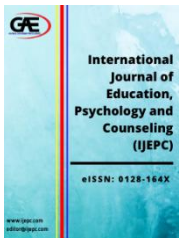
International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1832-1847



INTERNATIONAL JOURNAL OF EDUCATION, PSYCHOLOGY AND COUNSELLING (IJEPC)
https://gaexcellence.com/ijepc

KEBERKESANAN PENGGUNAAN NOTEBOOKLM SEBAGAI ALAT SOKONGAN PEMBELAJARAN BERASASKAN AI TERHADAP PENGUASAAN NAHU DAN SARF PELAJAR STAM: SATU KAJIAN TINDAKAN

THE EFFECTIVENESS OF USING NOTEBOOKLM AS AN AI-BASED LEARNING SUPPORT TOOL ON STAM STUDENTS' MASTERY OF NAHU AND SARF: AN ACTION RESEARCH

Muhammad Shamie Che Mahamood[1]*, Hazrati Husnin[2]

[1]Fakulti Pendidikan, Universiti Kebangsaan Malaysia, Malaysia
p165846@siswa.ukm.edu.my https://orcid.org/0009-0002-2133-8708
[2]Fakulti Pendidikan, Universiti Kebangsaan Malaysia, Malaysia
hazrati@ukm.edu.my https://orcid.org/0000-0002-6165-2411
*Corresponding Author

Article Info:

Article history:
Received date: 13.08.2026
Revised date: 30.08.2026
Accepted date: 10.09.2026
Published date: 30.09.2026

To cite this document:

Mahamood, M. S. C., & Husnin, H. (2026). Keberkesanan Penggunaan NotebookLM Sebagai Alat Sokongan Pembelajaran Berasaskan AI Terhadap Penguasaan Nahu Dan Sarf Pelajar Stam: Satu Kajian Tindakan. *International Journal*

Abstrak:

Penguasaan Nahu dan Sarf merupakan cabaran berterusan dalam kalangan pelajar Sijil Tinggi Agama Malaysia (STAM), sedangkan kaedah pengajaran konvensional kurang menyediakan bahan sokongan interaktif yang boleh digunakan pelajar semasa mengulang kaji di luar bilik darjah. Kajian tindakan ini menilai keberkesanan penggunaan NotebookLM, iaitu aplikasi kecerdasan buatan yang berasaskan sumber (*source-grounded*), sebagai alat sokongan pembelajaran terhadap penguasaan Nahu dan Sarf. Seramai 22 orang pelajar kelas 6 Ibnu Batutah di Kolej Tahfiz Sains Nurul Aman dipilih melalui persampelan bertujuan. Reka bentuk kajian menggunakan Model Gelung Kemmis dan McTaggart yang melibatkan fasa perancangan, tindakan, pemerhatian dan refleksi, dengan dapatan refleksi digunakan untuk menambah baik pelaksanaan intervensi secara berperingkat. Instrumen kajian ialah ujian pra dan ujian pasca dalam bentuk selari, manakala intervensi dijalankan selama lima minggu menggunakan ciri Infografik, Video Overview, Slide Deck, Kuiz dan Kad Imbasan secara berperingkat. Data dianalisis menggunakan statistik deskriptif serta ujian-t berpasangan bagi menentukan kesignifikanan statistik peningkatan markah. Dapatan menunjukkan markah 20 daripada 22

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1832-1847

of Education, Psychology and Counselling, 11(64), 1832-1847.

DOI: 10.35631/IJEPC.1164101

responden (90.9%) meningkat, dengan skor min naik daripada 15.00 (37.5%) kepada 23.41 (58.5%), iaitu peningkatan kira-kira 21 mata peratus. Ujian-t berpasangan menunjukkan peningkatan ini signifikan secara statistik, t(21) = 6.997, p < .001, dengan saiz kesan yang besar (d = 1.49). Seorang responden menurun manakala seorang lagi kekal pada markah asal. Bilangan responden pada tahap Sangat Lemah berkurangan daripada 14 orang kepada 2 orang, dan 9 orang mencapai tahap Baik. Kajian ini menyimpulkan bahawa NotebookLM berpotensi menyokong pembelajaran Nahu dan Sarf kerana responsnya berasaskan sumber rujukan yang dibekalkan oleh pengguna. Walau bagaimanapun, ketepatan sumber tidak menjamin ketepatan tulisan Arab yang dipaparkan, jadi semakan guru terhadap bahan yang dijana kekal diperlukan.

Kata Kunci:

Alat Sokongan Pembelajaran, Kajian Tindakan, Kecerdasan Buatan, Nahu Dan Sarf, NotebookLM, STAM

Abstract:

Mastery of Nahu and Sarf (Arabic grammar and morphology) remains a persistent challenge among Sijil Tinggi Agama Malaysia (STAM) students, while conventional teaching provides limited interactive support material for students revising outside the classroom. This action research evaluates the effectiveness of using NotebookLM, a *source-grounded* artificial intelligence application, as a learning support tool for the mastery of Nahu and Sarf. Twenty-two students from the 6 Ibnu Batutah class at Kolej Tahfiz Sains Nurul Aman were selected through purposive sampling. The study used the Kemmis and McTaggart model, covering the planning, action, observation, and reflection phases, with reflection findings used to refine the intervention in stages. The instruments were a pre-test and a post-test in parallel forms, while the five-week intervention introduced the Infographic, Video Overview, Slide Deck, Quiz, and Flashcard features in stages. Data were analysed using descriptive statistics and a paired-samples t-test to determine the significance of the score increase. The findings show that the scores of 20 of the 22 respondents (90.9%) improved, with the mean score rising from 15.00 (37.5%) to 23.41 (58.5%), an increase of about 21 percentage points. The paired-samples t-test showed that this increase was statistically significant, t(21) = 6.997, p < .001, with a large effect size (d = 1.49). One respondent declined while another remained at the same score. The number of respondents at the Very Weak level fell from 14 to two, and nine reached the good level. The study concludes that NotebookLM has the potential to support the learning of Nahu and Sarf because its responses are grounded in the reference sources supplied by the user. Nevertheless, source accuracy does not guarantee the accuracy of the Arabic script displayed, so teacher verification of the generated materials remains necessary.

