



Developing and Evaluating an Arabic Speaking Curriculum in an Indonesian Islamic Boarding School: A 4-D Model-Based Quasi-Experimental Study

تطوير منهج لتعليم مهارة الكلام وتقويم فاعليته في معهد إسلامي داخلي بإندونيسيا: دراسة

قائمة على نموذج الأبعاد الأربعة بتصميم شبه تجريبي

Gamil Taher Ahmed^{1*}, Dindin Solahudin², Isop Syaifei³, Dedih Wahyudin⁴

¹Al-Saeed University, Yemen

^{2,3,4}Postgraduate Program in Islamic Education, UIN Sunan Gunung Djati Bandung, Indonesia

*Corresponding E-mail: gameelazzani8080@gmail.com

ABSTRACT

Arabic curricula for non-native learners may include all four language skills while providing insufficient structure for sustained oral production in classroom implementation. This study developed and evaluated speaking-focused instructional units for first-year middle-school students at Al-Wadi Al-Mubarak Institute, Bogor, Indonesia. Research and development followed the define, design, and develop stages of the Thiagarajan 4-D model and was combined with a nonrandomized quasi-experimental pretest-posttest control-group design. The available field sample comprised 32 students, with 16 in each group; no documented a priori power analysis was available. Data were collected through interviews, observation, curriculum analysis, expert-validation forms, a speaking test, and an analytic oral-performance rubric. The surviving study records do not document independent multiple rating or an inter-rater reliability coefficient; this measurement limitation is therefore acknowledged. The product comprised speaking units for Part One of Arabic Between Your Hands in a student book and teacher guide, integrating communicative language teaching, task-based learning, cooperative work, role-play, formative assessment, self-assessment, and portfolios. The experimental-group mean increased from 9.75 to 17.19, compared with 9.38 to 10.63 in the control group. The adjusted posttest difference was 6.272 points (95% CI [5.688, 6.856]); ANCOVA yielded $F(1, 29) = 481.853$, $p < .001$, partial $\eta^2 = .943$. This effect size describes variance associated with group membership after adjustment within this sample, not the percentage of improvement caused by the curriculum. The study offers a methodological and contextual contribution by operationalizing needs analysis, the 4-D process, communicative tasks, and performance assessment within a boarding-school curriculum. Findings remain limited by the small, single-institution, nonrandomized sample and undocumented inter-rater reliability.

Keywords: Arabic curriculum development; speaking skills; communicative language teaching; task-based learning; Thiagarajan 4-D model

مستخلص البحث

يواجه تعليم العربية للناطقين بغيرها فجوة بين اشتغال المنهج على المهارات الأربع وبين محدودية الوقت والتنظيم المخصصين للإنتاج الشفهي في التطبيق الصفّي. هدفت هذه الدراسة إلى تطوير وحدات تعليمية مركزة على مهارة الكلام وتقويم فاعليتها لدى طلاب الصف الأول المتوسط في معهد الوادي المبارك ببوجور. استخدمت الدراسة منهج البحث والتطوير وفق مراحل التعريف والتصميم والتطوير من نموذج ثياغاراجان رباعي الأبعاد، مع تصميم شبه تجريبي غير عشوائي ذي مجموعتين واختبارين قبلي وبعدي. بلغ عدد المشاركين المتاحين للتطبيق ٣٢ طالباً، بواقع ١٦ طالباً في كل مجموعة. ولم يتوافر في سجلات الدراسة تحليل قدرة إحصائية قبلي موثق. جمعت البيانات بالمقابلات والملاحظة وتحليل المنهج وتحكيم الخبراء واختبار الكلام وبطاقة تقدير تحليلية للأداء الشفهي. ولا توثق السجلات المتاحة تعدد المقومين المستقلين أو معامل اتفاق بينهم؛ ولذلك صرح بهذه المحدودية القياسية. وتمثل المنتج في وحدات الكلام في الجزء الأول من كتاب «العربية بين يديك»، في صورة كتاب للطالب ودليل للمعلم، مع دمج المدخل التواصل والتعلم القائم على المهام والعمل التعاوني ولعب الأدوار والتقويم التكويني والذاتي وملف الإنجاز. ارتفع متوسط

Citation:

Ahmed, G. T., Salahudin, D., Syaifei, I., & Wahyudin, D. (2026). Developing and Evaluating an Arabic Speaking Curriculum in an Indonesian Islamic Boarding School: A 4-D Model-Based Quasi-Experimental Study. *Al-Muyassar: Journal of Arabic Education*, 5 (3), 1302–1317.

التجريبية من ٩,٧٥ إلى ١٧,١٩، مقابل ارتفاع الضابطة من ٩,٣٨ إلى ١٠,٦٣. وبلغ الفرق المعدل ٦,٢٧٢ درجات بفاصل ثقة ٩٥٪ بين ٥,٦٨٨ و ٦,٨٥٦، وأظهر تحليل التباين المصاحب أثرا دالا للمجموعة ($F(١, ٢٩) = ٤٨١,٨٥٣$ ، $p < ٠,٠٠١$ ، $\eta^2 = ٠,٩٤٣$). وتمثل الدراسة إسهاما منهجيا وسياقيا في تشغيل تحليل الاحتياجات ونموذج الأبعاد الأربعة والمهام التواصلية والتقويم الأدائي داخل منهج معهد إسلامي، لا ابتكارا نظريا مستقلا. وتفسر النتائج في حدود العينة الصغيرة غير العشوائية والمؤسسة الواحدة وعدم توثيق ثبات الاتفاق بين المقومين.

الكلمات المفتاحية: تطوير المنهج؛ مهارة الكلام؛ المدخل التواصلية؛ التعلم القائم على المهام؛ نموذج ثياغاراجان

المقدمة

تمثل مهارة الكلام الوجه الإنتاجي المباشر للكفاية اللغوية؛ ففيها تتحول المعرفة بالأصوات والمفردات والتراكيب إلى أداء شفهي يحقق غرضا تواصليا. ولا يكفي أن يعرف المتعلم القاعدة أو يتعرف إلى المفردة في سياق مكتوب، بل ينبغي أن يستدعيها في زمن محدود، وينظمها في خطاب مفهوم، ويراعي المخاطب والموقف، ويدير التفاعل بما يتناسب مع المستوى اللغوي. ولهذا تعد تنمية الكلام عملية مركبة تتداخل فيها الدقة والطلاقة والنطق والموارد اللغوية وإدارة الحوار (Bygate, ١٩٨٧; Goh & Burns, ٢٠١٢; Luoma, ٢٠٠٤).

