PIJAR: Jurnal Pendidikan dan Pengajaran
Vol. 5, No 1 Desember 2026, pp. 60-75
E-ISSN: 2964-7460 , DOI: 10.58540/pijar.v5i1.2691

PENGARUH PEMANFAATAN MEDIA PUZZLE KATA TERHADAP HASIL BELAJAR PENYUSUNAN KALIMAT SEDERHANA BAHASA ARAB SISWA SEKOLAH DASAR: STUDI KUASI-EKSPERIMEN

Fitri Setyo Rini[1], Muhammad Wahyudi [2], Muthmainnah Choliq[3], Nadhilla Claryzha Sifana Dewi[4]

[1] Universitas Darussalam Gontor, Indonesia; fitrisetyo2@unida.gontor.ac.id
[2] Universitas Darussalam Gontor, Indonesia; muh.wahyudi@unida.gontor.ac.id
[3] Universitas Darussalam Gontor, Indonesia; Muthmainnahcholiq@gmail.com
[4] Universitas Darussalam Gontor, Indonesia; nadhilaclarissa@pba.unida.gontor.ac.id

ARTICLE INFO

Article history:

Received; 12 September 2026
Revised; 23 Oktober 2026
Accepted; 11 November 2026

ABSTRAK

Kesulitan menyusun kalimat sederhana masih menjadi persoalan pembelajaran bahasa Arab di sekolah dasar, terutama ketika media pembelajaran kurang bervariasi. Penelitian ini bertujuan menguji pengaruh media puzzle kata terhadap hasil belajar penyusunan kalimat sederhana bahasa Arab siswa kelas VI di Sekolah Islam Terpadu (SIT) Latansa Cendekia, Kabupaten Tangerang, tahun ajaran 2025/2026, dibandingkan dengan media kartu kata. Penelitian kuantitatif ini menggunakan desain kuasi-eksperimen *Non-Equivalent Control Group Design*. Dua kelas dipilih secara purposif: 27 siswa kelas eksperimen belajar dengan puzzle kata dan 28 siswa kelas kontrol belajar dengan kartu kata. Perlakuan diberikan dalam dua pertemuan pada September 2025 dengan materi kosakata dan kalimat verbal (*jumlah fi'liyyah*) bertema aktivitas di sekolah. Hasil belajar diukur dengan tes tertulis objektif 11 butir valid (skala 0–100; α = 0,666 pada butir uji coba) sebagai pretest dan posttest, didukung observasi, wawancara, dan dokumentasi. Karena sebagian besar data tidak berdistribusi normal, analisis memakai uji Wilcoxon dan uji Mann–Whitney dengan SPSS. Rata-rata nilai kelas eksperimen naik dari 59,67 menjadi 78,67 (median posttest 73), sedangkan kelas kontrol naik dari 57,25 menjadi 64,96 (median posttest 64). Uji Mann–Whitney menunjukkan perbedaan posttest yang signifikan antara kedua kelas (U = 184,000; Z = −3,352; p = 0,001) dengan ukuran efek sedang (r = 0,45). Secara praktis, selisih rata-rata 13,71 poin setara sekitar 1,5 butir soal, dan peluang siswa kelas puzzle kata bernilai lebih tinggi daripada siswa kelas kartu kata sekitar 76%. Hasil ini menunjukkan perbedaan hasil belajar yang signifikan pada sampel penelitian, bukan bukti efektivitas universal puzzle kata; durasi perlakuan yang singkat, desain tanpa randomisasi, dan hanya satu kelas per kondisi menuntut kehati-hatian dalam generalisasi.

Kata Kunci: Puzzle Kata; Media Pembelajaran; Hasil Belajar; Penyusunan Kalimat; Bahasa Arab; Kuasi-Eksperimen; Sekolah Dasar

ABSTRACT

Difficulty in composing simple sentences remains a persistent problem in Arabic learning at primary schools, particularly when instructional media are limited. This study examined the effect of word puzzle media on the simple-sentence-construction learning outcomes in Arabic of sixth-grade

students at SIT Latansa Cendekia, Tangerang Regency, in the 2025/2026 academic year, compared with word cards. A quantitative quasi-experimental study with a non-equivalent control group design was conducted. Two intact classes were selected through purposive sampling: 27 students in the experimental class learned with word puzzles, and 28 students in the control class learned with word cards. The treatment lasted two meetings in September 2025 and covered vocabulary and simple verbal sentences (jumlah fi'liyyah) on school activities. Learning outcomes were measured with an 11-item valid objective written test (scored 0–100; α = 0.666 on the trial items) as pretest and posttest, supported by observation, interviews, and documentation. Because most distributions were non-normal, the data were analyzed using the Wilcoxon signed-rank and Mann–Whitney U tests in SPSS. The mean score of the experimental class increased from 59.67 to 78.67 (posttest median 73), while that of the control class rose from 57.25 to 64.96 (posttest median 64). The Mann–Whitney test showed a significant posttest difference between the two classes (U = 184.000; Z = −3.352; p = 0.001) with a medium effect size (r = 0.45). In practical terms, the 13.71-point mean difference equals about 1.5 test items, and the probability that a word-puzzle student outscored a word-card student was about 76%. The results indicate a significant difference in learning outcomes within this sample, not evidence of universal effectiveness of word puzzles; the short treatment, the non-randomized design, and the use of one class per condition call for cautious generalization.

Keywords: *word puzzle; learning media; learning outcomes; sentence construction; Arabic language; quasi-experiment; primary school*

This is an open access article under the CC BY license.

Corresponding Author:
Fitri Setyo Rini
Universitas Darussalam Gontor, Ponorogo, Indonesia; fitrisetyo2@unida.gontor.ac.id

1. PENDAHULUAN

Pembelajaran bahasa Arab di sekolah dasar menuntut penguasaan dasar kebahasaan yang dibangun secara bertahap. Belajar pada hakikatnya adalah proses interaksi antara peserta didik, guru, dan sumber belajar yang diharapkan menghasilkan perubahan perilaku (Faizah & Kamal, 2024). Perubahan itu hanya terjadi bila peserta didik terlibat aktif, sehingga guru perlu merancang pengalaman belajar yang menarik dan bermakna. Dalam pembelajaran bahasa Arab, keterlibatan tersebut berkaitan dengan empat keterampilan berbahasa, yaitu menyimak, berbicara, membaca, dan menulis (Alhamdi & Afril, 2025). Pada tingkat pemula, kemampuan menulis ditandai antara lain oleh kemampuan menyusun kata menjadi kalimat yang utuh (Munawarah & Zulkiflih, 2021).