Keywords:

Action Research, Artificial Intelligence, Learning Support Tool, Nahu and Sarf, NotebookLM, STAM

© 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 ijepc@gaexcllence.com

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1832-1847

Pengenalan

Perkembangan teknologi kecerdasan buatan (AI) berkembang pesat dalam bidang pendidikan sejak beberapa tahun kebelakangan ini. Perkembangan ini turut mendapat sokongan pada peringkat dasar. Konsensus Beijing menyeru penyepaduan AI dan pendidikan secara sistematik bagi memacu inovasi dalam pengajaran dan pembelajaran (UNESCO, 2019), manakala Dasar Pendidikan Digital menyasarkan integrasi teknologi dalam ekosistem pendidikan negara (Kementerian Pendidikan Malaysia, 2023). UNESCO (2024) seterusnya menggariskan kerangka kompetensi AI untuk guru bagi memastikan integrasi ini dilaksanakan secara berkesan dan beretika.

Dalam persekitaran pendidikan, AI generatif kini dilihat sebagai pemangkin pedagogi abad ke-21 dan bukan sekadar pelengkap pengajaran (Manikam et al., 2025). Alat seperti ChatGPT dan NotebookLM digunakan secara meluas oleh guru dan pelajar kerana keupayaannya menyediakan pembelajaran diperibadikan, sokongan berperingkat dan maklum balas segera (Andriyana et al., 2026). Sharjeel Ahmed et al. (2025) menegaskan bahawa keberkesanan integrasi AI banyak bergantung kepada literasi dan kesediaan guru, bukan semata-mata kepada keupayaan teknologi itu sendiri.

Di sebalik potensi tersebut, penggunaan AI generatif turut membawa cabaran yang perlu diberi perhatian. Kebergantungan berlebihan boleh menyebabkan pelajar berhenti berusaha menganalisis maklumat secara sendiri, dan Adnan (2025) memberi amaran bahawa keadaan ini boleh membawa kepada pengabaian rujukan yang lebih muktabar. Isu ketepatan turut menjadi kebimbangan, kerana AI generatif terbuka menjana jawapan berdasarkan pengetahuan latihannya dan tidak sentiasa boleh dijejaki kepada sumber yang sah.

Cabaran ini menjadi lebih ketara dalam pembelajaran bahasa Arab. Bahasa Arab mempunyai sistem morfologi yang kaya, struktur sintaksis yang berlapis serta penggunaan iʿrab yang menentukan makna ayat. Borham et al. (2022) mengenal pasti kekaburan ortografi, morfologi dan sintaksis bahasa Arab sebagai cabaran asas dalam pembangunan perisian pendidikan berasaskan AI. Haryati Jaacob et al. (2026) pula mendapati ChatGPT mampu menghasilkan ayat yang relatif gramatis pada peringkat asas, tetapi kurang konsisten apabila berhadapan dengan struktur ayat yang rumit dan penggunaan iʿrab yang tepat. Sungguhpun demikian, Rahem Samiya (2025), Fitri (2025) serta Ali dan Rani (2024) melaporkan bahawa AI berpotensi menyokong pembelajaran bahasa Arab melalui maklum balas segera, penyesuaian kandungan dan latihan kendiri.

NotebookLM menawarkan pendekatan yang berbeza daripada AI generatif terbuka. Andriyana et al. (2026) menjelaskan bahawa aplikasi ini menggunakan kaedah *Retrieval-Augmented Generation* yang mengikat jawapan kepada dokumen sumber yang dimuat naik pengguna, sekali gus membolehkan jawapan dijejaki kepada sumber asal. Aplikasi ini turut menyediakan sokongan pelbagai media seperti infografik, ringkasan video, slaid, kuiz dan kad imbasan. Ciri bersandarkan sumber ini menjadikannya relevan bagi bidang yang menuntut ketepatan hukum seperti Nahu dan Sarf.

Dalam konteks pengajaran pengkaji sendiri, penguasaan Nahu dan Sarf dalam kalangan pelajar Sijil Tinggi Agama Malaysia (STAM) berada pada tahap yang membimbangkan. Purata markah Kertas 1 bagi 22 orang pelajar kelas 6 Ibnu Batutah hanya 37.5%. Keadaan ini kritikal kerana Nahu dan Sarf menentukan ketepatan pembinaan ayat dan kefahaman teks Arab,

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1832-1847

sedangkan sukatan STAM secara keseluruhannya menuntut kemahiran tersebut. Kaedah konvensional yang diamalkan sebelum ini pula tidak menyediakan bahan sokongan yang membolehkan pelajar menyemak kefahaman sendiri apabila mengulang kaji di luar bilik darjah.

Kajian sedia ada mengenai AI dalam pembelajaran bahasa Arab sebahagian besarnya menumpukan kepada alat AI umum tanpa membezakan cara jawapan dijana, dan dapatannya bercanggah antara potensi pedagogi dengan ketepatan hukum. Setakat semakan pengkaji, belum ada kajian yang menguji sama ada AI berasaskan *Retrieval-Augmented Generation* memberi implikasi berbeza terhadap pembelajaran tatabahasa Arab berbanding AI yang menjana jawapan secara bebas. Kajian mengenai NotebookLM pula masih berbentuk tinjauan literatur (Andriyana et al., 2026). Kajian ini mengisi jurang tersebut, dengan konteks Nahu dan Sarf STAM dipilih kerana bidang ini menuntut ketepatan hukum yang tinggi.

Latar Belakang Kajian

Sijil Tinggi Agama Malaysia (STAM) merupakan peperiksaan awam bagi pelajar aliran agama yang dilaksanakan sejak tahun 2000 hasil kerjasama Kementerian Pendidikan Malaysia dengan Universiti al-Azhar, Kaherah (Zainal Rafit et al., 2021). Pelajar dituntut menguasai sukatan Nahu dan Sarf yang luas pada aras yang tinggi dalam tempoh pengajian yang singkat.

Kajian ini dijalankan di Kolej Tahfiz Sains Nurul Aman dan melibatkan 22 orang pelajar kelas 6 Ibnu Batutah. Markah Kertas 1 dalam Peperiksaan Pertengahan Tahun 2026 pada bulan Mei menunjukkan 14 orang pelajar (64%) memperoleh markah di bawah 40%, manakala markah tertinggi hanya 57.5% dan tiada seorang pun mencapai 60% ke atas. Markah rendah inilah yang mencetuskan kajian ini, dan markah tersebut dijadikan ujian pra.

