membina item-item penilaian yang menyentuh aspek kebertanggungjawaban pelajar semasa menggunakan AI, termasuk dalam konteks penyediaan tugas akademik.

Dalam erti kata lain, pembinaan instrumen dalam era kecerdasan buatan perlu melangkaui penilaian terhadap pengetahuan dan kebolehan pelajar semata-mata. Ia juga perlu merangkumi unsur afektif dan reflektif yang menilai bagaimana pelajar menggunakan AI dalam proses pembelajaran mereka serta mengapa mereka memilih untuk menggunakan teknologi tersebut. Instrumen yang komprehensif harus mampu mengenal pasti tahap kesedaran etika pelajar dalam menggunakan AI, termasuk keupayaan mereka membuat keputusan yang berasaskan prinsip moral dan tanggungjawab akademik. Dengan memasukkan dimensi etika ini dalam reka bentuk instrumen, pendidik bukan sahaja dapat mengukur pencapaian pembelajaran pelajar secara holistik, malah dapat memupuk budaya penggunaan teknologi yang berintegriti, selari dengan aspirasi pendidikan abad ke-21.

Keseluruhannya, asas teori dan konseptual dalam pembangunan instrumen bukan sekadar memberikan arah kepada pembangunan konstruk dan item, malah menjadi tunjang utama dalam memastikan instrumen itu benar-benar relevan, sah dan mampu menyokong amalan pendidikan yang beretika dan berpandangan jauh dalam era kecerdasan buatan.

Metodologi

Perancangan dan pembangunan instrumen penilaian dalam era kecerdasan buatan menuntut pendekatan yang sistematik, berasaskan teori, dan sejajar dengan perubahan landskap pendidikan digital masa kini. Proses pembangunan yang teliti bukan sahaja penting untuk menjamin kesahan dan kebolehpercayaan instrumen, malah memastikan ia benar-benar relevan dengan konteks penggunaan AI di peringkat sekolah.

Menurut Delcker et al.(2024), pembangunan instrumen yang berkesan perlu berpaksikan kepada rangka kerja yang terancang dan berterusan, yang menggabungkan elemen teori, pengalaman praktikal, serta kepekaan terhadap perkembangan teknologi semasa. Pendekatan ini amat penting bagi memastikan instrumen yang dibangunkan mampu menilai dan menyokong pembangunan kompetensi AI dalam kalangan pendidik dan pelajar, selaras dengan keperluan pendidikan abad ke-21. Berikut adalah cadangan langkah- langkah pembinaan instrumen penilaian dalam era kecerdasan buatan.

Penentuan Matlamat dan Dimensi Pengukuran

Langkah pertama adalah penentuan skop dan matlamat instrumen berdasarkan objektif penilaian. Dalam konteks AI seperti penggunaan ChatGPT, dimensi seperti literasi AI, kejujuran akademik, kemahiran penilaian maklumat dan tanggungjawab digital perlu diberi perhatian. Penyelidik perlu mengenal pasti konstruk utama yang ingin diukur dan memastikan ia mempunyai hubungan langsung dengan fenomena pendidikan yang sedang berlaku.

Kajian Literatur dan Penilaian Instrumen Sedia Ada

Pembangunan instrumen perlu berpandukan dapatan daripada kajian lepas yang berkaitan termasuklah mengenal pasti jurang dalam instrumen yang sedia ada atau kelemahan dari aspek kesahan konstruk. Kajian oleh Hamed Abdelreheem Ead (2024) menyatakan banyak instrumen yang digunakan cenderung menumpukan kepada pengukuran kompetensi digital tanpa secara eksplisit menilai elemen penting seperti integriti akademik dalam penggunaan AI. Ia menunjukkan keperluan untuk pembangunan alat penilaian yang lebih holistik yang meliputi aspek kejujuran akademik dan etika penggunaan teknologi.