في تعليم العربية للناطقين بغيرها، تزداد هذه المركبات تعقيدا بسبب الفروق الصوتية والصرفية والتركيبية بين لغة المتعلم والعربية، وقلة التعرض الطبيعي للغة خارج الفصل، وقلق الوقوع في الخطأ، ومحدودية الوقت المخصص لكل متعلم. وتشير الأدبيات إلى أن البيئة اللغوية والممارسة المتكررة والتغذية الراجعة المنظمة عناصر حاسمة في نقل المعرفة اللغوية إلى استعمال شفهي فعلي، ولا سيما في البيئات الداخلية التي يمكن أن تجعل العربية جزءا من الحياة التعليمية اليومية (Atabik et al., ٢٠٢٣; Fakhroh & Hikmah, ٢٠٢٣; Ilyas et al., ٢٠٢٤).

غير أن وجود كتاب شامل أو منهج معلن لا يضمن تحقق المنهج في صورته الصفية. فالمنهج المقصود قد يشتمل على المهارات الأربع، في حين يعاد تشكيله عند التنفيذ بفعل زمن الحصة، وأولويات المعلم، وكثافة الفصل، وأدوات التقويم، ومستويات الطلاب. ومن ثم ينبغي تحليل المنهج بوصفه أهدافا ومحتوى وأنشطة وتقويما وبيئة تنفيذ، لا بوصفه قائمة موضوعات فحسب. ويؤكد مجال تطوير مناهج اللغات أهمية البدء بتحليل الاحتياجات والسياق، ثم مواءمة المخرجات والمحتوى والإجراءات والتقويم (Graves, ٢٠٠٠; Nation & Macalister, ٢٠١٠; Richards, ٢٠٠١).

يعد كتاب «العربية بين يديك» من المواد واسعة الاستعمال في مؤسسات تعليم العربية، ويتميز بالتدرج وتنوع الموضوعات ودمج المهارات. إلا أن الملاحظة الأولية في معهد الوادي المبارك والمقابلات مع القائمين على البرنامج أظهرت أن التطبيق يميل إلى القراءة والكتابة والجوانب النظرية، بينما تظل فرص الكلام المنظم محدودة. وظهرت صعوبات تتصل بقلة زمن الممارسة، وتفاوت المستويات، وكثرة المتعلمين في بعض الفصول، وضعف الطلاقة والثقة، والخوف من

الخطأ. ولا تدل هذه الملاحظات على قصور الكتاب في ذاته، بل على حاجة إلى تكييف تنفيذي يجعل الكلام هدفا مركزيا له زمن وأنشطة وتقويم محدد.

يوفر المدخل التواصلي أساسا لهذا التكييف؛ إذ ينقل مركز التعلم من عرض البنية إلى استعمالها لتحقيق معنى، ويجعل التفاعل والمعلومات الناقصة والاختيار اللغوي والتغذية الراجعة مكونات للدرس (Brown & Lee, ٢٠١٥; Richards, ٢٠٠٦). ويكمل التعلم القائم على المهام هذا التوجه من خلال تنظيم التعلم حول مهمة لها غاية ونتيجة قابلة للملاحظة، مثل التعارف أو طلب المعلومات أو وصف شخص أو أداء دور، بحيث تصبح المفردات والتراكيب موارد لإنجاز المهمة لا غاية منفصلة (Ellis, ٢٠٠٣; Long, ٢٠١٥; Nunan, ٢٠٠٤).

وتسمح الأنشطة الثنائية والجماعية بزيادة زمن كلام المتعلم وتوزيع الأدوار، بينما يوفر لعب الأدوار سياقاً آمناً للتجريب اللغوي وإعادة الأداء. كما يدعم التقويم التكويني وبطاقة التقدير والتحقق الذاتي وملف الإنجاز مراقبة تطور الأداء بدل الاقتصار على اختبار معرفي نهائي. فالتغذية الراجعة تكون أكثر فائدة عندما توضح الهدف ومستوى الأداء والخطوة التالية، وعندما يشارك المتعلم في قراءة أدائه (Nicol, ٢٠٠٧; Hattie & Timperley, ٢٠٠٩; Black & Wiliam, ٢٠٠٩; Macfarlane-Dick, ٢٠٠٦).

يمكن تنظيم الأدبيات التي استندت إليها الدراسة في ثلاثة مسارات. تناول المسار الأول بناء مناهج الأداء العربي في ضوء أطر معيارية مثل ACTFL وتصوّر براون، وقدم أساساً لتحديد الكفايات ومؤشرات الأداء، لكنه لم يجمع بالضرورة بين تطوير منتج متعدد الوحدات واختباره بتصميم مضبوط قبلياً (Hasanah et al., ٢٠٢١). وركز المسار الثاني على البيئة اللغوية والبرامج المكثفة وتنمية المهارات داخل المعاهد الإسلامية، فأوضح أهمية الممارسة والدعم المؤسسي، غير أن اهتمامه انصب غالباً على وصف البرنامج أو آليات التنفيذ أكثر من توثيق سلسلة متكاملة تبدأ بتحليل الاحتياجات وتنتهي بكتاب طالب ودليل معلم وتقويم أداء (Atabik et al., ٢٠٢٣; Aziz et al., ٢٠٢٥; Ilyas et al., ٢٠٢٤; Siregar et al., ٢٠٢٣). أما المسار الثالث، وهو أدبيات المدخل التواصلي والتعلم القائم على المهام، فقد وفر مبادئ تصميم راسخة، لكنه لا يحدد بذاته كيفية تكييف وحدات «العربية بين يديك» في سياق المعهد محل الدراسة (Ellis, ٢٠٠٣; Long, ٢٠١٥; Nunan, ٢٠٠٤; Richards, ٢٠٠٦). وبناء على هذه المقارنة، لا تدعي الدراسة ابتكار مكونات نظرية جديدة؛ بل تعالج فجوة منهجية وسياقية تتمثل في تشغيل تحليل الاحتياجات ونموذج D-٤ والمدخل التواصلي والتعلم القائم على المهام والتقويم الأدائي ضمن منتج متعدد الوحدات، ثم تقويمه بتحكيم الخبراء وتصميم شبه تجريبي ذي ضبط قبلي.