Sebelum ini pengajaran dilaksanakan secara konvensional melalui kaedah *chalk and talk* yang tertumpu kepada terjemahan teks dan menjawab soalan pelajar. Pelajar memang mengulang kaji dan menyiapkan kerja rumah di rumah, tetapi mereka tidak mempunyai bahan sokongan yang membolehkan mereka menyemak kefahaman sendiri apabila menghadapi kekeliruan di luar waktu pengajaran.

Bagi menangani masalah ini, NotebookLM dipilih untuk diuji sebagai alat sokongan pembelajaran. Guru memuat naik Buku Teks Nahu dan Sarf serta DSKP Nahu dan Sarf STAM ke dalam aplikasi supaya respons yang dijana adalah berasaskan dokumen sumber yang selaras dengan sukatan. Pelajar seterusnya menggunakan ciri Infografik, Video Overview, Slide Deck, Kuiz dan Kad Imbasan untuk mengulang kaji. Kajian ini menilai sama ada ciri tersebut dikaitkan dengan perubahan pencapaian pelajar.

Objektif Kajian

Objektif kajian ini adalah untuk:

i. Membandingkan pencapaian pelajar dalam ujian pra dan ujian pasca.
ii. Mengenal pasti keberkesanan penggunaan NotebookLM berdasarkan perubahan tahap penguasaan pelajar.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1832-1847

Persoalan Kajian

Kajian ini dipandu oleh dua persoalan berikut:

i. Apakah perbezaan pencapaian pelajar antara ujian pra dan ujian pasca?
ii. Sejauh manakah keberkesanan penggunaan NotebookLM berdasarkan perubahan tahap penguasaan pelajar?

Sorotan Literatur

Kajian mengenai AI semakin mendapat perhatian kerana potensinya sebagai alat sokongan pembelajaran (Manikam et al., 2025; Sharjeel Ahmed et al., 2025). Manikam et al. (2025) merumuskan bahawa alat berasaskan AI berpotensi menyokong pembelajaran diperibadikan, namun kejayaannya bergantung kepada kecekapan guru dan sokongan institusi. Sharjeel Ahmed et al. (2025) pula menegaskan bahawa keberkesanan integrasi AI banyak bergantung kepada literasi dan kesediaan guru. Kedua-dua kajian ini bersifat konseptual dan tidak mengukur kesan sebenar terhadap pencapaian pelajar.

Dalam konteks bahasa Arab, Rahem Samiya (2025) mendapati AI menawarkan pembelajaran diperibadikan, maklum balas segera dan peningkatan motivasi. Fitri (2025) menyatakan pembelajaran adaptif berasaskan AI berpotensi menangani kerumitan struktur bahasa Arab. Ali dan Rani (2024) mengenal pasti lima tema penggunaan AI, iaitu pembentukan ayat, terjemahan, penandaan baris, pembelajaran interaktif dan pembelajaran kendiri. Khazri Osman et al. (2025) pula mengenal pasti manfaat AI dalam menjelaskan kandungan pelajaran menerusi perspektif guru. Keempat-empat kajian ini menumpu kepada mekanisme yang hampir sama, tetapi berbeza dari segi sumber data. Rahem Samiya (2025) dan Fitri (2025) menggunakan tinjauan literatur, Khazri Osman et al. (2025) menggunakan pandangan guru, manakala Ali dan Rani (2024) menggunakan analisis kandungan. Tiada satu pun menggunakan markah pencapaian pelajar. Dakwaan tentang keberkesanan AI dalam bidang ini masih belum diuji secara langsung.

Ketepatan AI menjadi persoalan penting dalam bidang yang menuntut kepatuhan kepada hukum tatabahasa. Haryati Jaacob et al. (2026) mendapati ChatGPT mampu menghasilkan ayat yang relatif gramatis pada peringkat asas, tetapi kurang konsisten apabila berhadapan dengan struktur ayat yang rumit dan penggunaan iʿrab. Borham et al. (2022) mengenal pasti kekaburan ortografi, morfologi dan sintaksis bahasa Arab sebagai cabaran utama. Adnan (2025) pula memberi amaran bahawa kebergantungan terhadap AI boleh menyebabkan pelajar meninggalkan kitab-kitab klasik. Sehubungan itu, alat AI yang responsnya berasaskan dokumen sumber terkawal lebih wajar digunakan.

NotebookLM menawarkan pendekatan yang berbeza. Andriyana et al. (2026) menjelaskan bahawa aplikasi ini menggunakan kaedah *Retrieval-Augmented Generation* (RAG) yang mengikat jawapan kepada dokumen sumber yang dimuat naik pengguna. Pendekatan ini mengurangkan risiko halusinasi dan membolehkan jawapan dijejaki kepada sumber asal. Aplikasi ini turut menyediakan sokongan pelbagai media seperti infografik, ringkasan video, slaid, kuiz dan kad imbasan. Literatur memperlihatkan dua kecenderungan yang berbeza. Satu kumpulan kajian merumuskan AI menyokong pembelajaran bahasa Arab (Rahem Samiya, 2025; Fitri, 2025; Ali dan Rani, 2024). Satu kumpulan lagi menunjukkan AI generatif terbuka kurang konsisten dari segi hukum (Haryati Jaacob et al., 2026) dan berisiko menjejaskan