Namun, menyusun kalimat sederhana sering menjadi hambatan. Banyak siswa sekolah dasar masih kesulitan memahami struktur kalimat sederhana, memperkaya kosakata, dan

menggunakan kosakata Arab dalam kehidupan sehari-hari. Hal ini disebabkan oleh cara mengajar yang kurang menarik (Sanwil et al., 2021). Problematika pembelajaran bahasa Arab pada siswa madrasah ibtidaiyah juga dilaporkan oleh Wijaya dan Hikmah (2023), sementara Ghofur dan Riski (2024) menegaskan bahwa pembelajaran bahasa Arab menghadapi tantangan yang menuntut strategi yang efektif. Laporan empiris terbaru menunjukkan pola serupa. Qadhafi et al. (2026) melaporkan lemahnya penguasaan kosakata dan terbatasnya pemahaman sebagian kaidah tata bahasa pada siswa kelas VI sebuah sekolah dasar Islam di Jakarta Timur. Pada jenjang yang lebih tinggi, Fahimah dan Saleh (2026) memakai tes diagnostik 10 butir pada 25 siswa kelas VIII MTs dan menemukan bahwa butir penyusunan kalimat *jumlah fi'liyyah* paling sulit: butir 8 dan butir 10 hanya dijawab benar oleh 2 dari 25 siswa, dan butir 10 menguji kesesuaian jenis dan jumlah antara kata kerja dan pelakunya.

Kondisi tersebut ditemukan pula di Sekolah Islam Terpadu (SIT) Latansa Cendekia, Kabupaten Tangerang. Studi pendahuluan melalui observasi kelas (21 Juni 2025) dan wawancara dengan guru bahasa Arab kelas VI (22 Juni 2025) menunjukkan bahwa banyak siswa tampak bosan, berbicara sendiri, dan tidak dapat menjawab pertanyaan guru. Menurut guru, sebagian besar siswa sulit menyusun kalimat dari kosakata yang telah mereka miliki sehingga nilai ujian dan nilai harian mereka kurang stabil. Guru mengenal bentuk dan kegunaan puzzle kata, tetapi belum pernah menggunakannya dalam pembelajaran. Data kuantitatif awal memperkuat temuan tersebut: pada pretest sebelum perlakuan, 55 siswa dari dua kelas yang diteliti hanya memperoleh rata-rata 59,67 (kelas eksperimen) dan 57,25 (kelas kontrol) pada skala 0–100, dengan nilai tertinggi 91 dan tanpa satu pun nilai sempurna (Tabel 6). Persoalannya dengan demikian tidak semata-mata kekurangan kosakata, melainkan belum adanya media yang menjembatani kosakata dengan penyusunan kalimat secara menyenangkan.

Media pembelajaran berfungsi sebagai perantara yang membantu penyampaian pesan agar proses belajar lebih efektif dan menarik (Zahwa & Syafi'i, 2022). Salah satu media visual berbentuk permainan yang relevan adalah puzzle kata, yaitu potongan kata yang harus disusun peserta didik hingga membentuk kalimat bermakna (Listryanto & Widiyono, 2023; Taqiyah, 2023). Media ini melatih siswa memahami urutan kalimat secara visual dan logis (Rini & Sukartaningsih, 2021). Bukti yang lebih luas menunjukkan bahwa pembelajaran berbasis permainan berkontribusi positif terhadap capaian belajar bahasa. Meta-analisis Dixon et al. (2022) melaporkan efek kecil hingga sedang ($d = 0{,}50$) pada rancangan antarkelompok untuk pembelajaran bahasa kedua berbasis permainan digital. Lee dan Baek (2023) menemukan efek sedang ($g = 0{,}517$) gamifikasi terhadap kemahiran bahasa Inggris dan mencatat bahwa intervensi tanpa teknologi cenderung menghasilkan ukuran efek lebih besar. Kedua meta-analisis itu terutama menelaah bahasa Inggris, tetapi arah temuannya relevan bagi sekolah dengan sarana terbatas karena puzzle kata dapat dibuat secara manual.

Dalam pembelajaran bahasa Arab, media puzzle telah dikaji pada beberapa aspek. Dali et al. (2023) melaporkan pengaruh media puzzle terhadap hasil belajar bahasa Arab siswa kelas IV madrasah ibtidaiyah negeri; Jamilah dan Wijaya (2025) melaporkan peningkatan

penguasaan kosakata yang signifikan; dan Silvia dan Yusra (2026) melalui kuasi-eksperimen menyimpulkan bahwa puzzle meningkatkan penguasaan kosakata siswa kelas IV secara signifikan. Pada jenjang menengah, puzzle dilaporkan berpengaruh terhadap kemampuan berbicara (Hasibuan et al., 2025) serta keterampilan membaca dan menulis (Putri et al., 2025), sedangkan Maghfiroh (2025) mengkaji puzzle untuk keterampilan berbicara di madrasah ibtidaiyah. Adapun puzzle kata lebih banyak diteliti pada pelajaran bahasa Indonesia: untuk menyusun kalimat sederhana di kelas III SD (Putri, 2021), menyusun struktur kalimat S-P-O-K pada siswa dengan hambatan pendengaran (Aulia, 2016), membaca permulaan (Listryanto & Widiyono, 2023; Rini & Sukartaningsih, 2021), dan menggunakan kata baku (Mustikasari et al., 2023), dengan rancangan yang mencakup penelitian tindakan kelas, pengembangan produk, dan studi kasus kualitatif. Perbandingan ringkas studi-studi tersebut dengan penelitian ini disajikan pada Tabel 1.

Penelusuran tersebut, yang tidak dimaksudkan sebagai tinjauan sistematis, memperlihatkan dua celah. Pertama, kajian puzzle pada bahasa Arab terutama menyoroti kosakata, berbicara, dan membaca atau menulis, sedangkan pemanfaatan puzzle kata untuk penyusunan kalimat sederhana (*tarkib*) yang telah dikaji pada pelajaran bahasa Indonesia (Aulia, 2016; Putri, 2021) dan pengaruhnya terhadap hasil belajar bahasa Arab di sekolah dasar masih jarang diuji secara kuantitatif; Silvia dan Yusra (2026) sendiri menyatakan bahwa kajian efektivitas puzzle pada jenjang sekolah dasar Islam masih terbatas. Kedua, sebagian studi yang dapat ditelusuri membandingkan puzzle dengan pembelajaran konvensional, sehingga belum jelas apakah keunggulannya berasal dari format puzzle atau sekadar dari penggunaan media. Penelitian ini mengisi celah tersebut secara terbatas melalui dua kebaruan kontekstual: (a) berfokus pada penyusunan kalimat verbal sederhana bahasa Arab di kelas VI sekolah dasar; dan (b) membandingkan puzzle kata dengan pembanding aktif, yaitu kartu kata, bukan sekadar ceramah. Kebaruan ini bersifat kontekstual dan terbatas pada literatur yang dapat ditelusuri penulis.