اختير نموذج ثياغاراجان رباعي الأبعاد لأنه يوفر مساراً إجرائياً يبدأ بالتعريف، ثم التصميم،

فالتطوير، ثم النشر (Thiagarajan et al., ١٩٧٤). ونفذت الدراسة المراحل الثلاث الأولى ميدانياً، أما النشر الواسع فاعتبر مرحلة لاحقة تتطلب تحققاً في مؤسسات وعينات أخرى. ولم يعامل التقويم مرحلة منفصلة في النهاية، بل عملية مصاحبة للتحليل والتصميم والتحكيم والتنفيذ والمراجعة. وهذه الصياغة تجعل النموذج وسيلة لبناء منتج قابل للفحص والتحسين، لا مخططاً شكلياً.

بناء على ذلك، سعت الدراسة إلى الإجابة عن أربعة أسئلة: ما خصائص المنهج المستخدم واحتياجات تطوير الكلام في المعهد؟ وكيف تبني وحدات الكلام المطورة وفق تحليل الاحتياجات والمدخل التواصلي والتعلم القائم على المهام؟ وكيف تنفذ هذه الوحدات في الفصل؟ وما مقدار التغير في أداء الكلام لدى المجموعة التجريبية مقارنة بالمجموعة الضابطة؟ وتهدف الدراسة إلى تقديم منتج تعليمي قابل للاستخدام وإسهام منهجي وسياقي يبين كيفية تشغيل مناهج قائمة داخل بيئة معهدية محددة، مع الفصل بين دلالة النتائج داخل العينة وبين الادعاء بابتكار نظري أو تعميم سببي أوسع.

الأساس النظري لتطوير منهج الكلام

ينطلق تطوير منهج الكلام من تصور الكفاية التواصلية بوصفها قدرة على اختيار اللغة المناسبة للمقصد والمخاطب والسياق، إلى جانب سلامة البنية. ومن ثم لا يختزل النجاح في خلو الجملة من الخطأ؛ فقد تكون الجملة صحيحة ولكنها لا تحقق الاستجابة المطلوبة أو لا تستمر في التفاعل. ويقتضي المنهج لذلك الجمع بين الدقة والطلاقة والملاءمة وإدارة الدور، مع تدرج يتناسب مع مستوى المبتدئ. وفي المستوى A١ مثلاً، تكون الأهداف الواقعية هي تقديم معلومات شخصية بسيطة، وفهم أسئلة مألوفة، وإدارة تبادل قصير، لا إلقاء خطاب معقد (Council of Europe, ٢٠٢٠).

ويؤثر تعريف الكلام مباشرة في اختيار أداة القياس. فإذا عرف بوصفه حفظ مفردات أو ترديد نص، أمكن الاكتفاء بأسئلة منفصلة، أما إذا عرف بوصفه أداء تواصلية فيلزم وضع المتعلم أمام مهمة تتطلب إنتاجاً واستجابة. ولهذا قسمت بطاقة الدراسة الأداء إلى النطق والموارد اللغوية والتفاعل والتعبير. وتسمح هذه البنية بتحديد موضع التحسن والضعف، كما تقلل أثر الانطباع العام للمقوم. ويؤكد مجال تقويم الكلام ضرورة وضوح المهمة ومعايير التقدير واتساق المقومين، لأن الدرجة الشفهية تنتج من تفاعل المتعلم والمهمة والمقوم (Fulcher, ٢٠١٥; Luoma, ٢٠٠٤).

تحتاج الطلاقة إلى توازن بين التخطيط والضغط الزمني والتكرار. فالمتعلم المبتدئ قد يعرف العبارة ولكنه يتأخر في استدعائها؛ وكلما تكررت العبارة في أغراض متقاربة أصبحت أكثر جاهزية. غير أن التكرار الميكانيكي وحده لا يكفي، لأنه قد ينتج أداء محفوظاً لا ينتقل إلى موقف جديد.

لذلك انتقلت الوحدات من التريد إلى تبديل المعلومات ثم إلى المهمة، وأتاحت إعادة النشاط مع شريك أو سياق مختلف. ويجمع هذا بين الأمان الذي توفره الممارسة الموجهة والمرونة التي تتطلبها الكلام الحقيقي (Bygate, ٢٠١٠; Nation & Newton, ٢٠٢٠).

كما يعد التفاعل بين الأقران موردا تعليميا لا مجرد طريقة لتنظيم المقاعد. ففي الحوار الثنائي يضطر المتعلم إلى الإصغاء والاستجابة وطلب التوضيح وإصلاح سوء الفهم، وهي عمليات لا تظهر في الحديث الأحادي. ويمكن للقرين أن يقدم نموذجا قريبا من مستوى المتعلم، وأن يتيح وقتا أطول للكلام من نمط السؤال الذي يطرحه المعلم على طالب واحد. إلا أن فاعلية التفاعل تتطلب توزيع الأدوار وتحديد الناتج ومراقبة استخدام اللغة، وإلا تحولت المجموعة إلى عمل غير متوازن أو إلى استعمال اللغة الأولى (Sato & Ballinger, ٢٠١٦).

ويرتبط اختيار المحتوى بتحليل الاحتياجات لا بالتدرج اللغوي المجرد وحده. فموضوع التحية والتعارف مثلا يصبح مهما لأن الطالب يحتاج إليه في المعهد، ويمكن أن توظف داخله الضمائر وأدوات الاستفهام والجنسية والبلد. وتتحقق الكفاءة عندما تخدم المفردات والتراكيب هذا الحدث التواصل، وعندما يكرر المتعلم استعمالها في تقديم نفسه وتقديم زميله والسؤال عن معلومات جديدة. وبذلك يجمع تنظيم المحتوى بين الموقف والوظيفة والبنية بدلا من تقديم كل منها في مسار منفصل.

يؤدي دليل المعلم دورا أساسيا في أمانة التنفيذ. فقد تكون وحدة الطالب جيدة، لكن المعلم قد يحول المهمة إلى شرح أو يختصر وقت التفاعل إذا لم يتضح له الهدف والزمن والناتج وطريقة التصحيح. لذلك لا ينبغي أن يقتصر الدليل على إجابات التدريبات، بل يوضح المنطق التربوي وتسلسل الدعم وكيفية تشكيل الأزواج والمجموعات ومتى يؤجل التصحيح ومتى يتدخل. وهذه التفاصيل تجعل المنتج قابلا للنقل بين معلمين وتقل اعتماد نجاحه على الاجتهاد الفردي.