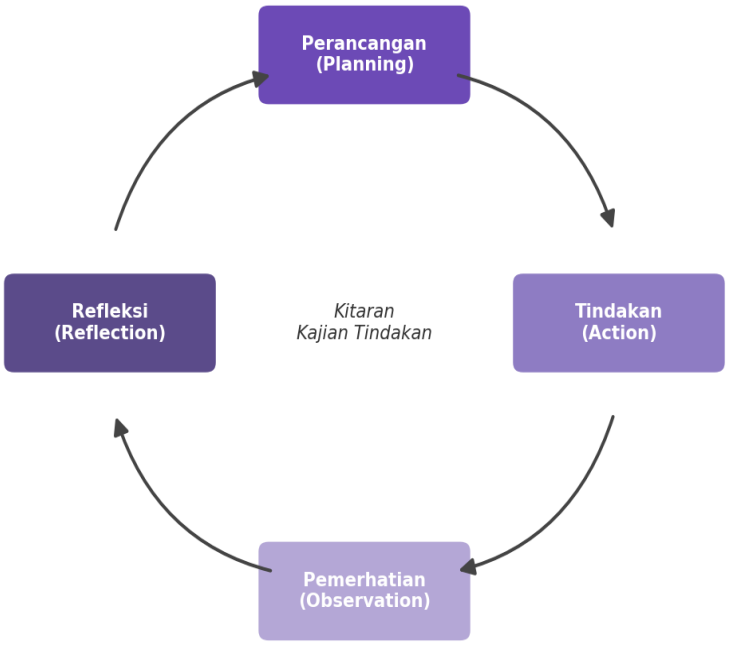
International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1832-1847

rujukan muktabar (Adnan, 2025). Borham et al. (2022) pula menunjukkan kekaburan bahasa Arab itu sendiri merupakan cabaran asas dalam pembangunan perisian berasaskan AI. Kedua-dua kumpulan ini sukar didamaikan selagi ketepatan sumber tidak dijadikan pemboleh ubah kajian.

Dari segi reka bentuk, Azmi dan Ruzanna (2026) menggunakan kajian tindakan Model Kemmis dan McTaggart dengan ujian pra dan ujian pasca bagi menilai ChatGPT dalam pembelajaran tatabahasa Bahasa Melayu, dan mendapati markah 77.8% responden meningkat. Kajian tersebut menjadi rujukan reka bentuk kepada kajian ini, tetapi ia menggunakan AI terbuka dan tidak melibatkan bidang yang menuntut ketepatan hukum. Aminun Nabil Ahmad Syahir (2025) turut menegaskan bahawa kajian integrasi AI dalam pengajian al-Quran masih terhad dan bersifat deskriptif.

Metodologi Kajian

Kajian ini menggunakan reka bentuk kajian tindakan Model Gelung Kemmis dan McTaggart (1988) yang melibatkan empat fasa, iaitu perancangan, tindakan, pemerhatian dan refleksi, seperti dalam Rajah 1 dan diringkaskan dalam Jadual 1. Bahagian ini menghuraikan pelaksanaan kajian mengikut keempat-empat fasa tersebut.



Rajah 1: Model Gelung Kemmis dan McTaggart

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1832-1847

Jadual 1: Proses Kajian Menggunakan Model Kemmis dan McTaggart

Fasa	Huraian
1. Perancangan (*Planning*)	Mengenal pasti masalah penguasaan Nahu dan Sarf berdasarkan markah Kertas 1 Peperiksaan Pertengahan Tahun 2026. Merancang penggunaan NotebookLM dan menyediakan sumber terkawal iaitu Buku Teks dan DSKP Nahu dan Sarf STAM.
2. Tindakan (*Action*)	Mengintegrasikan NotebookLM dalam pengajaran Nahu dan Sarf selama lima minggu, mengikut ciri aplikasi secara berperingkat.
3. Pemerhatian (*Observation*)	Mengumpul markah ujian pra dan ujian pasca serta memerhatikan penglibatan pelajar sepanjang sesi intervensi.
4. Refleksi (*Reflection*)	Menganalisis perubahan pencapaian dan tahap penguasaan, seterusnya merumuskan penambahbaikan bagi kitaran seterusnya.

Fasa Perancangan

Pengkaji mengenal pasti masalah penguasaan Nahu dan Sarf yang lemah berdasarkan markah Kertas 1 dalam Peperiksaan Pertengahan Tahun 2026 pada bulan Mei. Pengkaji seterusnya merancang penggunaan NotebookLM sebagai alat sokongan dan menyediakan sumber terkawal, iaitu Buku Teks Nahu dan Sarf serta DSKP Nahu dan Sarf STAM.

Kajian melibatkan 22 orang pelajar STAM kelas 6 Ibnu Batutah di Kolej Tahfiz Sains Nurul Aman. Persampelan bertujuan digunakan, iaitu pemilihan responden berdasarkan kriteria yang ditetapkan pengkaji. Responden dipilih kerana mereka diajar oleh pengkaji dan menunjukkan penguasaan yang lemah.

Instrumen kajian ialah ujian pra dan ujian pasca dalam bentuk selari, dengan markah penuh masing-masing 40 markah (*parallel forms*). Ujian pra ialah markah Kertas 1 Peperiksaan Pertengahan Tahun 2026 yang telah diduduki pelajar pada bulan Mei. Ujian pasca menggunakan kertas Kertas 1 Peperiksaan Pertengahan Tahun 2025 yang dijalankan selepas intervensi. Kedua-dua kertas disediakan oleh Majlis Pengetua Sekolah-Sekolah Maahad/Agama Yayasan Islam Kelantan.

Kertas berpusat dipilih atas tiga pertimbangan. Pertama, kertas digubal berdasarkan sukatan dan format Kertas 1 Nahu dan Sarf STAM, jadi item yang diuji mewakili kandungan yang menjadi fokus kajian. Kedua, kertas berpusat digubal dan disemak oleh panel di peringkat majlis pengetua, iaitu proses yang lebih berstruktur berbanding ujian binaan guru. Ketiga, penggunaan kertas peperiksaan sebenar memastikan tugasan setara dengan tuntutan pentaksiran STAM.

Penggunaan dua kertas berbeza bertujuan mengelakkan kesan ingatan soalan. Kesetaraan aras kesukaran kedua-dua kertas diandaikan berdasarkan keseragaman sukatan dan format, tanpa pengesahan empirik seperti analisis indeks kesukaran item.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1832-1847

Fasa Tindakan

NotebookLM diintegrasikan dalam pengajaran Nahu dan Sarf selama lima minggu, bermula pada bulan Jun iaitu sebaik selepas Peperiksaan Pertengahan Tahun 2026. Pengkaji memuat naik sumber terkawal ke dalam aplikasi supaya respons yang dijana adalah berasaskan dokumen sumber yang boleh dijejaki. Setiap pelajar mengakses aplikasi menggunakan akaun Google masing-masing, jadi kuota penggunaan harian tidak dikongsi antara responden. Pelaksanaan dijalankan secara berperingkat mengikut ciri NotebookLM seperti dalam Jadual 2. Ujian pasca dijalankan selepas tamat intervensi.