Oleh itu, kajian literatur perlu dijalankan secara menyeluruh bagi mengenal pasti teori, model, dan prinsip etika serta pedagogi yang dapat menyokong pembangunan konstruk dan item instrumen. Antara prinsip etika yang boleh diguna pakai termasuk keadilan, ketelusan, privasi, dan akauntabiliti seperti yang digariskan oleh Nguyen et al. (2023). Selain itu, teori pengukuran seperti *Classical Test Theory* (CTT) dan *Item Response Theory* (IRT) boleh dijadikan asas dalam penilaian kebolehpercayaan dan kesahan instrumen. Kajian literatur juga dapat membantu membina pemahaman terhadap kesesuaian konstruk yang telah digunakan dalam kajian terdahulu, sekali gus memperkukuh asas teori pembangunan instrumen baharu.

Pembinaan Item Berasaskan Teori dan Konteks

Pada asasnya, teori berfungsi sebagai panduan utama bagi pembinaan instrumen penilaian. Teori membantu kita mengenal pasti dengan jelas apakah aspek penting atau konstruk yang ingin diukur dan bagaimana kaitan antara konstruk tersebut. Sebagai contoh, teori pengukuran seperti *Classical Test Theory* (CTT) dan *Item Response Theory* (IRT) telah lama digunakan untuk memastikan kualiti item penilaian berada pada tahap yang boleh dipercayai dan sah. Teori-teori ini memberi kita panduan tentang cara membina, menapis dan menganalisis item dengan betul.

Dalam konteks pendidikan masa kini yang semakin dipengaruhi oleh teknologi, khususnya kecerdasan buatan (AI), teori pengukuran klasik sahaja tidak lagi mencukupi. Pembinaan item perlu mengambil kira konteks semasa dan keperluan etika, seperti penggunaan AI dalam penyediaan tugas, isu plagiarisme, serta kesedaran pelajar dan pendidik terhadap akauntabiliti akademik. Justeru itu, instrumen yang ingin dibina harus juga selari dengan prinsip etika AI seperti keadilan, ketelusan, privasi dan tanggungjawab, sebagaimana yang diketengahkan oleh Nguyen et al.

(2023) dalam kajian sistematik mereka tentang amalan AI dalam pendidikan. Penyataan disokong oleh kajian antarabangsa terkini seperti oleh Ziyang Wang et al. (2025) dan Guo et al. (2024) yang telah membangunkan instrumen penilaian menggabungkan unsur kognitif, teknikal, dan etika AI, serta diuji menggunakan pendekatan seperti *Confirmatory Factor Analysis* (CFA) dan *Cronbach's Alpha*. Kajian ini menunjukkan keperluan untuk membina item yang bukan sahaja mengukur tahap pengetahuan atau kemahiran teknikal, tetapi juga kefahaman dan refleksi etika dalam penggunaan AI. Dengan itu, gabungan antara teori pengukuran, prinsip etika serta konteks pendidikan digital semasa menjadi asas penting bagi pembangunan item yang relevan, sah, dan boleh dipercayai. Instrumen yang dibangunkan melalui pendekatan ini lebih berupaya untuk menangkap realiti penggunaan AI dalam pendidikan dan menyumbang kepada pembangunan penilaian yang lebih menyeluruh.

Penjanaan Item Instrumen

Penjanaan item merupakan satu fasa penting dalam pembangunan instrumen kerana ia menentukan sejauh mana konstruk yang telah dikenal pasti dapat diukur secara tepat dan berkesan. Setiap item yang dibina mestilah berasaskan indikator yang jelas dan mencerminkan aspek yang ingin diukur. Dalam proses ini, pemilihan kata kunci, struktur ayat, dan bentuk soalan perlu diberi perhatian agar item yang dihasilkan bersesuaian dengan tahap kefahaman dan konteks kumpulan sasaran, sama ada pelajar, guru, atau pengguna pendidikan lain. Penggunaan bahasa yang ringkas, jelas dan tidak mengelirukan adalah kunci kepada keberkesanan item.