ويشكل التقويم التكويني حلقة وصل بين المنهج المكتوب والمنهج المنفذ. فعندما يلاحظ المعلم أن معظم الطلاب يفهمون المفردات ولكنهم لا يبادرون بالسؤال، يمكنه تعديل النشاط التالي ليزيد فرص طرح الأسئلة. وعندما يكشف التسجيل الذاتي عن مشكلات نطق متكررة، يمكن إضافة تدريب قصير قبل إعادة المهمة. وهكذا لا يقدم التقويم رقما في نهاية الوحدة فقط، بل يولد معلومات تعيد توجيه التدريس، وهو جوهر التقويم من أجل التعلم (Black & Wiliam, ٢٠٠٩).

استخدمت الدراسة منهج البحث والتطوير لبناء وحدات تعليمية موجهة لتنمية الكلام، ودمجت ذلك بتصميم شبه تجريبي ذي مجموعتين واختبارين قبلي وبعدي. اعتمد مسار التطوير على نموذج ثياغاراجان D-4 في مراحل التعريف والتصميم والتطوير، ولم تنفذ مرحلة النشر الواسع؛ لأن هدف الدراسة كان إنتاج نموذج صالح للسياق والتحقق الأولي من فاعليته. واستخدم التقويم بصورة مستمرة في كل مرحلة لمراجعة القرارات والمواد والأدوات.

في مرحلة التعريف، حللت خصائص الطلاب ومستوياتهم واحتياجاتهم، ودرست الكتاب والخطة الزمنية وممارسات التدريس والتقويم، وجمعت آراء الإدارة والمعلمين بالمقابلة والملاحظة. وفي مرحلة التصميم، حددت نواتج الكلام، ونظمت المفردات والتراكيب، وبنيت تسلسل الأنشطة، وأعدت مسودتي كتاب الطالب ودليل المعلم وأداة الاختبار وبطاقة التقدير. وفي مرحلة التطوير، عرض المنتج والأدوات على خبراء في تعليم العربية والمناهج والقياس، وأجريت التعديلات اللازمة قبل التطبيق الميداني.

أجريت الدراسة في معهد الوادي المبارك ببوجور على طلاب الصف الأول المتوسط من الناطقين بغير العربية. بلغ عدد أفراد العينة المتاحة للتطبيق ٣٢ طالبا، قسموا إلى مجموعتين متساويتين: ١٦ في المجموعة التجريبية و١٦ في المجموعة الضابطة. ولا توثق سجلات الدراسة المتاحة إجراء توزيع فردي عشوائي؛ لذلك تعاملت الدراسة المجموعتين بوصفهما مجموعتين غير عشوائيتين ضمن تصميم شبه تجريبي، ولا تستبعد احتمال تحيز الاختيار على الرغم من تقارب المتوسطين القبليين. كما لم يتوافر تحليل قدرة إحصائية قبلي موثق يحدد حجم العينة؛ ومن ثم يعد العدد الصغير قيذا على دقة التقدير والتعميم. درست الضابطة بالطريقة القائمة، في حين طبقت على التجريبية جميع وحدات الكلام المطورة. وحافظ التنفيذ على الخطة الزمنية المعتمدة في المعهد، وخصصت لكل وحدة ثلاث حصص، بما يسمح بالانتقال من المدخل الموجه إلى الأداء المستقل.

تكون المنتج من كتاب للطالب ودليل للمعلم يغطيان وحدات الكلام في الجزء الأول من «العربية بين يديك». اشتملت كل وحدة على بيانات المستوى والزمن، ونواتج التعلم، والمفردات والتراكيب، والوسائل، وخطوات التدريس، والمهمة التواصلية، والتقويم والواجب. وبدأ التسلسل بالتهيئة وتنشيط المعرفة السابقة، ثم عرض المفردات والاستماع والترديد، فالحوار، فالعمل الثنائي أو الجماعي، فلعب الأدوار، وانتهى بنشاط «أتحدث بطلاقة» ومهمة قابلة للتقويم.

جمعت البيانات النوعية بالمقابلات والملاحظة وتحليل محتوى المنهج وملاحظات الخبراء، واستخدمت في تشخيص المشكلة وتوجيه التصميم وتفسير التنفيذ. أما البيانات الكمية فجمعت

باختبار أداء شفهي قبلي وبعدي وبطاقة تقدير تحليلية. وشملت معايير الأداء أربعة مجالات رئيسية: المهارات الصوتية والنطقية، والمهارات اللغوية والتركيبية، والمهارات التواصلية والتفاعلية، ومهارات التعبير الشفهي. وبذلك قيس الكلام بوصفه أداء مركباً، لا مجرد معرفة بالمفردات أو القواعد. غير أن السجلات البحثية المتاحة لا تحدد ما إذا كانت الاستجابات قد قومت بواسطة مقوم مستقل واحد أو أكثر، ولا تقدم معامل ثبات بين المقومين مثل ICC؛ لذلك لا تدعي الدراسة ثباتاً بين المقومين، وتعد قابلية الدرجة للتأثر بحكم المقوم من حدود القياس التي ينبغي معالجتها في الدراسات اللاحقة.

حللت البيانات باستخدام المتوسطات والانحرافات المعيارية واختبار t للعينات المترابطة داخل المجموعة، واختبار t للعينات المستقلة للمقارنة بين المجموعتين، وتحليل التباين المصاحب ANCOVA لضبط أثر الاختبار القبلي. وسبقت ANCOVA اختبارات افتراض تجانس ميل الانحدار وتجانس التباين. كما حسب مربع إيتا الجزئي لتقدير حجم الأثر في نموذج التحليل، وحسب فاصل الثقة ٩٥٪ للفرق المعدل من تباين الخطأ ودرجات الحرية في نموذج ANCOVA. واعتمد مستوى الدلالة ٠,٠٥، مع تفسير النتائج في حدود التصميم شبه التجريبي غير العشوائي والعينة والسياق.