Jadual 2: Prosedur Pelaksanaan Intervensi

Minggu	Aktiviti
Minggu 1	Pengenalan aplikasi NotebookLM dan cara meneroka sumber pembelajaran.
Minggu 2	Penggunaan ciri Infografik bagi membina peta minda konsep Nahu dan Sarf.
Minggu 3	Penggunaan ciri Video Overview bagi menjelaskan konsep secara visual.
Minggu 4	Penggunaan ciri Slide Deck bagi menyusun isi pembelajaran.
Minggu 5	Penggunaan ciri Kuiz dan Kad Imbasan bagi ulang kaji dan penilaian kendiri.

Fasa Pemerhatian

Pengkaji mengumpul data berkaitan proses dan kesan pelaksanaan intervensi. Data utama ialah markah ujian pra dan ujian pasca. Pengkaji turut memerhatikan tingkah laku dan penglibatan pelajar, namun pemerhatian ini tidak direkodkan menerusi senarai semak atau nota lapangan. Ia dijadikan asas kepada refleksi retrospektif guru penyelidik. Dengan itu, dapatan kuantitatif kajian bersandarkan markah ujian, manakala refleksi berfungsi sebagai maklumat sokongan.

Fasa Refleksi

Pengkaji menganalisis dan menilai semula keberkesanan 1839ormalit, seterusnya merumuskan penambahbaikan bagi kitaran seterusnya. Data markah dianalisis menggunakan 1839ormality deskriptif iaitu skor min, peratusan dan taburan tahap penguasaan, serta ujian-t berpasangan bagi menentukan kesignifikanan 1839ormality peningkatan markah berikutan andaian 1839ormality disahkan menerusi ujian Shapiro-Wilk. Data dipersembahkan dalam bentuk jadual dan carta bar. Markah pelajar ditafsirkan mengikut tahap penguasaan seperti dalam Jadual 3.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1832-1847

Jadual 3: Julat Markah dan Tahap Penguasaan

Julat Markah (%)	Tahap Penguasaan
80 – 100	Cemerlang
65 – 79	Baik
50 – 64	Sederhana
40 – 49	Lemah
0 – 39	Sangat Lemah

Dapatan Kajian

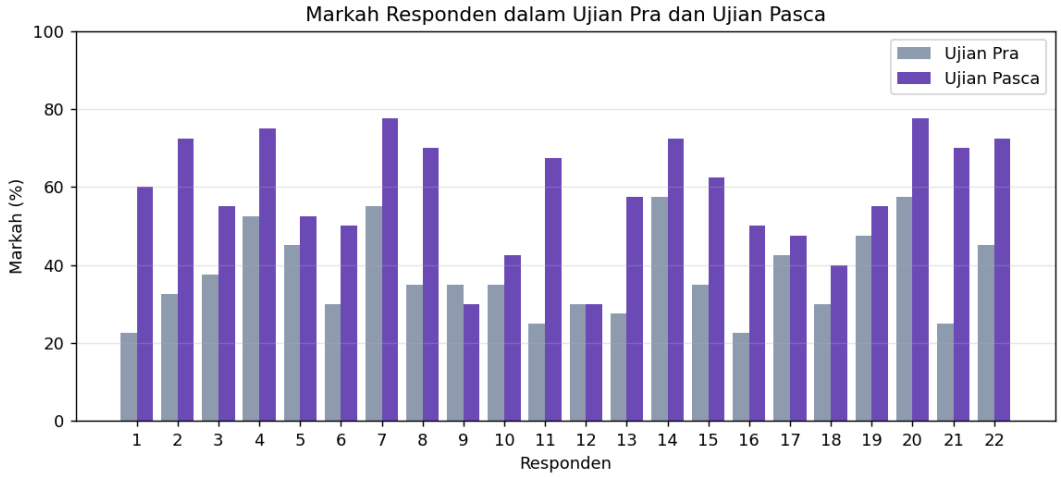
Daripada analisis data kajian, Jadual 4 dan Rajah 2 menunjukkan perbandingan markah responden dalam ujian pra dan ujian pasca. Markah dilaporkan dalam bentuk peratusan berdasarkan markah penuh 40 markah bagi Kertas 1 Nahu dan Sarf.

Jadual 4: Markah Responden dalam Ujian Pra dan Ujian Pasca

Responden	Ujian Pra (%)	Ujian Pasca (%)
1	22.5	60.0
2	32.5	72.5
3	37.5	55.0
4	52.5	75.0
5	45.0	52.5
6	30.0	50.0
7	55.0	77.5
8	35.0	70.0
9	35.0	30.0
10	35.0	42.5
11	25.0	67.5
12	30.0	30.0
13	27.5	57.5
14	57.5	72.5
15	35.0	62.5
16	22.5	50.0
17	42.5	47.5
18	30.0	40.0
19	47.5	55.0

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1832-1847

Responden	Ujian Pra (%)	Ujian Pasca (%)
20	57.5	77.5
21	25.0	70.0
22	45.0	72.5
Purata	**37.5**	**58.5**



Rajah 2: Markah Responden dalam Ujian Pra dan Ujian Pasca

Berdasarkan Jadual 4 dan Rajah 2, seramai 20 daripada 22 responden (90.9%) mencatatkan markah yang lebih tinggi dalam ujian pasca. Seorang responden menurun dan seorang lagi kekal pada markah yang sama. Statistik deskriptif bagi kedua-dua ujian ditunjukkan dalam Jadual 5.

Jadual 5: Statistik Deskriptif Ujian Pra dan Ujian Pasca

Ujian	N	Minimum	Maksimum	Min	Sisihan Piawai
Ujian Pra	22	9	23	15.00	4.49
Ujian Pasca	22	12	31	23.41	5.85

Jadual 5 menunjukkan skor min meningkat daripada 15.00 markah (37.5%) kepada 23.41 markah (58.5%), iaitu perbezaan sebanyak 8.41 markah atau 21.0 mata peratus. Markah minimum meningkat daripada 9 kepada 12, manakala markah maksimum meningkat daripada 23 kepada 31. Sisihan piawai turut meningkat daripada 4.49 kepada 5.85.