Dalam konteks era kecerdasan buatan (AI), pendekatan tradisional dalam penjanaan item perlu dikembangkan agar sejajar dengan realiti pendidikan semasa. Item-item yang dibina bukan sahaja perlu menilai tahap kognitif pengguna, tetapi juga kemahiran digital, kesedaran etika dan keupayaan reflektif terhadap penggunaan teknologi AI. Sebagai contoh, item boleh dirangka dalam bentuk senario dunia sebenar yang melibatkan penggunaan alat AI seperti ChatGPT dalam penyediaan tugas, penulisan karangan, atau penyelidikan dalam talian. Dalam senario ini, responden boleh ditanya tentang pilihan tindakan, pertimbangan etika, atau justifikasi keputusan mereka dalam menggunakan AI, bagi menilai integriti akademik dan literasi AI mereka.

Secara keseluruhannya, penjanaan item bukan sahaja menuntut kefahaman teori dan konstruk, tetapi juga kreativiti dan kepekaan terhadap perubahan dalam ekosistem pendidikan digital. Instrumen yang relevan dan kontemporari lahir daripada proses penjanaan item yang menyeluruh, teliti dan berteraskan prinsip etika serta realiti penggunaannya.

Semakan Pakar dan Penambahbaikan Awal

Semakan pakar dijalankan setelah draf awal instrumen disediakan, dengan tujuan utama untuk menilai sejauh mana setiap item mewakili konstruk yang ingin diukur secara tepat, serta meneliti aspek kejelasan bahasa, ketepatan format dan kesesuaian kandungan dengan konteks penggunaan sebenar. Dalam konteks instrumen yang membabitkan penggunaan teknologi seperti kecerdasan buatan (AI), proses semakan ini menjadi lebih penting kerana ia membantu mengesan isu teknikal dan keperluan konteks yang mungkin tidak timbul dalam pembangunan instrumen konvensional.

Proses ini biasanya melibatkan panel pakar yang terdiri daripada individu berpengalaman dalam bidang pengukuran dan penilaian, pendidikan, serta teknologi atau kandungan berkaitan. Semakan dilakukan menggunakan pendekatan berstruktur, sama ada melalui borang penilaian atau temubual, bagi mendapatkan maklum balas mengenai kesesuaian setiap item, sama ada dari segi struktur ayat, tahap kesukaran, serta kerelevanan terhadap konstruk dan populasi sasaran. Maklum balas ini kemudiannya digunakan oleh penyelidik untuk membuat penambahbaikan terhadap item yang lemah, mengubahsuai istilah atau frasa yang mengelirukan, serta memperkemas susunan dan format instrumen.

Kajian oleh menunjukkan bahawa proses semakan pakar memainkan peranan penting dalam pembinaan instrumen berkaitan penerimaan AI dalam pendidikan. Menurut Roza García et al. (2024) dalam kajiannya menyatakan, proses semakan pakar memainkan peranan penting dalam membangun dan menyempurnakan instrumen, terutama dalam konteks penilaian kandungan dan memastikan kejelasan serta relevansi item. Melalui proses ini, penyelidik bukan sahaja dapat meningkatkan mutu kandungan item, malah dapat memastikan bahawa instrumen yang dibangunkan bersifat praktikal, relevan, dan dapat diterima baik oleh kumpulan sasar. Justeru, semakan pakar dan penambahbaikan awal dianggap sebagai satu fasa penting yang menyumbang secara langsung kepada pembinaan instrumen yang sah, boleh dipercayai dan responsif terhadap keperluan semasa dalam pendidikan digital.

Ujian Rintis dan Analisis Statistik

Ujian rintis atau *pilot testing* merujuk kepada proses pengujian awal instrumen ke atas sampel kecil daripada populasi sasaran. Langkah ini bertujuan untuk mengenal pasti sama ada item yang dibina difahami dengan jelas, sesuai dari segi konteks dan format, serta berfungsi sebagaimana yang dimaksudkan oleh pembina instrumen. Ia juga merupakan fasa awal dalam mendapatkan data empirikal yang diperlukan untuk menjalankan analisis psikometrik.

Dalam ujian rintis, fokus diberikan kepada beberapa aspek penting seperti kefahaman arahan dan skala, kesesuaian masa menjawab, serta kesediaan teknikal platform jika instrumen dijalankan secara digital. Penyelidik juga berpeluang mengesan masalah teknikal atau logistik yang mungkin berlaku semasa pengumpulan data sebenar. Kajian oleh Ziyang Wang et al. (2025), misalnya dalam membina instrumen literasi teknologi dalam kalangan guru, menunjukkan bahawa ujian rintis telah membantu dalam mengenal pasti istilah yang mengelirukan serta kekaburan pada format skala yang digunakan.