الجدول ١. عناصر وحدات الكلام المطورة وتسلسلها التعليمي

العنصر	المضمون الإجرائي	وظيفته في تنمية الكلام
نواتج التعلم	صياغة سلوك شفهي قابل للملاحظة والقياس	توجيه المحتوى والنشاط والتقويم إلى أداء محدد
الهيئة والمدخل	صورة أو موقف، مفردات، استماع وترديد	تنشيط المعرفة وبناء الموارد الصوتية والمعجمية
الممارسة الموجهة	حوار، أسئلة وأجوبة، تبديل معلومات	نقل المتعلم من المحاكاة إلى الاستعمال المضبوط
المهمة التواصلية	عمل ثنائي أو جماعي، لعب أدوار، حل موقف	خلق غرض تواصلية وحاجة فعلية إلى الكلام
الإنتاج المستقل	حديث قصير أو تقديم أو وصف دون قراءة	تنمية الطلاقة والتنظيم والثقة
التقويم	بطاقة تقدير، تقويم تكويني وذاتي، ملف إنجاز	متابعة الأداء وتقديم تغذية راجعة قابلة للتطبيق

تشخيص المنهج والحاجة إلى التطوير

أظهر تحليل المنهج أن الكتاب المستخدم متدرج ومتكامل في بنيته العامة ويشتمل على المهارات الأربع، وهو أساس مناسب لتعليم العربية للناطقين بغيرها. إلا أن التنفيذ الفعلي لم يوزع الوقت والأنشطة توزيعاً متوازناً؛ فقد غلبت القراءة والكتابة والشرح النظري، وبقيت فرص الكلام محدودة وغير منتظمة. واتفق هذا التشخيص بين الملاحظة والمقابلات وتحليل المواد، مما وجه التطوير إلى إعادة تنظيم التنفيذ لا إلى إلغاء الكتاب أو استبداله.

تمثلت أبرز الاحتياجات في زيادة فرص الكلام لكل طالب، ومراعاة تفاوت المستوى، وبناء تدرج من النطق والجملة القصيرة إلى الحوار والتعبير المستقل، وتخفيف الخوف من الخطأ، وربط التقويم بالسلوك الشفهي. وتؤيد دراسات المعاهد أهمية البيئة اللغوية والممارسة المتكررة والدعم المؤسسي، لكنها تبين كذلك أن مجرد وجود العربية في البرنامج لا يضمن مشاركة جميع الطلاب في إنتاجها (Ilyas et al., ٢٠٢٤; Fakhroh & Hikmah, ٢٠٢٣; Aziz et al., ٢٠٢٥).

يقود هذا التشخيص إلى تمييز مهم بين المنهج كمادة والمنهج كخبرة منفذة. فالمادة قد تكون غنية بالحوار والمفردات، لكن أثرها في الكلام يظل محدوداً إذا استهلك الشرح معظم الحصّة، أو بقي الأداء مقتصرًا على التردد الجماعي، أو قيس النجاح بأسئلة مكتوبة. ولذلك أعاد المنتج توزيع سلطة الوقت والكلام والتقويم داخل الفصل: وقت أطول لأداء الطلاب، وأدوار تفاعلية أوضح، وتقويم يصف الأداء الفعلي.

تكامل تحليل الاحتياجات والتصميم والتقويم

تكشف مراحل التطوير أن تحليل الاحتياجات لم يكن إجراءً افتتاحياً ينتهي عند كتابة التقرير، بل مرجعاً للقرارات اللاحقة. فقد تحولت مشكلة قلة الممارسة إلى قرار بزيادة العمل الثنائي والجماعي، وتحول تفاوت المستوى إلى تدرج في مقدار الدعم، وتحول الخوف من الخطأ إلى أنشطة تبدأ بالمحاكاة الآمنة قبل الأداء المستقل. كما تحولت محدودية التقويم الشفهي إلى بطاقة تقدير وملف إنجاز. وبهذا أمكن تتبع الصلة بين كل مشكلة مشخصة ومكون في المنتج.

وتعد هذه القابلية للتتبع معياراً مهماً لجودة التطوير؛ لأنها تمنع إضافة أنشطة جذابة لا تخدم هدفاً محدداً. فكل نشاط في الوحدات يسأل عنه من ثلاث جهات: ما الأداء الذي يستهدفه؟ وما المورد اللغوي الذي يحتاج إليه؟ وما الدليل الذي سيظهر تحقق التعلم؟ فإذا لم تكن الإجابات واضحة، يعاد تصميم النشاط أو يحذف. ويقارب هذا المبدأ فكرة المحاذاة بين المخرجات والتعلم والتقويم في تصميم المناهج (Nation & Macalister, ٢٠١٠; Graves, ٢٠٠٠).

كما أظهر التحكيم أن صلاحية المنتج لا تتحدد بصحة اللغة وحدها. فقد شملت المراجعة وضوح الهدف، ومناسبة المحتوى للمستوى، وتسلسل الخطوات، وواقعية المهمة، وقابلية التقويم، ووضوح إرشادات المعلم. وأدت ملاحظات الخبراء إلى تعديلات قبل التجربة، مما خفض احتمال أن تعزى صعوبات التنفيذ إلى غموض المادة. ولا يغني التحكيم عن التجربة الميدانية، لكنه يوفر فحصاً منطقياً ومهنيّاً قبل استثمار وقت المتعلمين في التطبيق.

أسهمت دورة التطوير كذلك في التمييز بين التعديل والتطوير. فالتعديل قد يقتصر على تغيير مثال أو إضافة تمرين، أما التطوير فيعيد النظر في العلاقة بين عناصر الوحدة. وقد أبقت الدراسة على موضوعات الكتاب وتسلسله العام، لكنها أعادت صياغة النواتج، وزادت المهمات، وغيرت توزيع وقت الكلام، وأدخلت معايير الأداء. لذلك كان الناتج تكييفاً عميقاً داخل الإطار القائم، لا كتاباً منفصلاً لا صلة له بالمنهج المؤسسي.

تصميم المنتج وخصائصه التربوية

أسفر التطوير عن كتاب للطالب ودليل للمعلم يغطيان جميع وحدات الكلام المستهدفة في الجزء الأول، لا وحدة تجريبية منفردة. وحددت لكل وحدة أهداف تبدأ بصيغة أداء، ومفردات وتراكيب مرتبطة بالموقف، وتسلسلاً من التهيئة إلى المهمة، ومعايير للتغذية الراجعة. وراجع الخبراء اتساق الأهداف والمحتوى والأنشطة والمهارات والتقويم، ثم عدلت المواد قبل التنفيذ.

تميز التصميم بربط كل مورد لغوي بوظيفة تواصلية. فعبارات التحية والتعريف لا تقدم للحفظ فقط، بل تستخدم في مهمة يتعرف فيها الطالب إلى زميل ثم يقدمه أمام الفصل. وأدوات الاستفهام لا تدرب في جمل معزولة، بل تستعمل للحصول على معلومة ناقصة. وتسمح هذه البنية للطالب بأن يلاحظ الصيغة ويتمرن عليها ثم يوظفها في نتيجة تواصلية، وهو ما يوافق مبادئ المدخل التواصلية والتعلم القائم على المهام (Ellis, ٢٠٠٣; Long, ٢٠١٥; Richards, ٢٠٠٦).