Tabel 1. Ringkasan penelitian terdahulu tentang puzzle dalam pembelajaran bahasa

Penelitian	Bahasa dan jenjang	Aspek yang dikaji	Catatan pembeda
Dali et al. (2023)	Bahasa Arab, kelas IV MI	Hasil belajar bahasa Arab	Jenjang kelas IV; sasaran hasil belajar umum
Jamilah & Wijaya (2025)	Bahasa Arab, kelas IV MI	Penguasaan kosakata	Fokus kosakata, bukan penyusunan kalimat
Silvia & Yusra (2026)	Bahasa Arab, kelas IV SD Islam	Penguasaan kosakata (kuasi-eksperimen)	Fokus kosakata; menyatakan kajian di SD Islam masih terbatas
Maghfiroh (2025)	Bahasa Arab, MI	Keterampilan berbicara	Fokus berbicara
Hasibuan et al. (2025)	Bahasa Arab, SMP IT	Keterampilan berbicara	Jenjang menengah; kelompok kontrol konvensional

Penelitian	Bahasa dan jenjang	Aspek yang dikaji	Catatan pembeda
Putri et al. (2025)	Bahasa Arab, SMP Islam	Membaca dan menulis (puzzle suku kata)	Jenjang menengah; media berbeda
Putri (2021)	Bahasa Indonesia, kelas III SD	Menyusun kalimat sederhana (pengembangan media)	Bukan bahasa Arab
Aulia (2016)	Bahasa Indonesia, kelas VI SLB B/C	Struktur kalimat S-P-O-K	Bukan bahasa Arab; peserta dengan hambatan pendengaran
Listryanto & Widiyono (2023); Rini & Sukartaningsih (2021); Mustikasari et al. (2023)	Bahasa Indonesia, kelas I SD dan SMP	Membaca permulaan; penggunaan kata baku	Bukan bahasa Arab; bukan penyusunan kalimat
Penelitian ini	Bahasa Arab, kelas VI SD	Hasil belajar penyusunan kalimat verbal sederhana	Kuasi-eksperimen dengan pembanding aktif (kartu kata)

Tujuan penelitian ini ialah menguji pengaruh media puzzle kata terhadap hasil belajar penyusunan kalimat sederhana bahasa Arab siswa kelas VI SIT Latansa Cendekia tahun ajaran 2025/2026 dibandingkan dengan media kartu kata. Hipotesis nol (H_0) menyatakan bahwa puzzle kata tidak berpengaruh terhadap hasil belajar bahasa Arab (tidak ada perbedaan posttest antara kelas puzzle kata dan kelas kartu kata), sedangkan hipotesis alternatif (H_a) menyatakan bahwa puzzle kata berpengaruh terhadap hasil belajar bahasa Arab (ada perbedaan posttest).

2. METODE

Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen semu (*quasi-experiment*) dan desain *Non-Equivalent Control Group Design* (Anantasia & Rindrayani, 2025). Variabel bebas adalah penggunaan media puzzle kata, sedangkan variabel terikat adalah hasil belajar bahasa Arab. Kelas eksperimen belajar menyusun kalimat sederhana dengan puzzle kata, sedangkan kelas kontrol mempelajari materi yang sama dengan kartu kata. Kedua kelas diberi pretest sebelum dan posttest sesudah perlakuan (Tabel 2). Variabel terikat, yaitu hasil belajar bahasa Arab, didefinisikan secara operasional sebagai skor (0–100) pada tes tertulis objektif yang mengukur kosakata dan penyusunan kalimat verbal sederhana (*jumlah fi'liyyah*) bertema aktivitas di sekolah; indikatornya dirinci pada Tabel 5 sehingga selaras dengan perlakuan yang melatih siswa menyusun kosakata menjadi kalimat.

Tabel 2. Desain penelitian

Kelompok	n	Pretest	Perlakuan	Posttest
Eksperimen (VI Imam Bukhari)	27	O_1	Puzzle kata (X)	O_2
Kontrol (VI Imam Muslim)	28	O_3	Kartu kata	O_4

Populasi dan Sampel

Penelitian dilaksanakan di SIT Latansa Cendekia, Kabupaten Tangerang, pada tahun ajaran 2025/2026. Populasinya adalah 113 siswa kelas VI yang terbagi dalam empat kelas, yaitu VI Imam Muslim (28 siswa), VI Imam Tirmidzi (30), VI Imam Bukhari (27), dan VI Imam Syafi'i (28). Sampel ditentukan dengan teknik *purposive sampling* (Suriani et al., 2023) berdasarkan kemampuan awal bahasa Arab, karakteristik siswa, dan jumlah siswa yang relatif setara, dengan mempertimbangkan hasil observasi dan rekomendasi guru mata pelajaran. Kelas VI Imam Bukhari (27 siswa) ditetapkan sebagai kelas eksperimen dan kelas VI Imam Muslim (28 siswa) sebagai kelas kontrol, sehingga sampel berjumlah 55 siswa. Kedua kelas berasal dari tingkat kelas dan sekolah yang sama (Tabel 3). Penetapan kelas tidak dilakukan secara acak. Kesetaraan kemampuan awal dinilai dari pretest: rata-rata 59,67 (SB = 13,53) pada kelas eksperimen dan 57,25 (SB = 12,87) pada kelas kontrol, dengan rentang yang sama (37–91). Selisih 2,42 poin itu tergolong kecil (d = 0,18), dan uji-t Welch yang dihitung dari statistik ringkas tersebut tidak signifikan (t = 0,68; p = 0,50). Angka ini hanya indikasi kesetaraan, karena pretest kelas kontrol tidak berdistribusi normal (Tabel 8) dan kesetaraan tanpa randomisasi tidak dapat dijamin.

Tabel 3. Kelas dalam populasi dan karakteristik pretest kelas sampel

Kelas	n	Peran dalam penelitian	Pretest: rata-rata (SB)	Pretest: rentang
VI Imam Bukhari	27	Eksperimen (puzzle kata)	59,67 (13,53)	37–91
VI Imam Muslim	28	Kontrol (kartu kata)	57,25 (12,87)	37–91
VI Imam Tirmidzi	30	Tidak dilibatkan	–	–
VI Imam Syafi'i	28	Tidak dilibatkan	–	–
Total populasi	113	Sampel = 55 siswa		

Prosedur Perlakuan

Sebelum perlakuan, peneliti mewawancarai guru bahasa Arab (22 Juni 2025), mengurus izin penelitian, dan menyusun rencana pembelajaran. Pretest diberikan kepada kedua kelas sebelum perlakuan dan posttest setelah perlakuan berakhir, dengan butir tes yang sama. Perlakuan diberikan dalam dua pertemuan pada bulan September 2025 oleh peneliti di kedua kelas. Kedua kelas mempelajari materi yang sama, yaitu kosakata dan kalimat verbal sederhana (*jumlah fi'liyyah*) bertema aktivitas di sekolah dengan memperhatikan kesesuaian *mudzakkar–muannats*. Kerangka kegiatan (pendahuluan, penyajian, evaluasi, penerapan, dan penutup) serta alokasi waktu disamakan; perbedaan hanya terletak pada kegiatan inti (Tabel 4). Pada kelas eksperimen, kegiatan inti berupa permainan lima putaran dengan batas waktu maksimal 5 menit per putaran; setiap putaran menuntut setiap siswa menyusun dua kalimat sesuai gambar (10 kalimat per permainan). Pada kelas kontrol, siswa menggabungkan kartu kata menjadi satu kalimat bersama temannya. Perbedaan jumlah latihan ini merupakan bagian dari perlakuan dan dibahas sebagai keterbatasan.