Setelah data daripada ujian rintis diperoleh, ia perlu dianalisis secara statistik untuk menilai kualiti psikometrik setiap item. Analisis ini merangkumi dua aspek utama iaitu kesahan konstruk dan kebolehpercayaan. Untuk menilai kesahan konstruk, penyelidik lazimnya menggunakan pendekatan seperti Analisis Faktor Eksploratori (EFA) bagi mengenal pasti struktur faktor asas dalam data, dan Analisis Faktor Sahkan (CFA) jika struktur konstruk telah ditentukan terlebih dahulu berdasarkan teori. Kaedah EFA membolehkan penyelidik meneliti sama ada item-item berkumpul di bawah dimensi yang sama seperti yang dijangkakan, manakala CFA pula menguji sama ada model yang dicadangkan sesuai dengan data sebenar. Bagi menilai kebolehpercayaan pula, nilai Alpha Cronbach digunakan untuk mengukur konsistensi dalaman sesuatu dimensi. Nilai kebolehpercayaan yang melebihi 0.70 secara umumnya dianggap memuaskan, manakala nilai antara 0.60 hingga 0.69 masih boleh diterima dalam peringkat eksploratori. Beberapa kajian juga mula menggunakan indeks kebolehpercayaan alternatif seperti Composite Reliability (CR) dan Average Variance Extracted (AVE), terutama dalam model pengukuran berasaskan CFA.

Keseluruhannya, proses ujian rintis dan analisis statistik bukan sahaja membantu membuktikan bahawa instrumen tersebut sah dan boleh dipercayai, malah memberikan asas yang kukuh bagi memuktamadkan versi akhir instrumen sebelum digunakan secara rasmi dalam kajian utama. Dalam era kecerdasan buatan, ketepatan analisis ini menjadi lebih penting kerana konstruk yang diukur selalunya melibatkan konsep baharu seperti etika digital, kebolegunaan AI, atau kesedaran teknologi yang masih berkembang. Oleh itu, penyelidik perlu memastikan bahawa setiap item benar-benar berupaya mengukur konstruk yang dimaksudkan dengan cara yang stabil, jelas, dan sesuai dengan konteks pendidikan digital semasa.

Penambahbaikan dan Pemuktamadan Instrumen

Setelah analisis statistik dijalankan ke atas data ujian rintis, penyelidik akan melaksanakan penambahbaikan terhadap item-item instrumen berdasarkan prestasi psikometrik dan maklum balas responden. Item yang tidak menunjukkan korelasi memuaskan, mempunyai beban faktor rendah, atau menurunkan nilai kebolehpercayaan keseluruhan akan dipertimbangkan untuk digugurkan, diubah suai, atau diganti. Penilaian semula ini juga mengambil kira aspek kejelasan bahasa, kesesuaian konteks, serta kesetaraan interpretasi oleh kumpulan sasaran. Langkah ini penting untuk memastikan semua item benar-benar mewakili konstruk secara sah dan konsisten, seiring dengan objektif pengukuran.

Pemuktamadan akhir instrumen bukan hanya bersifat teknikal, tetapi turut dipengaruhi oleh pertimbangan pedagogi dan nilai etika, terutama dalam konteks instrumen berkaitan kecerdasan buatan. Kajian oleh Yueqiao Jin et al. (2024) dan Ziyang Wang et al. (2025) menunjukkan bahawa proses ini memperkukuh keabsahan instrumen sebelum digunakan dalam kajian utama atau penilaian berskala besar. Namun begitu, pemuktamadan bukanlah titik akhir; pemantauan berterusan masih diperlukan bagi menyesuaikan instrumen dengan perubahan semasa dalam teknologi, dasar pendidikan, dan keperluan pengguna sebenar.