أدى العمل الثنائي والجماعي إلى توسيع زمن الكلام مقارنة بنمط السؤال والجواب المتمركز حول المعلم. كما وفر لعب الأدوار مسافة نفسية آمنة تسمح بالتجريب وإعادة الأداء، بينما أتاح نشاط الطلاقة الانتقال من النص المقروء إلى الكلام المنظم. وتنسجم هذه الخصائص مع التصور الذي يرى أن تطوير الطلاقة يحتاج إلى تكرار ذي معنى، ومحتوى مألوف، وضغط تواصلية مناسب، وفرصة لتحسين الأداء في محاولات لاحقة (Bygate, ٢٠١٠; Nation & Newton, ٢٠٢٠).

لم يعامل التقويم خاتمة منفصلة؛ فقد استخدمت الملاحظة وبطاقة التقدير أثناء الأداء، وأضيف التقويم الذاتي وملف الإنجاز لمتابعة التسجيلات والحوارات والتقديمات. وأسهم وضوح المعايير في تحويل التغذية الراجعة من حكم عام إلى معلومات عن النطق والتراكيب والتفاعل

وتنظيم التعبير. وهذا النوع من التغذية يدعم التعلم عندما يعرف المتعلم أين وصل وما الذي ينبغي تعديله (Nicol & Macfarlane-Dick, ٢٠٠٦; Hattie & Timperley, ٢٠٠٧).

تفسير تربوي للفروق الكمية

يمكن فهم الزيادة الكبيرة في متوسط التجريبية في ضوء كثافة فرص الاسترجاع والإنتاج. ففي كل وحدة تعرض الطالب للمفردات والتراكيب، ثم استدعاها في حوار، ثم أعاد تنظيمها في موقف أو دور، ثم قدم ناتجا مستقلا. وهذه الدورات المتتابعة تقلل الفجوة بين المعرفة المتاحة في الذاكرة والمعرفة القابلة للاستخدام السريع. كما أن تكرار الأداء أمام شركاء مختلفين يمنع ارتباط العبارة بنص واحد ويزيد مرونتها.

وقد يكون لتوزيع المشاركة أثر مواز لأثر المحتوى. ففي الصف المتمركز حول المعلم يمكن أن يكون مجموع زمن الكلام كبيرا بينما لا يحصل الطالب الفرد إلا على ثوان قليلة. أما الأزواج المتزامنة فتسمح لعدد أكبر بالكلام في الوقت نفسه. وعندما تقرر الأدوار بناتج واضح، يصبح على كل طالب أن يسأل ويجيب أو يقدم معلومة. لذلك لا يقاس تجديد المنهج بعدد الصفحات الجديدة، بل بمقدار النشاط اللغوي الهادف الذي يتيح لكل متعلم.

كما يمكن أن تكون التغذية الراجعة القصيرة عاملا في التحسن؛ إذ ركزت على أخطاء تعيق المعنى أو تتكرر، وأجل بعضها إلى ما بعد المهمة حفاظا على انسياب الكلام. ويختلف ذلك عن تصحيح كل خطأ فور حدوثه، وهو ما قد يقطع التفاعل ويزيد التردد. وتسمح بطاقة التقدير بجمع الملاحظات تحت مجالات واضحة، ثم تقديم هدف صغير للمحاولة التالية، مثل تحسين أداة استفهام أو زيادة ترابط الإجابة.

أما التحسن المحدود في الضابطة فيمكن تفسيره باستمرار التعلم العادي والتعرض لمحتوى الكتاب والاعتیاد على بنية الاختبار. وهذا مهم لأن وجود زيادة في المجموعتين يمنع نسبة كل تغير إلى التدخل تلقائيا. غير أن الفارق الكبير في مقدار الزيادة، واستمرار الفرق بعد ضبط القياس القبلي، يدعمان ارتباط النتيجة بالمعالجة المركبة. ومع ذلك، يظل احتمال أثر المعلم أو دافعية المشاركة أو تفاعل أفراد المجموعة قائما في التصميم شبه التجريبي.

وتشير دلالة الاختبار القبلي في ANCOVA إلى أن الأداء السابق ظل عاملا مهما في النتيجة البعدية. وهذه الملاحظة تربويا أهم من اعتبار القياس القبلي مجرد متغير إحصائي؛ فهي تعني أن الطلاب لا يدخلون التدخل بالموارد نفسها، وأن التصميم ينبغي أن يوفر مسارات دعم إضافية لمن يبدأ بمستوى أدنى. ويمكن في التطبيقات اللاحقة استخدام نتائج القياس التشخيصي لتكوين أزواج مرنة أو تزويد بعض الطلاب بمدخلات صوتية وتمارين إضافية.

ولا ينبغي قراءة حجم الأثر بمعزل عن خصائص المقياس. فالاختبار الكلي من عشرين درجة، والعينة متجانسة نسبيا، والتدخل مركز على الأداء الذي يقيسه الاختبار. وقد تسهم هذه العوامل في ارتفاع القيمة. لذلك يكون التقرير العلمي الأدق هو عرض المتوسطات والانحرافات وقيم الاختبار وحجم الأثر معا، ثم وصف الحدود، بدلا من استخدام مصطلح «فاعلية عالية» دون سياق. وقد اتبعت الدراسة هذا النهج بفصل الدلالة الإحصائية عن التعميم التربوي.

تقدم النتائج، مع ذلك، دليلا عمليا مهما للمؤسسة؛ فهي تبين أن إعادة تنظيم الوحدات القائمة قد تحدث فرقا قابلا للقياس دون الحاجة إلى تغيير الكتاب كليا. وهذا يجعل التطوير أكثر قابلية للتبني، لأن المعلمين والإدارة يواصلون العمل ضمن المحتوى المعروف مع الحصول على دليل إجرائي أوضح. وتزداد فرص الاستدامة عندما يرتبط التجديد بالخطوة الزمنية والموارد المتاحة والتدريب، لا بمشروع إضافي يصعب دمجها في البرنامج.