Bagi menentukan kesignifikanan statistik peningkatan markah ini, ujian normaliti Shapiro-Wilk dijalankan ke atas skor perbezaan (ujian pasca tolak ujian pra) untuk mengesahkan andaian normaliti sebelum ujian-t berpasangan dijalankan. Keputusan ujian normaliti ditunjukkan dalam Jadual 6.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1832-1847

Jadual 6: Keputusan Ujian Normaliti Shapiro-Wilk bagi Skor Perbezaan

N	Statistik Shapiro-Wilk (W)	Nilai p
22	0.972	0.767

Jadual 6 menunjukkan nilai p bagi ujian Shapiro-Wilk ialah 0.767 ($p > .05$), yang membuktikan skor perbezaan antara ujian pra dan ujian pasca adalah bertaburan normal. Oleh itu, ujian-t berpasangan (paired-samples t-test) sesuai digunakan untuk menguji kesignifikanan perbezaan min markah. Keputusan ujian-t berpasangan ditunjukkan dalam Jadual 7.

Jadual 7: Keputusan Ujian-t Berpasangan bagi Markah Ujian Pra dan Ujian Pasca

Min Perbezaan	SP Perbezaan	t	df	Nilai p	Cohen's d
8.41 markah	5.64	6.997	21	< .001	1.49

Nota. Min Perbezaan dikira sebagai Ujian Pasca tolak Ujian Pra.

Jadual 7 menunjukkan ujian-t berpasangan mendapati peningkatan markah min daripada ujian pra ($M = 15.00$, $SP = 4.49$) kepada ujian pasca ($M = 23.41$, $SP = 5.85$) adalah signifikan secara statistik, $t(21) = 6.997$, $p < .001$. Saiz kesan yang dikira menggunakan Cohen's d ialah 1.49, iaitu tergolong dalam kategori besar (large effect) menurut piawaian Cohen (1988). Dapatan ini memberikan bukti statistik bahawa peningkatan markah bukan sekadar perubahan deskriptif, sebaliknya mewakili peningkatan yang signifikan selepas pelaksanaan intervensi NotebookLM.

Taburan responden mengikut tahap penguasaan ditunjukkan dalam Jadual 8.

Jadual 8: Taburan Responden Mengikut Tahap Penguasaan

Tahap Penguasaan	Ujian Pra (Bil.)	Ujian Pra (%)	Ujian Pasca (Bil.)	Ujian Pasca (%)
Cemerlang	0	0.0	0	0.0
Baik	0	0.0	9	40.9
Sederhana	4	18.2	8	36.4
Lemah	4	18.2	3	13.6
Sangat Lemah	14	63.6	2	9.1
Jumlah	**22**	**100.0**	**22**	**100.0**

Berdasarkan Jadual 8, dalam ujian pra seramai 14 orang responden (63.6%) berada pada tahap Sangat Lemah, 4 orang (18.2%) pada tahap Lemah dan 4 orang (18.2%) pada tahap Sederhana. Dalam ujian pasca, 9 orang (40.9%) berada pada tahap Baik, 8 orang (36.4%) pada tahap Sederhana, 3 orang (13.6%) pada tahap Lemah dan 2 orang (9.1%) kekal pada tahap Sangat Lemah.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1832-1847

Perbincangan

Persoalan pertama kajian ini berkaitan perbezaan pencapaian antara ujian pra dan ujian pasca. Dapatan menunjukkan 20 daripada 22 responden (90.9%) mencatatkan peningkatan markah. Skor min meningkat daripada 15.00 (37.5%) kepada 23.41 (58.5%), iaitu perbezaan 8.41 markah atau 21.0 mata peratus. Ujian-t berpasangan mengesahkan bahawa peningkatan ini signifikan secara statistik, $t(21) = 6.997$, $p < .001$, dengan saiz kesan yang besar ($d = 1.49$). Seorang responden mencatatkan penurunan dan seorang lagi kekal pada markah asal.

Persoalan kedua pula berkaitan keberkesanan berdasarkan perubahan tahap penguasaan. Bilangan responden pada tahap Sangat Lemah berkurangan daripada 14 orang kepada 2 orang, manakala 9 orang mencapai tahap Baik yang sebelum ini tidak dicapai oleh sesiapa pun. Namun begitu, tiada seorang pun mencapai tahap Cemerlang, dan 5 orang masih berada pada tahap Lemah dan Sangat Lemah. Corak ini menunjukkan penggunaan NotebookLM dikaitkan dengan pengurangan bilangan pelajar pada tahap terendah, disokong oleh bukti statistik yang signifikan, tetapi belum memindahkan majoriti kepada tahap penguasaan yang tinggi.

Corak dapatan ini serupa dengan Azmi dan Ruzanna (2026) yang mendapati markah 77.8% responden meningkat, termasuk kewujudan responden yang markahnya tidak berubah atau menurun. Persamaan ini mengukuhkan rumusan mereka bahawa alat AI tidak memberi kesan yang seragam. Kadar peningkatan yang lebih tinggi dalam kajian ini (90.9%) berkemungkinan berkaitan dengan tahap penguasaan awal, kerana responden mereka terdiri daripada murid pintar dan berbakat yang ruang peningkatannya lebih terhad.

Dari sudut pedagogi, terdapat beberapa ciri yang berkemungkinan membantu pelajar. Kajian ini tidak mengumpul data penggunaan aplikasi seperti kekerapan capaian atau bilangan kuiz yang dijawab. Penjelasan berikut merupakan tafsiran berdasarkan ciri aplikasi dan bukan mekanisme yang disahkan daripada data. Ciri Kuiz menuntut pelajar mengingat semula maklumat dan bukan sekadar membaca nota. Ciri Kad Imbasan membolehkan pengulangan istilah dan tanda iʿrab mengikut kadar sendiri. Ciri Infografik, Video Overview dan Slide Deck pula menyampaikan kandungan yang sama dalam bentuk berbeza. Kesemuanya selari dengan pandangan Rahem Samiya (2025) serta Ali dan Rani (2024) tentang peranan maklum balas segera dan latihan kendiri. Oleh sebab pelajar bebas memilih ciri, sumbangan setiap ciri tidak dapat diasingkan.