Tabel 4. Perbandingan kegiatan inti pada kedua kelas

Aspek	Kelas eksperimen (puzzle kata)	Kelas kontrol (kartu kata)
Pengelompokan	Kelompok kecil beranggotakan lima siswa	Siswa bekerja bersama teman; setiap siswa menerima satu potongan kartu kata
Media	Satu map per kelompok berisi 10 gambar, 10 papan puzzle, dan satu set potongan kata untuk membentuk kalimat verbal (*jumlah fi'liyyah*)	Potongan kartu kata
Aktivitas inti	Lima putaran; setiap siswa menyusun dua kalimat per putaran sesuai gambar (maksimal 5 menit per putaran); gambar yang sudah dipilih tidak boleh dipilih lagi	Siswa mencari teman yang memegang kartu lain sehingga kartu-kartu itu membentuk kalimat bermakna
Umpan balik dan penilaian	Putaran 1–2 diperiksa dan diperbaiki guru; putaran 3–5 dinilai (skor 1 untuk jawaban benar); kelompok dengan skor tertinggi menang	Guru memeriksa kalimat setelah tersusun
Penguatan	Guru membahas kalimat yang benar dan salah, menjelaskan ulang materi, dan memberi hadiah	Guru membahas kalimat yang benar dan salah, menjelaskan ulang materi, dan memberi hadiah

Instrumen dan Pengumpulan Data

Hasil belajar diukur dengan tes tertulis objektif yang disusun berdasarkan kompetensi dasar (KD) tema aktivitas di sekolah. Kumpulan butir uji coba mencakup tiga ranah kognitif, yaitu pengetahuan (C1), pemahaman (C2), dan penerapan (C3), dalam enam bentuk soal: menjodohkan, melengkapi kosakata, menyusun kalimat, mencocokkan gambar, pilihan ganda berdasarkan gambar, dan benar–salah. Instrumen diuji coba pada 30 siswa di luar sampel dan divalidasi oleh guru bahasa Arab kelas VI. Validitas butir diuji dengan korelasi *Pearson product moment* (r tabel = 0,361 untuk n = 30; Anggraini et al., 2022): dari 20 butir uji coba yang dianalisis, 11 butir valid (r hitung 0,374–0,579) dan digunakan sebagai pretest dan posttest, sedangkan 9 butir lainnya digugurkan. Butir yang gugur mencakup seluruh butir menjodohkan dan melengkapi kosakata, sehingga tes final terdiri atas empat bentuk soal dan 11 butir: 3 butir menyusun kalimat (C3), 3 butir mencocokkan gambar (C1), 2 butir pilihan ganda berdasarkan gambar (C2), dan 3 butir benar–salah (C3). Dengan demikian 6 dari 11 butir (54,5%) berada pada ranah penerapan, yaitu menyusun dan menilai kalimat *jumlah fi'liyyah* sesuai *mudzakkar–muannats*, sehingga isi tes selaras dengan perlakuan; indikator tiap bentuk soal dirinci pada Tabel 5. Skor dinyatakan pada skala 0–100; karena butirnya hanya 11, selisih satu butir sekitar 9 poin (nilai yang teramati: 37, 46, 55, 64, 73, 82, 91, dan 100). Reliabilitas diuji dengan *Cronbach's alpha* pada butir uji coba dan diperoleh 0,666, yang melampaui batas minimum 0,60; reliabilitas 11 butir final tidak dihitung terpisah. Data penunjang dikumpulkan melalui observasi kelas, wawancara semi terstruktur dengan guru bahasa Arab, dan dokumentasi (foto, video, serta profil sekolah).

Tabel 5. Kisi-kisi tes hasil belajar final (11 butir valid)

Bentuk soal	KD	Ranah	Indikator	Butir final
Menyusun kalimat	4.2	C3 Penerapan	Siswa menyusun kosakata yang diberikan menjadi kalimat sederhana bertema aktivitas di sekolah, sesuai mudzakkar–muannats	3
Mencocokkan gambar	4.4	C1 Pengetahuan	Siswa mengingat kosakata Arab dan menjawab pertanyaan sederhana tentang aktivitas di sekolah dengan benar, sesuai mudzakkar–muannats	3
Pilihan ganda berdasarkan gambar	3.1	C2 Pemahaman	Siswa memahami dan mencocokkan kalimat sederhana yang sesuai dengan gambar aktivitas di sekolah, sesuai mudzakkar–muannats	2
Benar–salah	4.1	C3 Penerapan	Siswa memilih struktur kalimat sederhana yang benar dan bermakna dalam tema aktivitas di sekolah, sesuai mudzakkar–muannats	3
Menjodohkan; melengkapi kosakata	3.3; 3.2	C1; C2	Seluruh butir tidak valid pada uji coba sehingga tidak dipakai	0
Total butir final				11

Analisis Data

Ancaman terhadap validitas internal berupa *history* dan *maturation* dikendalikan dengan melaksanakan eksperimen dalam waktu singkat dan menyamakan waktu belajar kedua kelas; kemampuan awal diketahui melalui pretest (Siska et al., 2024). Ancaman seleksi (*selection*) tidak dapat dikendalikan karena kelas tidak ditetapkan secara acak. Media puzzle belum pernah digunakan di sekolah tersebut sebelum penelitian.

Data dianalisis dengan SPSS. Statistik deskriptif (n, minimum, maksimum, rata-rata, dan simpangan baku) disajikan untuk setiap kelompok, dilengkapi median, kuartil, rentang, dan distribusi frekuensi untuk posttest. Normalitas diuji dengan Kolmogorov–Smirnov dan Shapiro–Wilk (Isnaini et al., 2025), dan homogenitas varians posttest dengan uji Levene (Usmadi, 2020). Karena asumsi normalitas tidak terpenuhi pada sebagian besar data, perubahan nilai dalam kelas diuji dengan *Wilcoxon signed-rank* (Astuti et al., 2021) dan perbedaan antarkelas pada posttest diuji dengan *Mann–Whitney U* (Qolby, 2020) pada taraf signifikansi 0,05 (dua sisi); arah perbedaan dibaca dari peringkat rata-rata dan median. Ukuran efek dihitung sebagai $r = |Z|/\sqrt{N}$ (N = jumlah pengamatan; pada uji Wilcoxon $N = 2n$) dan ditafsirkan dengan batas 0,10, 0,30, dan 0,50 untuk efek kecil, sedang, dan besar (Fritz et al., 2012). Makna praktis dijelaskan dengan tiga indikator: (a) peluang keunggulan (*probability of superiority*, $PS = U/(n_1 \cdot n_2)$), yaitu peluang seorang siswa kelas eksperimen bernilai lebih tinggi daripada seorang siswa kelas kontrol dengan nilai yang sama dihitung setengah

(McGraw & Wong, 1992), beserta interval kepercayaan bootstrap persentil 95% (5.000 resampel); (b) selisih rata-rata dalam satuan butir soal; dan (c) gain ternormalisasi kelompok, N-gain = (rata-rata posttest − rata-rata pretest)/(100 − rata-rata pretest), dengan kategori rendah (< 0,3), sedang (0,3–0,7), dan tinggi (> 0,7) (Hake, 1998). Ketiganya dihitung dari distribusi frekuensi posttest (Tabel 7) dan statistik ringkas. H_0 ditolak apabila $p < 0,05$.