Pelaksanaan Instrumen dan Penilaian

Setelah instrumen dimuktamadkan melalui proses penambahbaikan dan pengesahan psikometrik, langkah terakhir dalam kitaran pembangunan ialah pelaksanaannya dalam konteks sebenar. Pelaksanaan ini merujuk kepada penggunaan instrumen untuk mengumpul data dalam skala lebih besar dan dalam suasana yang mencerminkan persekitaran dunia

sebenar, seperti bilik darjah, latihan guru, atau kajian lapangan. Tujuan utama pelaksanaan adalah untuk menilai keberkesanan instrumen dalam mengukur konstruk sasaran serta mendapatkan maklumat bermakna bagi tujuan penilaian, intervensi, atau penyelidikan lanjut. Dalam konteks pendidikan berasaskan kecerdasan buatan, pelaksanaan instrumen juga membolehkan pendidik dan pembuat dasar memahami tahap literasi AI, etika penggunaan teknologi, dan kebolehpayaan digital dalam kalangan pengguna pendidikan.

Kesimpulan

Perkembangan pesat teknologi kecerdasan buatan (AI) telah membawa transformasi signifikan terhadap pendekatan pendidikan, khususnya dalam aspek penilaian. Penggunaan aplikasi seperti ChatGPT dalam kalangan pelajar telah mencetuskan keperluan baharu terhadap instrumen yang mampu menilai bukan sahaja tahap pengetahuan dan kemahiran, tetapi juga nilai integriti akademik serta keupayaan membuat pertimbangan etika secara kritis. Justeru, pembinaan instrumen dalam era AI perlu dilaksanakan secara sistematik, berasaskan teori dan disokong oleh pendekatan reka bentuk penilaian yang holistik. Kertas konsep ini telah menghuraikan pelbagai komponen utama dalam pembinaan instrumen termasuk kerangka teori seperti konstruktivisme dan etika digital, pemilihan model pembangunan seperti ADDIE, penjanaan dan pengujian item, serta pelaksanaan dan penilaian instrumen dalam konteks dunia sebenar. Implikasi daripada usaha ini adalah besar terhadap ekosistem pendidikan masa kini. Instrumen yang dibangunkan berpotensi untuk digunakan secara meluas dalam sistem pentaksiran sekolah, program latihan guru, dan juga dalam penyelidikan pendidikan yang berkaitan dengan keberkesanan penggunaan AI. Ia boleh menjadi alat diagnostik dalam menilai tahap kesedaran etika pelajar, mengenal pasti keperluan intervensi nilai, serta membimbing pembangunan kurikulum yang lebih seimbang antara kemahiran digital dan tanggungjawab moral. Di samping itu, instrumen ini dapat menyokong pembuat dasar dalam merangka panduan penggunaan AI di peringkat sekolah yang selaras dengan prinsip akauntabiliti dan keadilan pendidikan.

Sehubungan itu, kajian lanjutan dicadangkan agar memberi tumpuan kepada perancangan dan pembinaan instrumen ini dalam pelbagai latar pendidikan serta kumpulan sasaran yang berbeza, termasuk pelajar sekolah rendah, menengah, dan institusi pengajian tinggi. Kajian jangka panjang juga wajar dilaksanakan untuk menelusuri perubahan dalam sikap, pemahaman, dan amalan pelajar terhadap penggunaan AI dari semasa ke semasa. Di samping itu, pendekatan kualitatif seperti temu bual mendalam atau penulisan reflektif pelajar boleh digunakan untuk memperkayakan kefahaman terhadap bagaimana konstruk seperti integriti akademik dan kebertanggungjawaban diterjemahkan dalam tindakan sebenar. Kajian-kajian lanjutan ini bukan sahaja penting untuk menilai kesahan dan kebolegunaan instrumen secara berterusan, malah dapat memastikan instrumen yang dibangunkan kekal relevan, kontekstual dan responsif terhadap cabaran pendidikan digital yang sentiasa berkembang.