Sifat berasaskan sumber (*source-grounded*) aplikasi ini relevan dengan konteks kajian. Respons yang dijana adalah berasaskan dokumen sumber Buku Teks dan DSKP Nahu dan Sarf STAM yang dimuat naik guru (Andriyana et al., 2026). Ini mengurangkan risiko pelajar menerima maklumat yang tidak tepat dari segi hukum, iaitu kelemahan yang dikenal pasti Haryati Jaacob et al. (2026) pada AI generatif terbuka.

Jaminan ini bagaimanapun terhad kepada kandungan dan bukan kepada persembahan visual. Pemerhatian guru penyelidik mendapati teks Arab dalam Infografik kadangkala dijana dengan baris yang berlebihan, titik yang tersalah letak dan bentuk huruf yang tidak menepati tulisan yang betul. Bagi Nahu dan Sarf, kesilapan sebegini bukan perkara kecil. Baris menentukan kedudukan iʿrab, manakala titik dan bentuk huruf menentukan identiti kata. Ketepatan sumber tidak menjamin ketepatan tulisan yang dipaparkan. Peranan guru sebagai penyemak bahan kekal diperlukan walaupun aplikasi yang digunakan bersandarkan sumber terkawal.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1832-1847

Penggunaan AI turut membawa risiko. Kebergantungan berlebihan boleh menyebabkan pelajar berhenti berusaha menganalisis hukum secara sendiri. Adnan (2025) memberi amaran bahawa keadaan ini boleh membawa kepada pengabaian rujukan muktabar seperti kitab-kitab klasik. Kemudahan menjana ringkasan dan kuiz secara automatik juga berisiko mengurangkan latihan membina nota dan membuat iʿrab secara manual. Sifat bersandarkan sumber mengurangkan risiko maklumat tidak tepat, tetapi tidak menghapuskan risiko kebergantungan. Atas sebab itu, guru mengekalkan peranan menuntut pelajar merujuk semula sumber asal.

Refleksi dan Penambahbaikan

Fasa refleksi menuntut pengkaji menilai semula tindakan dan menentukan penambahbaikan bagi kitaran berikutnya. Bahagian ini membentangkan refleksi retrospektif guru penyelidik terhadap pelaksanaan intervensi.

Cabaran paling ketara berlaku pada minggu pertama. Pelajar bukan sahaja perlu membiasakan diri dengan antara muka aplikasi, malah keliru tentang cara membina *prompt* yang betul. Prompt yang terlalu umum menghasilkan jawapan yang kurang berfokus. Guru terpaksa membimbing pelajar merumuskan soalan yang lebih khusus. Berdasarkan pemerhatian ini, kemahiran membina *prompt* perlu diajar secara langsung dan tidak boleh diandaikan sedia ada dalam kalangan pelajar.

Penglibatan pelajar semakin meningkat sepanjang intervensi berdasarkan pemerhatian guru. Pelajar lebih kerap bertanya tentang hukum Nahu dan Sarf, lebih banyak mencuba membina *prompt* sendiri, dan mula berbincang sesama rakan tentang jawapan yang dijana aplikasi. Menjelang minggu akhir, bimbingan guru dapat dikurangkan kerana pelajar sudah mampu menggunakan aplikasi secara berdikari. Peranan guru berubah daripada penyampai maklumat kepada pembimbing kemahiran menggunakan sumber. Pemerhatian ini bersifat kualitatif dan tidak dikuantifikasikan, jadi ia dilaporkan sebagai refleksi amalan.

Satu lagi pemerhatian penting berkaitan ciri Infografik. Teks Arab yang dijana kadangkala tidak tepat dari segi penulisan, seperti baris yang berlebihan, titik yang tersalah letak dan bentuk huruf yang tidak menepati tulisan yang betul. Guru terpaksa menyemak setiap bahan sebelum diedarkan kepada pelajar. Tanpa semakan tersebut, pelajar berkemungkinan menerima input yang salah dan membentuk kefahaman yang keliru terhadap hukum yang sedang dipelajari.

Dari segi hasil, majoriti responden menunjukkan peningkatan. Namun seorang responden menurun, seorang lagi kekal pada markah asal, dan 2 orang masih berada pada tahap Sangat Lemah. Intervensi berjaya mengangkat sebahagian besar pelajar tetapi belum menjangkau kesemuanya. Penggunaan aplikasi sahaja tidak memadai tanpa bimbingan yang lebih individu kepada pelajar yang ketinggalan.

Beberapa penambahbaikan dicadangkan bagi kitaran seterusnya. Pertama, analisis item mengikut topik perlu dijalankan bagi mengenal pasti komponen yang masih lemah. Kedua, tempoh intervensi wajar dipanjangkan. Ketiga, catatan pemerhatian dan refleksi pelajar perlu direkodkan secara lebih sistematik. Keempat, kesetaraan instrumen perlu disahkan secara empirik sebelum kitaran seterusnya. Kelima, setiap bahan visual yang dijana aplikasi perlu disemak guru terlebih dahulu bagi memastikan ketepatan tulisan Arab sebelum diedarkan kepada pelajar.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1832-1847

Kesimpulan

Dapatan kajian menunjukkan penggunaan NotebookLM dikaitkan dengan peningkatan penguasaan Nahu dan Sarf pelajar STAM. Markah 20 daripada 22 responden (90.9%) meningkat, dengan skor min naik daripada 37.5% kepada 58.5%, iaitu peningkatan yang signifikan secara statistik ($t(21) = 6.997$, $p < .001$, $d = 1.49$). Bilangan responden pada tahap Sangat Lemah berkurangan daripada 14 orang kepada 2 orang, manakala 9 orang mencapai tahap Baik. Namun begitu, seorang responden menurun dan seorang lagi kekal pada markah asal. Kesan intervensi tidak seragam bagi semua pelajar.

Sifat berasaskan sumber (*source-grounded*) aplikasi ini menjadikannya sesuai bagi bidang yang menuntut ketepatan hukum. Ciri pelbagai media seperti Infografik, Video Overview, Slide Deck, Kuiz dan Kad Imbasan pula menyediakan bahan sokongan yang boleh dirujuk pelajar semasa mengulang kaji di luar waktu pengajaran formal. Dari sudut sumbangan akademik, kajian ini memperluaskan bukti empirikal mengenai penggunaan AI generatif bersandarkan sumber dalam pembelajaran bahasa Arab, iaitu kategori alat yang kurang diuji berbanding AI generatif terbuka. Kajian ini turut menyumbangkan satu model pelaksanaan lima minggu yang boleh diadaptasi oleh guru lain.