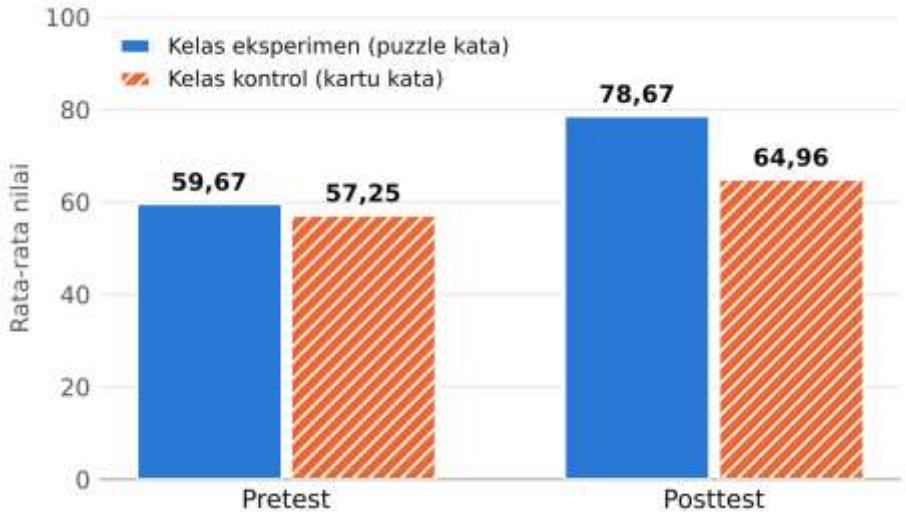
3. HASIL DAN PEMBAHASAN

Hasil Penelitian

Tabel 6 meringkas nilai pretest dan posttest kedua kelas. Kemampuan awal kedua kelas relatif sebanding: rata-rata pretest kelas eksperimen 59,67 dan kelas kontrol 57,25, dengan nilai terendah dan tertinggi yang sama (37 dan 91). Setelah perlakuan, rata-rata kelas eksperimen naik 19,00 poin menjadi 78,67, sedangkan kelas kontrol naik 7,71 poin menjadi 64,96 (Gambar 1). Selisih rata-rata posttest kedua kelas adalah 13,71 poin. Rentang posttest kelas eksperimen 55–100 dan kelas kontrol 46–100.

Tabel 6. Statistik deskriptif nilai pretest dan posttest

Kelompok data	n	Minimum	Maksimum	Rata-rata	Simpangan baku
Pretest eksperimen	27	37	91	59,67	13,528
Posttest eksperimen	27	55	100	78,67	14,587
Pretest kontrol	28	37	91	57,25	12,874
Posttest kontrol	28	46	100	64,96	13,713



Gambar 1. Rata-rata nilai pretest dan posttest kelas eksperimen dan kelas kontrol

Sebaran nilai posttest pada Tabel 7 memperlihatkan pergeseran ke kategori yang lebih tinggi pada kelas eksperimen. Nilai modus kedua kelas sama, yaitu 64 (8 siswa di kelas eksperimen dan 11 siswa di kelas kontrol), tetapi 13 dari 27 siswa kelas eksperimen (48,1%) memperoleh nilai 82 atau lebih, dibandingkan 4 dari 28 siswa kelas kontrol (14,3%). Enam siswa kelas eksperimen (22,2%) memperoleh nilai maksimum 100, sedangkan di kelas kontrol hanya satu siswa (3,6%). Sebaliknya, tidak ada siswa kelas eksperimen yang berada pada nilai

46, sementara 4 siswa kelas kontrol (14,3%) berada pada nilai tersebut. Median posttest kelas eksperimen adalah 73 (Q1–Q3 = 64–91) dan kelas kontrol 64 (Q1–Q3 = 55–73). Satu siswa eksperimen (3,7%) bernilai 55 atau kurang, dibandingkan 9 siswa kontrol (32,1%).

Tabel 7. Distribusi frekuensi nilai posttest

Nilai	Eksperimen (n = 27) f (%)	Kontrol (n = 28) f (%)
46	0 (0,0)	4 (14,3)
55	1 (3,7)	5 (17,9)
64	8 (29,6)	11 (39,3)
73	5 (18,5)	4 (14,3)
82	5 (18,5)	1 (3,6)
91	2 (7,4)	2 (7,1)
100	6 (22,2)	1 (3,6)
Median	73	64
Q1–Q3	64–91	55–73
Rentang	55–100	46–100

Uji Shapiro–Wilk (Tabel 8) menunjukkan bahwa hanya pretest kelas eksperimen yang berdistribusi normal (p = 0,220). Pretest kelas kontrol serta posttest kedua kelas menyimpang dari normalitas (p antara 0,005 dan 0,011), dan hasil Kolmogorov–Smirnov searah. Karena itu analisis inferensial memakai uji nonparametrik. Uji Levene terhadap nilai posttest kedua kelas menunjukkan varians yang homogen (statistik Levene = 1,550; db1 = 1; db2 = 53; p = 0,219; berdasarkan median: 1,332; p = 0,254).

Tabel 8. Hasil uji normalitas

Kelompok data	n	Kolmogorov–Smirnov statistik (p)	Shapiro–Wilk statistik	p
Pretest eksperimen	27	0,153 (0,103)	0,950	0,220
Posttest eksperimen	27	0,176 (0,031)	0,879	0,005
Pretest kontrol	28	0,284 (< 0,001)	0,897	0,010
Posttest kontrol	28	0,242 (< 0,001)	0,899	0,011

Uji Wilcoxon (Tabel 9) menunjukkan bahwa nilai posttest kedua kelas lebih tinggi daripada pretest. Di kelas eksperimen seluruh 27 siswa mengalami kenaikan nilai (Z = −4,598; p < 0,001), sedangkan di kelas kontrol 25 siswa naik, 1 siswa turun, dan 2 siswa tetap (Z = −4,707; p < 0,001). Jumlah peringkat positif kelas eksperimen 378,00 (rata-rata peringkat 14,00, tanpa peringkat negatif), sedangkan pada kelas kontrol 337,50 (rata-rata 13,50) berbanding 13,50 untuk peringkat negatif; ukuran efek r = 0,63 pada kedua kelas (besar). Dengan demikian kedua perlakuan diikuti peningkatan hasil belajar yang signifikan.