Rujukan

- Delcker, J., Heil, J., & Ifenthaler, D. (2024). Evidence-based development of an instrument for the assessment of teachers' self-perceptions of their artificial intelligence competence. *Educational Technology Research and Development*, 73(1), 115–133. <https://doi.org/10.1007/s11423-024-10418-1>
- Guo, S., Shi, L., & Zhai, X. (2024). Validating an Instrument for Teachers' Acceptance of Artificial Intelligence in Education. *ArXiv E-Prints*, 1–10. <http://arxiv.org/abs/2406.10506>
- Hamed Abdelreheem Ead. (2024). Exploring the Impact of Artificial Intelligence on Academic Writing: Perspectives of Ph.D. Students in the Faculty of Science at Cairo University. *SunText Review of Medical & Clinical Research*, 5(4), 1–9. <https://doi.org/10.51737/2766-4813.2024.108>
- Knoth, N., Decker, M., Laupichler, M. C., Pinski, M., Buchholtz, N., Bata, K., & Schultz, B. (2024). Developing a holistic AI literacy assessment matrix – Bridging generic, domain-specific, and ethical competencies. *Computers and Education Open*, 6(April), 1–14. <https://doi.org/10.1016/j.caeo.2024.100177>
- Mohd Alif Aineh, & Wirawati Ngui. (2024). Teachers' And Students' Perceptions Towards The Use Of Chatgpt To Improve Writing In The Malaysian Secondary School Context. *International Journal on E-Learning Practices (IJELP)*, 7(1). <https://doi.org/10.51200/ijelp.v7i1.5399>

- Natalia Sevchenko, Tobias Appel, Manuel Ninaus, Moeller, K., & Peter Gerjets. (2023). Theory-based approach for assessing cognitive load during time-critical resource-managing human–computer interactions: an eye-tracking study. *Journal on Multimodal User Interfaces*, 17(1), 1–19. <https://doi.org/10.1007/s12193-022-00398-y>
- Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Bich-Phuong Thi Nguyen. (2023). Ethical principles for artificial intelligence in K-12 education. *Education and Information Technologies*, 4(28), 100131. <https://doi.org/10.1016/j.caeai.2023.100131>
- Rozo García, H., Alcantar Nieblas, C., & Ramírez Montoya, M. S. (2024). Scale to measure student perception in collaborative online international learning experiences: design and validation. *Frontiers in Education*, 9, 1–10. <https://doi.org/10.3389/feduc.2024.1401295>
- Samuel Boateng, Gilbert Kalonde Gilbert, & Caludia Duedu. (2024). Augmenting Academic Efficiency: The Integration of Assessments Within the Addie Model for Pedagogical Development. *Proceedings of The International Conference on Advanced Research in Education, Teaching, and Learning*, 1(1), 1–12. <https://doi.org/10.33422/aretl.v1i1.186>
- Sigit Wibowo, Muhammad Nur Wangid, & Fery Muhamad Firdaus. (2025). The relevance of Vygotsky's constructivism learning theory with the differentiated learning primary schools. *Journal of Education and Learning*, 19(1), 431–440. <https://doi.org/10.11591/edulearn.v19i1.21197>
- Tang, C. M., & Chaw, L. Y. (2023). What Have People Discussed about ChatGPT in Malaysian Education? A Qualitative Content Analysis of News Articles. *European Conference on E-Learning*, 22(1), 314–321. <https://doi.org/10.34190/ecel.22.1.1557>
- Yeliena Prokhorova, Rashmi Gujrati, & Hayri Uygün. (2024). The Use Of Ai Chatbots In Higher Education: The Problem Of Plagiarism. *Review of Artificial Intelligence in Education*, 5(May), 1–8. <https://doi.org/https://doi.org/10.37497/rev.artif.intell.educ.v5i00.31>
- Yueqiao Jin, Roberto Martinez Maldonado, Dragan Gašević, & Lixiang Yan. (2024). *GLAT: The Generative AI Literacy Assessment Test*. 4984, 1–20. <https://doi.org/10.1016/j.caeai.2025.100436>
- Ziying Wang, Ching Sing Chai, Jiajing Li, & Vivian Wing Yan. (2025). Assessment of AI ethical reflection: the development and validation of the AI ethical reflection scale (AIERS) for university students. *International Journal of Educational Technology in Higher Education*, 22(19), 1–16. <https://doi.org/10.1186/s41239-025-00519-z>