تنفيذ الوحدات وتحول أدوار الصف

أظهر التطبيق أن التنظيم المتدرج ساعد على انتقال الطلاب من الاستجابة المقيدة إلى الإنتاج الأكثر استقلالا. بدأ المعلم بنموذج صوتي ومعجمي، ثم خفض مقدار الدعم تدريجيا، في حين انتقل الطالب من التردد إلى إكمال الحوار وتبديل المعلومات وأداء الدور والحديث دون قراءة. لم يختف دور المعلم، بل تحول من مصدر وحيد للغة إلى منظم للوقت وميسر للتفاعل ومقدم لتغذية راجعة قصيرة.

برزت أهمية التدرج لأن الدفع المباشر إلى حديث حر قد يزيد القلق ويجعل المتعلم يعتمد على لغته الأولى. أما الانتقال من المدخل المسموع إلى الممارسة الموجهة ثم المهمة، فقد وفر أساسا لغويا قبل طلب الإنتاج. وسمح تكرار البنية في مواقف مختلفة بتثبيت الصيغة دون فصلها عن المعنى. كما جعلت المهام القصيرة المشاركة ممكنة لطلاب متفاوتي المستوى، إذ أمكن تعديل مقدار المعلومات وطول الدور وفق قدرة المتعلم.

تكشف الملاحظة أيضا أن نجاح الوحدة لم يرتبط بالمحتوى وحده، بل بإدارة الصف وتوزيع الزمن وقواعد التفاعل. فالعمل الجماعي من دون أدوار واضحة قد يسمح بسيطرة طالب واحد، ولعب الأدوار من دون هدف قد يتحول إلى تمثيل شكلي. لذلك تضمن دليل المعلم تعليمات للخطوات، وزمنا لكل مرحلة، ونوع الناتج المتوقع، وطريقة تقديم التغذية الراجعة. وهذه التفاصيل الإجرائية هي التي تحول المبدأ التواصلي إلى ممارسة قابلة للتكرار.

نتائج الاختبارات القبليّة والبعديّة

قبل التدخل كان متوسط المجموعة الضابطة ٩,٣٨ بانحراف معياري ١,٧٤٠، ومتوسط

المجموعة التجريبية ٩,٧٥ بانحراف معياري ١,٥٧١. ولم يكن الفرق دالا إحصائيا ($p = ٠,٦٣٩$, $t = ٠,٥٢٨$)، مما يشير إلى تقارب الأداء الأولي. بعد التطبيق ارتفع متوسط الضابطة إلى ١٠,٦٣، في حين ارتفع متوسط التجريبية إلى ١٧,١٩. وبلغت زيادة التجريبية ٧,٤٤ درجات، مقابل ١,٢٥ درجة في الضابطة.

الجدول ٢. المتوسطات القبلية والبعديّة للمجموعتين

مقدار التغير	الانحراف البعدي	المتوسط البعدي	الانحراف القبلي	المتوسط القبلي	ن	المجموعة
١,٢٥	١,٦٦٨	١٠,٦٣	١,٧٤٠	٩,٣٨	١٦	الضابطة
٧,٤٤	١,٣٢٨	١٧,١٩	١,٥٧١	٩,٧٥	١٦	التجريبية

أظهر اختبار العينات المترابطة فرقا دالا داخل المجموعة التجريبية بين القياسين ($t = ٣٦,٥٥١$, $p < ٠,٠٠١$). كما أظهرت المقارنة البعدية بين المجموعتين فرقا لصالح التجريبية ($t = ١٢,٣١$, $p < ٠,٠٠١$). وتشير هذه النتيجة إلى أن التحسن لم يقتصر على التعرض العادي أو الاعتياد على الاختبار، لأن الضابطة تحسنت تحسنا محدودا فقط، في حين كان تغير التجريبية أكبر بوضوح.

نتائج تحليل التباين المصاحب وحجم الأثر

تحققت افتراضات تجانس ميل الانحدار وتجانس التباين؛ إذ تجاوز مستوى الدلالة في الاختبارين ٠,٠٥. وبعد ضبط الاختبار القبلي، بلغ المتوسط المعدل للضابطة ١٠,٧٧٠ وللجريبية ١٧,٠٤٢، بفارق معدل قدره ٦,٢٧٢ درجات لصالح التجريبية. وبلغ فاصل الثقة ٩٥٪ لهذا الفرق من ٥,٦٨٨ إلى ٦,٨٥٦، بما يبين مقدار عدم اليقين حول التقدير داخل العينة. وأظهر ANCOVA أثرا دالا للمجموعة في الأداء البعدي ($F(١, ٢٩) = ٤٨١,٨٥٣$, $p < ٠,٠٠١$)، كما كان الاختبار القبلي متنبئا دالا بالأداء البعدي ($F(١, ٢٩) = ٧٦,٨٠٨$, $p < ٠,٠٠١$).

الجدول ٣. خلاصة تحليل التباين المصاحب للأداء البعدي

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	F	p	η^2 الجزئي
الاختبار القبلي	٤٩,٤٩٨	١	٤٩,٤٩٨	٧٦,٨٠٨	<.001	٠,٧٢٦
المجموعة	٣١٠,٥٢٩	١	٣١٠,٥٢٩	٤٨١,٨٥٣	<.001	٠,٩٤٣

بلغ مربع إيتا الجزئي لأثر المجموعة ٠,٩٤٣، وهو حجم أثر كبير جدا في نموذج التحليل المستخدم. إلا أن هذه القيمة لا تعني أن المنهج وحده يفسر ٩٤,٣ في المئة من تعلم الكلام في جميع الظروف؛ فهي تقدير مرتبط بالعينة الصغيرة، وأداة القياس، ومدة التدخل، والتباين داخل هذا السياق. ولذلك يستدل بها على قوة الفرق داخل النموذج، لا على نسبة سببية مطلقة قابلة للنقل إلى كل مؤسسة.

تتسق النتيجة مع المنطق التربوي للتدخل؛ فقد زادت الوحدات عدد دورات الاستماع والاستجابة، ووزعت فرص الكلام، وربطت التراكيب بمهمات، وقدمت تغذية راجعة على الأداء نفسه. ويمكن تفسير الفرق على أنه حصيلة لنظام متكامل من القرارات، لا أثر عنصر منفرد مثل لعب الأدوار أو العمل الجماعي. وهذا يفسر أهمية تطوير الأهداف والأنشطة والتقويم ودليل التنفيذ معا.