Tabel 9. Hasil uji Wilcoxon (nilai posttest − pretest) pada setiap kelas

Kelas	n	Naik / turun / tetap	Jumlah peringkat (+ / –)	Z	p	r
Eksperimen	27	27 / 0 / 0	378,00 / 0,00	−4,598	< 0,001	0,63
Kontrol	28	25 / 1 / 2	337,50 / 13,50	−4,707	< 0,001	0,63

Untuk menjawab apakah hasil belajar setelah perlakuan berbeda antarkelas, nilai posttest kedua kelas dibandingkan dengan uji Mann–Whitney (Tabel 10). Peringkat rata-rata kelas eksperimen (35,19) lebih tinggi daripada kelas kontrol (21,07), dengan U = 184,000, Z = −3,352, dan p = 0,001. Karena p < 0,05, H_0 ditolak dan H_a diterima: terdapat perbedaan hasil belajar bahasa Arab yang signifikan antara siswa yang belajar dengan puzzle kata dan siswa yang belajar dengan kartu kata pada sampel penelitian ini. Ukuran efeknya r = 3,352/√55 = 0,45, tergolong sedang (mendekati batas efek besar, 0,50).

Tabel 10. Hasil uji Mann–Whitney pada nilai posttest dan ukuran efek

Statistik	Hasil
Peringkat rata-rata (jumlah peringkat)	Eksperimen 35,19 (950,00); kontrol 21,07 (590,00)
Mann–Whitney U; Wilcoxon W	184,000; 590,000
Z; p (dua sisi)	−3,352; 0,001
Ukuran efek r = |Z|/√N	0,45 (sedang)
Peluang keunggulan (PS)	0,76 (IK 95% bootstrap 0,63–0,87)
Median posttest (Q1–Q3)	Eksperimen 73 (64–91); kontrol 64 (55–73)

Makna praktis ukuran efek dapat dibaca dari tiga hal. Pertama, selisih rata-rata posttest 13,71 poin setara sekitar 1,5 butir dari 11 butir tes (satu butir ≈ 9,09 poin), dan selisih median 9 poin setara satu butir. Kedua, peluang seorang siswa kelas puzzle kata bernilai lebih tinggi daripada seorang siswa kelas kartu kata adalah 0,76 (IK 95% bootstrap 0,63–0,87; nilai yang sama dihitung setengah). Ketiga, N-gain kelas eksperimen 0,47 (kategori sedang) berbanding 0,18 pada kelas kontrol (kategori rendah) menurut kategori Hake (1998). Perubahan dan sebaran nilai kedua kelas dirangkum pada Tabel 11. Karena tes hanya 11 butir dan 6 siswa eksperimen (22,2%) mencapai nilai maksimum, skor kurang peka membedakan siswa berkemampuan tinggi (kemungkinan efek langit-langit).

Tabel 11. Perubahan dan sebaran nilai kedua kelas

Indikator	Eksperimen (n = 27)	Kontrol (n = 28)
Selisih rata-rata (posttest − pretest)	19,00 poin	7,71 poin
N-gain kelompok (kategori)	0,47 (sedang)	0,18 (rendah)
Siswa naik / tetap / turun	27 / 0 / 0	25 / 2 / 1
Median posttest (Q1–Q3)	73 (64–91)	64 (55–73)

Indikator	Eksperimen (n = 27)	Kontrol (n = 28)
Siswa bernilai ≥ 82 pada posttest	13 (48,1%)	4 (14,3%)
Siswa bernilai ≤ 55 pada posttest	1 (3,7%)	9 (32,1%)
Siswa bernilai 100 pada posttest	6 (22,2%)	1 (3,6%)

Pembahasan

Hasil penelitian menjawab tujuan penelitian: pada sampel ini, siswa yang belajar dengan puzzle kata memperoleh hasil belajar penyusunan kalimat yang lebih tinggi secara signifikan daripada siswa yang belajar dengan kartu kata. Temuan ini menunjukkan perbedaan antara dua kelas yang diteliti, bukan bukti bahwa puzzle kata efektif secara universal. Dua temuan perlu dibaca bersama. Pertama, kedua kelas sama-sama meningkat secara signifikan (Tabel 9). Hal ini menunjukkan bahwa kegiatan menyusun kalimat dengan media konkret, apa pun bentuknya, membantu siswa, di samping kemungkinan pengaruh keakraban dengan soal yang sama pada pretest. Kedua, kenaikan kelas eksperimen (19,00 poin) jauh lebih tinggi dari kenaikan kelas kontrol (7,71 poin), dan perbedaan posttest antarkelas signifikan dengan ukuran efek sedang (r = 0,45). Karena pembanding dalam penelitian ini juga memakai media, perbedaan itu tidak dapat dijelaskan semata-mata oleh "adanya media".

Beberapa karakteristik perlakuan dapat menjelaskan perbedaan tersebut, tetapi perlu ditegaskan bahwa penelitian ini tidak memisahkan efek tiap komponen. Pertama, puzzle kata dilengkapi gambar sebagai petunjuk makna sehingga siswa mengaitkan kosakata, urutan kata, dan makna kalimat secara visual dan logis (Rini & Sukartaningsih, 2021). Kedua, struktur lima putaran, dengan pemeriksaan guru pada dua putaran awal dan penilaian pada tiga putaran akhir, memberi umpan balik segera dan latihan berulang: setiap siswa menyusun dua kalimat per putaran, sedangkan pada kartu kata siswa bersama temannya menghasilkan satu kalimat dari gabungan kartu. Ketiga, kerja kelompok dan persaingan sehat antartim kemungkinan meningkatkan keterlibatan, sebagaimana dilaporkan pada studi puzzle lain yang mencatat antusiasme dan partisipasi aktif siswa (Jamilah & Wijaya, 2025; Putri et al., 2025); keterlibatan ini tidak diukur langsung dalam penelitian ini. Perbedaan hasil belajar karena itu paling tepat dipahami sebagai efek gabungan dari format permainan, petunjuk gambar, dan jumlah latihan yang lebih banyak.

Arah temuan ini sejalan dengan penelitian puzzle pada pembelajaran bahasa Arab. Dali et al. (2023) dan Jamilah dan Wijaya (2025) melaporkan pengaruh positif puzzle pada hasil belajar dan kosakata siswa kelas IV madrasah ibtidaiyah, Silvia dan Yusra (2026) menyimpulkan peningkatan signifikan pada kuasi-eksperimen berdesain serupa, dan Hasibuan et al. (2025) serta Putri et al. (2025) menemukan perbedaan signifikan pada keterampilan berbicara serta membaca dan menulis di jenjang menengah. Pada pelajaran bahasa Indonesia, puzzle kata juga dilaporkan meningkatkan kemampuan siswa (Mustikasari et al., 2023). Secara umum arah ini konsisten dengan meta-analisis pembelajaran bahasa berbasis permainan yang melaporkan efek positif sedang (Dixon et al., 2022; Lee & Baek,

2023), meskipun metrik ukuran efeknya berbeda sehingga perbandingan angka hanya bersifat indikatif.

Terdapat satu perbedaan yang layak dicatat. Pada penelitian Hasibuan et al. (2025), kelompok kontrol yang memperoleh pembelajaran konvensional tidak meningkat signifikan pada uji berpasangan, sedangkan kelas kontrol dalam penelitian ini meningkat signifikan. Perbedaan itu masuk akal karena pembanding di sini adalah kartu kata, yaitu media aktif, bukan pembelajaran konvensional murni. Nilai tambah temuan ini terletak pada bukti kuantitatif awal untuk penyusunan kalimat sederhana bahasa Arab di sekolah dasar dengan pembanding aktif, sehingga selisih yang diperoleh dapat dibaca sebagai indikasi tambahan manfaat format puzzle kata pada sampel ini, dengan catatan bahwa jumlah latihan kedua kelas tidak setara.