Sebagai kajian tindakan, kajian ini bertujuan menyelesaikan masalah penguasaan yang dihadapi pelajar terlibat serta menambah baik amalan pengajaran pengkaji sendiri. Dalam konteks tersebut, NotebookLM wajar dilihat sebagai alat sokongan pedagogi yang perlu disertai bimbingan guru. Bagi kitaran seterusnya, tumpuan disarankan kepada topik yang masih lemah, tempoh yang lebih panjang, dan rekod pemerhatian yang lebih sistematik.

Penghargaan: Penulis merakamkan setinggi-tinggi penghargaan kepada Fakulti Pendidikan, Universiti Kebangsaan Malaysia atas sokongan dan kemudahan yang diberikan sepanjang kajian ini dijalankan

Penyataan Pembiayaan: Kajian ini tidak menerima sebarang pembiayaan.

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 Education, Psychology and Counselling (IJEPC)

Pernyataan Etika: Kajian ini telah dijalankan selaras dengan piawaian etika penyelidikan. Semua prosedur yang melibatkan responden manusia telah disemak dan diluluskan oleh Jawatankuasa Etika Penyelidikan UKM. Persetujuan bermaklumat telah diperoleh daripada semua responden sebelum pengumpulan data dijalankan. Penyertaan adalah secara sukarela, dan responden telah dijamin kerahsiaan serta anonimitas. Data yang dikumpulkan digunakan semata-mata bagi tujuan akademik.

Pernyataan Sumbangan Penulis: Semua penulis telah memberikan sumbangan yang signifikan terhadap pembangunan manuskrip ini. Muhammad Shamie bertanggungjawab terhadap pengkonsepan kajian, pembangunan metodologi serta penyeliaan pelaksanaan kajian. Beliau juga mengendalikan proses pengumpulan data, analisis dan tafsiran dapatan kajian, di samping menyumbang kepada penyediaan sorotan literatur serta penulisan draf awal manuskrip. Hazrati Husnin menyumbang kepada semakan kritikal manuskrip serta penyeliaan keseluruhan kajian. Semua penulis telah membaca, menyemak dan meluluskan versi akhir manuskrip sebelum penyerahan.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1832-1847

Rujukan

Adnan, W. M. S. W. (2025). Keberkesanan kecerdasan buatan dalam pembelajaran ilmu nahu: Suatu tinjauan umum. ATTARBAWIY: Malaysian Online Journal of Education, 9(1), 1–10.

Ali, M. F., & Rani, M. J. A. (2024). Kecerdasan buatan (AI) sebagai medium transformasi pembelajaran bahasa Arab. Jurnal Al-Haady, 5(1), 1–9.

Aminun Nabil Ahmad Syahir. (2025). Kecerdasan buatan dan transformasi digital dalam pengajian al-Quran: Satu tinjauan literatur sistematik. QURANICA: International Journal of Quranic Research, S.I.14, 543–570.

Andriyana, Hidayat, A., Hanifah, I., & Ramdona, D. C. (2026). Integrasi NotebookLM dalam pembelajaran: Kajian literatur terhadap peranannya sebagai dukungan profesional guru. Buletin Pengembangan Perangkat Pembelajaran, 8(1), 47–60.

Azmi, N., & Ruzanna, W. M. (2026). Keberkesanan ChatGPT dalam pembelajaran tatabahasa Bahasa Melayu terhadap murid pintar dan berbakat. International Journal of Education, Psychology and Counselling, 11(62), 1343–1354.

Borham, S. R., Ramli, S., & Abdul Ghani, M. T. (2022). Integrasi konsep kecerdasan buatan dalam reka bentuk kit E-Muhadathah untuk bukan penutur Arab. Journal of ICT in Education, 9(3), 1–10.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Fitri, A. T. (2025). Tinjauan literatur: Integrasi kecerdasan buatan (AI) dalam pembelajaran bahasa Arab. Jurnal Ilmiah Pendidikan Scholastic, 9(1), 16–26.

Haryati Jaacob et al. (2026). Penggunaan kecerdasan buatan (AI) dalam pembentukan pola ayat menulis bahasa Arab. Educalia: Journal of Educational Research, 5(1), 26–48.

Kementerian Pendidikan Malaysia. (2023). Dasar Pendidikan Digital. Kementerian Pendidikan Malaysia.

Kemmis, S., & McTaggart, R. (1988). The action research planner. Deakin University Press.

Khazri Osman, Al Dairi Salha Juma Mohamed, Marhamah Ulfa, Dayang Sakini Mohd Kadri, & Mohamad Yazid Abdul Majid. (2025). Integrasi AI dalam komunikasi pengajaran bahasa Arab: Tinjauan kritis dari perspektif guru. Al-Qalam International Journal of Arabic Studies, 7, 80–92.

Manikam, N., Omar, R., Mahmood, R., Poongavanam, K., & Parmasivam, P. (2025). Integrasi kecerdasan buatan (AI) generatif dalam pembangunan profesional guru: Analisis keberkesanan alat berasaskan AI terhadap kecekapan dan penglibatan dalam pengajaran. Jurnal Pendidikan Bitara UPSI, 18(2), 175–189. https://doi.org/10.37134/bitara.vol18.2.15.2025

Rahem Samiya. (2025). Artificial intelligence in Arabic language education: Current applications, challenges, and future perspectives. Studies in Education Sciences, 6(4), 1–21.

Sharjeel Ahmed, Muhammad Faisal, Syed Shan Fareh Hasan, & Muhammad Ayaz Ghazi. (2025). Impact of artificial intelligence (AI) on teacher self-efficacy: A systematic literature review. Social Science Review Archives, 3(4), 3406–3420.

UNESCO. (2019). Beijing consensus on artificial intelligence and education. UNESCO.

UNESCO. (2024). AI competency framework for teachers. UNESCO.

Zainal Rafit, Z. H., Zainudin, S., & Ali Othman, Z. (2021). Model Peramalan Bilangan Calon Tarik Diri dari Peperiksaan Awam Malaysia Menerusi Pendekatan Perlombongan Data dan Petua.