Wawancara dengan guru bahasa Arab menunjukkan sejumlah faktor yang dapat membatasi optimalisasi media ini: waktu pelajaran yang terbatas, sulitnya mengelola aktivitas kelompok di kelas, perbedaan motivasi dan dukungan keluarga antarsiswa, serta lemahnya dasar bahasa Arab pada sebagian siswa sehingga sulit beralih dari menata kata ke memahami struktur kalimat. Silvia dan Yusra (2026) juga menyoroti ketergantungan pada keterampilan guru mengelola kelas dan kemungkinan efek kebaruan. Hasil penelitian ini perlu ditafsirkan dengan mempertimbangkan keterbatasan berikut: (1) kelas tidak dipilih secara acak, dan kesetaraan awal hanya dinilai dari statistik deskriptif serta uji-t Welch indikatif yang dihitung dari statistik ringkas, tanpa uji formal pada data mentah atau penyesuaian kovariat; (2) hanya ada satu kelas per kondisi di satu sekolah, sehingga efek perlakuan tidak dapat dipisahkan dari efek kelas, dan siswa dianalisis sebagai unit independen meskipun sebagian kegiatan berlangsung berkelompok; (3) perlakuan hanya dua pertemuan sehingga ketahanan efek dan efek kebaruan belum diketahui; (4) peneliti terlibat langsung dalam pelaksanaan pembelajaran di kedua kelas sehingga bias peneliti tidak dapat dikesampingkan; (5) perlakuan bersifat paket (kelompok, gambar, putaran, dan skor) dengan dosis latihan yang berbeda sehingga kontribusi tiap komponen tidak dapat dipisahkan; (6) tes hanya 11 butir (selisih satu butir sekitar 9 poin), reliabilitas ($\alpha = 0{,}666$) diperoleh pada butir uji coba dan tergolong sedang, serta kemungkinan efek langit-langit karena 6 siswa kelas eksperimen mencapai nilai maksimum; dan (7) capaian belum dianalisis per ranah kognitif (C1–C3), sehingga belum dapat dipastikan apakah perbedaan lebih besar pada keterampilan menyusun kalimat daripada pemahaman konsep.

Bagi guru, temuan ini menunjukkan bahwa puzzle kata buatan sendiri dari bahan sederhana dapat menjadi media yang murah untuk melatih penyusunan kalimat; beban persiapan manual dapat dikurangi dengan membuat templat yang dipakai berulang. Agar pengaruhnya lebih kuat, puzzle sebaiknya diikuti penjelasan eksplisit kaidah kebahasaan, misalnya kesesuaian *mudzakkar–muannats*, serta pengelolaan waktu yang ketat. Bagi sekolah, pola kegiatan lima putaran dengan umpan balik guru dapat diadaptasi pada materi lain.

4. KESIMPULAN

Pada sampel penelitian ini, hasil belajar penyusunan kalimat sederhana bahasa Arab siswa kelas VI SIT Latansa Cendekia yang belajar dengan puzzle kata lebih tinggi secara signifikan daripada siswa yang belajar dengan kartu kata. Rata-rata nilai kelas eksperimen naik 19,00 poin (dari 59,67 menjadi 78,67), lebih besar daripada kenaikan kelas kontrol yang memakai kartu kata, yaitu 7,71 poin (dari 57,25 menjadi 64,96). Perbedaan posttest antarkelas signifikan (Mann–Whitney U = 184,000; p = 0,001; r = 0,45; ukuran efek sedang), sehingga H_0 ditolak dan H_a diterima. Secara praktis, selisih rata-rata posttest 13,71 poin setara sekitar 1,5 butir soal, dengan peluang sekitar 76% bahwa seorang siswa kelas puzzle kata bernilai lebih tinggi daripada seorang siswa kelas kartu kata. Temuan ini berlaku untuk dua kelas yang diteliti dengan perlakuan dua pertemuan dan bukan bukti efektivitas universal puzzle kata, mengingat kelas tidak dipilih secara acak, hanya ada satu kelas per kondisi, dan tes terdiri atas 11 butir. Kontribusi penelitian ini ialah bukti kuantitatif awal bahwa, pada penyusunan kalimat sederhana bahasa Arab di kelas VI sekolah dasar, puzzle kata berkaitan dengan hasil belajar yang lebih tinggi daripada media aktif lain (kartu kata).

Penelitian lanjutan disarankan: (1) memperpanjang perlakuan melebihi dua pertemuan dan menambahkan posttest tertunda untuk menguji ketahanan efek; (2) memakai penugasan acak pada beberapa kelas per kondisi atau analisis yang mengendalikan nilai awal, misalnya ANCOVA atau skor peningkatan; (3) merancang eksperimen yang memisahkan komponen puzzle (gambar, kerja kelompok, dan jumlah latihan); (4) menganalisis capaian per ranah kognitif, memperbanyak butir tes dan melaporkan reliabilitas tes final, serta menambahkan data kualitatif tentang motivasi dan keterlibatan; dan (5) mereplikasi penelitian di sekolah dan jenjang lain.

5. REFERENSI

Alhamdi, F., & Afril, R. (2025). Maharah lughawiyah dalam komponen pembelajaran bahasa Arab. *Al-Muyassar: Journal of Arabic Education*, *4*(1), 214–225.

Anantasia, G., & Rindrayani, S. R. (2025). Metodologi penelitian quasi eksperimen. *ADIBA: Journal of Education*, *5*(2), 183–192.

Anggraini, F. D. P., Aprianti, A., Setyawati, V. A. V., & Hartanto, A. A. (2022). Pembelajaran statistika menggunakan software SPSS untuk uji validitas dan reliabilitas. *Jurnal Basicedu*, *6*(4), 6491–6504.

Astuti, W., Taufiq, M., & Muhammad, T. (2021). Implementasi Wilcoxon signed rank test untuk mengukur efektifitas pemberian video tutorial dan PPT untuk mengukur nilai teori. *Jurnal Produktif*, *5*(1), 405–410.

Aulia, M. R. (2016). *Penggunaan media puzzle kata dalam meningkatkan kemampuan penyusunan struktur kalimat (SPOK) pada peserta didik hambatan pendengaran di kelas VI SLB B/C Nugraha Jakarta Utara* [Skripsi, Universitas Negeri Jakarta].

Dali, Z., Damhuri, D., & Iman, M. N. (2023). The effect of puzzle media on Arabic learning outcomes of grade IV students of Madrasah Ibtidaiyah 2 Limboto. *Al-Kalim: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, *2*(1), 102–118.

Dixon, D. H., Dixon, T., & Jordan, E. (2022). Second language (L2) gains through digital game-based language learning (DGBLL): A meta-analysis. *Language Learning & Technology*, *26*(1), 1–25. https://hdl.handle.net/10125/73464

Fahimah, N. H., & Saleh, N. (2026). Analisis learning obstacle materi jumlah fi'liyyah dalam pembelajaran bahasa Arab pada siswa kelas VIII MTs Ar Rohmah. *ARMADA: Jurnal Penelitian Multidisiplin, 4*(7), 3165–3177. https://doi.org/10.55681/armada.v4i7.3182

Faizah, H., & Kamal, R. (2024). Belajar dan pembelajaran. *Jurnal Basicedu*, *8*(1), 466–476.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18. https://doi.org/10.1037/a0024338

Ghofur, A., & Riski, R. B. (2024). Pendidikan bahasa Arab di era digital: Tantangan, peluang dan strategi menuju pembelajaran yang efektif. *El-Fusha: Jurnal Bahasa Arab dan Pendidikan*, *5*(1), 15–28.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74. https://doi.org/10.1119/1.18809

Hasibuan, H. S., Muslim, B., Ridha, M., & Ninoersy, T. (2025). تطبيق وسيلة Puzzle في تدريس مهارة الكلام بـ SMP IT Al Farabi Bilingual School Aceh Besar [Penerapan media puzzle dalam pembelajaran keterampilan berbicara di SMP IT Al Farabi Bilingual School Aceh Besar]. *Lisanuna: Jurnal Ilmu Bahasa Arab dan Pembelajarannya*, *15*(2), 175–192.

Isnaini, M., Afgani, M. W., Haqqi, A., & Azhari, I. (2025). Teknik analisis data uji normalitas. *J-CEKI: Jurnal Cendekia Ilmiah*, *4*(2), 1377–1384.

Jamilah, N. A., & Wijaya, M. (2025). Media puzzle as an interactive strategy to enhance Arabic vocabulary mastery among students: A study in class IV of MIN 1 Probolinggo. *El-Ibtikar: Jurnal Pendidikan Bahasa Arab*, *14*(1), 132–142.

Lee, J.-Y., & Baek, M. (2023). Effects of gamification on students' English language proficiency: A meta-analysis on research in South Korea. *Sustainability*, *15*(14), 11325. https://doi.org/10.3390/su151411325

Listryanto, D. P., & Widiyono, A. (2023). Implementasi media puzzle kata pada kurikulum merdeka untuk mengatasi kesulitan membaca siswa kelas I di sekolah dasar. *BASICA*, *3*(1), 65–80.

Maghfiroh, L. (2025). Media pembelajaran puzzle dalam meningkatkan kemampuan berbicara bahasa Arab di madrasah ibtidaiyah (MI). *At-Taklim: Jurnal Pendidikan Multidisiplin*, *2*(11), 197–213.

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, 111*(2), 361–365. https://doi.org/10.1037/0033-2909.111.2.361

Munawarah, M., & Zulkiflih, Z. (2021). Pembelajaran keterampilan menulis (maharah al-kitabah) dalam bahasa Arab. *Loghat Arabi: Jurnal Bahasa Arab dan Pendidikan Bahasa Arab*, *1*(2), 22–34.

Mustikasari, M. N. E., Hadi, P. K., & Hurustianit, A. (2023). Meningkatkan kemampuan penggunaan kata baku bahasa Indonesia dengan media pembelajaran "puzzle kata" pada siswa kelas VIII SMP Negeri 1 Mlarak tahun pelajaran 2022/2023. *Shambhasana: Seminar Nasional Bahasa, Sastra, dan Pengajarannya*, *2*, 159–166.

Putri, L. A., Baihaqi, M., Huda, H., & Rofiq, A. (2025). The effect of applying syllable puzzles on improving Arabic reading and writing skills among students at an Islamic junior high school. *Arabia: Jurnal Pendidikan Bahasa Arab*, *17*(2), 193–212. https://doi.org/10.21043/arabia.v17i2.34259

Putri, R. S. A. (2021). *Pengembangan media puzzle kata untuk meningkatkan kemampuan menyusun kalimat sederhana siswa kelas III SDIT Al-Uswah 02 Banyuwangi* [Skripsi, Universitas Islam Negeri Maulana Malik Ibrahim].

Qadhafi, M., Sundari, N., Hidayat, T., Wijaya, T., & Usman, S. M. (2026). Analysis of the difficulties in teaching Arabic language to students: Field study at "Sarj Al-Huda" Islamic primary school, East Jakarta. *Al-Muyassar: Journal of Arabic Education, 5*(2), 560–575. https://doi.org/10.31000/hfs4tb43

Qolby, B. S. (2020). Uji Mann Whitney dalam statistika non parametrik. *Jurnal Mahasiswa Teknik Informatika*, *2*(1), 3–16.

Rini, W. A., & Sukartaningsih, W. (2021). Pengembangan media puzzle kata bergambar menggunakan aplikasi Instagram untuk siswa kelas 1 sekolah dasar pada keterampilan membaca permulaan. *Jurnal Penelitian Pendidikan Guru Sekolah Dasar*, *9*(1), 2597–2609.

Sanwil, T., Utami, R., Hidayat, R., dkk. (2021). *Pembelajaran bahasa Arab untuk siswa SD/MI*. Yayasan Penerbit Muhammad Zaini.

Silvia, M., & Yusra, O. (2026). The effect of using puzzle media on Arabic vocabulary mastery among fourth-grade students at Kuttab An-Nahl Islamic Elementary School, Danjung-Danjung. *al-Ittijah: Jurnal Keilmuan dan Kependidikan Bahasa Arab*, *18*(1), 23–37. https://doi.org/10.15408/alittijah.v18i1.12918

Siska, M., Indah, M. R., & Wardani, R. (2024). Systematic literature review: Validitas internal dan eksternal dalam desain eksperimen. *EDUINOVASI: Journal of Basic Educational Studies*, *4*(3), 987–1001.

Suriani, N., Risnita, R., & Jailani, M. S. (2023). Konsep populasi dan sampling serta pemilihan partisipan ditinjau dari penelitian ilmiah pendidikan. *IHSAN: Jurnal Pendidikan Islam*, *1*(2), 24–36.

Taqiyah, S. Z. (2023). Implementasi manual media Kaana puzzle dalam pembelajaran tata bahasa Arab. *Tatsqifiy: Jurnal Pendidikan Bahasa Arab*, *4*(1), 34–51.

Usmadi, U. (2020). Pengujian persyaratan analisis (uji homogenitas dan uji normalitas). *Inovasi Pendidikan*, *7*(1), 51–53.

Wijaya, M., & Hikmah, F. (2023). Problematika pembelajaran maharah qiro'ah dalam pembelajaran bahasa Arab pada siswa madrasah ibtidaiyah. *Jurnal Educatio FKIP UNMA*, *9*(2), 858–864.

Zahwa, F. A., & Syafi'i, I. (2022). Pemilihan pengembangan media pembelajaran berbasis teknologi informasi. *Equilibrium: Jurnal Penelitian Pendidikan dan Ekonomi*, *19*(1), 61–78.