ويظل التفسير مقيدا بغياب توثيق التوزيع الفردي العشوائي، واحتمال تحيز الاختيار، وصغر العينة، وارتباطها بمؤسسة واحدة، وعدم وجود تحليل قدرة قبلي أو قياس مؤجل للتحقق من بقاء الأثر. كما أن غياب المعلومات الموثقة عن عدد مقومي الأداء وثبات الاتفاق بينهم يحد من قوة الاستدلال القياسي، وقد يستفيد التطبيق في المعهد الداخلي من بيئة مختلفة عن المدرسة النهارية. ومن ثم يلزم تكرار الدراسة بعينات أكبر وتوزيع أو ضبط أقوى، ومقومين مستقلين مع حساب ICC، ومدرسين متعددين، وقياسات مؤجلة، مع مراقبة أمانة التنفيذ.

الإسهام النظري والتطبيقي

لا تتمثل مساهمة الدراسة في ابتكار نموذج نظري جديد أو في الادعاء بأن نموذج D-٤ أو المدخل التواصل أو التعلم القائم على المهام مكونات مستحدثة. وإنما تقدم إسهاما منهجيا وسياقيا في دمج هذه المكونات ضمن سلسلة إجرائية قابلة للتتبع: تحليل حاجات متعلمي معهد إسلامي، وتكييف جميع وحدات الكلام المستهدفة في الجزء الأول من «العربية بين يديك»، وبناء كتاب للطالب ودليل للمعلم، وربط المهمات بمعايير أداء، ثم تقويم المنتج بالتحكيم والتطبيق شبه التجريبي المضبوط قبليا. وتوضح هذه السلسلة كيف يمكن تشغيل مبادئ معروفة في منتج متماسك يناسب سياقاً مؤسسيا محددا، من غير تحويل هذا التكييف إلى ادعاء بجدة نظرية عامة.

عمليا، يمكن للمؤسسات التي تستخدم كتابا جاهزا أن تطور طريقة تنفيذه دون التخلي عن تسلسله أو هويته. يبدأ التطوير بتحديد السلوك الشفهي المطلوب، ثم تخصيص وقت له، وبناء

مهمة تجعله ضروريا، ووضع معيار لملاحظته. كما ينبغي تدريب المعلم على إدارة الأزواج والمجموعات والتغذية الراجعة، وتوفير بيئة تسمح بإعادة المحاولة دون أن يتحول التصحيح إلى مصدر للقلق.

ولا يعني تطوير المنهج زيادة كمية الأنشطة فقط؛ فقد يؤدي ازدحام الوحدة إلى تقليل عمق الممارسة. المطلوب هو اتساق وظيفي بين الهدف والمدخل والمهمة والتقييم. فإذا كان الهدف إدارة حوار، وجب أن تتضمن الوحدة حاجة إلى تبادل المعلومات وأدوارا وتتابعا في الكلام، وأن يقيس التقييم الاستجابة والمتابعة، لا صحة الجملة منفصلة فحسب.

الخلاصة

أجابت الدراسة عن أسئلتها من خلال تشخيص عدم التوازن بين حضور المهارات في المادة وفرص الكلام في التنفيذ، ثم بناء وحدات كلام متدرجة في الجزء الأول من «العربية بين يديك»، وتطبيقها في الصف، ومقارنة تغير الأداء بين مجموعتين. وأنتجت مرحلة التطوير كتابا للطلاب ودليلا للمعلم يربطان تحليل الاحتياجات بالنواتج والمفردات والتراكيب والمهام التواصلية والتقييم التكويني والأدائي. وأظهر التنفيذ إمكان زيادة مشاركة الطلاب والانتقال من التردد الموجه إلى الحوار والمهمة والإنتاج الأكثر استقلالا.

كميا، ارتفع متوسط المجموعة التجريبية من ٩,٧٥ إلى ١٧,١٩، مقابل ارتفاع الضابطة من ٩,٣٨ إلى ١٠,٦٣، وبلغ الفرق المعدل بعد ضبط الاختبار القبلي ٦,٢٧٢ درجات بفواصل ثقة ٩٥٪ بين ٥,٦٨٨ و ٦,٨٥٦. ويمثل ذلك دليلا سياقيا قويا داخل العينة، لا تقديرا سببيا نهائيا قابلا للتعميم غير المشروط. ونظريا، لا تقترح الدراسة نظرية جديدة، بل تبين إمكان وصل تطوير المنهج بتعليم الكلام والتقييم الأدائي. ومنهجيا، تقدم مسارا قابلا للتتبع يدمج تحليل الاحتياجات ومراحل D-٤ والمهام التواصلية والتحكيم والاختبار شبه التجريبي. وتطبيقيا، توفر وحدات ودليلا يمكن تكييفهما في معاهد مشابهة مع مراعاة اختلاف المتعلمين والزمن والموارد. وتبقى النتائج مقيدة بصغر العينة، وغياب توثيق التوزيع العشوائي وثبات الاتفاق بين المقومين، والمؤسسة الواحدة، وقصر مدة القياس؛ لذلك ينبغي أن تستخدم الدراسات اللاحقة عينات أكبر، وإجراءات تخصيص موثقة، ومقومين مستقلين، وقياسات مؤجلة ومتعددة المؤسسات.

المراجع

- Atabik, A., Yahya, M. S., & Mustajab. (2023). Life skills approach in Arabic language learning at Islamic boarding school-based Madrasah Aliyah. *Al-Hayat: Journal of Islamic Education*, 7(2), 327–342. <https://doi.org/10.35723/ajie.v7i2.378>
- Aziz, M. F., Mulyati, A., Ramdhani, Z. R., & Syahputra, I. (2025). The implementation of an intensive Arabic language instruction program: A case study at Al-Aqsa Islamic Boarding School. *Tadris Al-'Arabiyyah*, 4(1), 97–113. <https://doi.org/10.15575/ta.v4i1.44921>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Brown, H. D., & Lee, H. (2015). *Teaching by principles: An interactive approach to language pedagogy* (4th ed.). Pearson.
- Bygate, M. (1987). *Speaking*. Oxford University Press.
- Bygate, M. (2010). Speaking. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (2nd ed., pp. 412–423). Oxford University Press.
- Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Fakhiroh, N., & Hikmah, K. (2023). Enhancing Arabic language proficiency in Islamic boarding schools: A comprehensive study. *Indonesian Journal of Islamic Studies*, 11(3). <https://doi.org/10.21070/ijis.v11i3.1673>
- Fulcher, G. (2015). *Re-examining language testing: A philosophical and social inquiry*. Routledge.
- Goh, C. C. M., & Burns, A. (2012). *Teaching speaking: A holistic approach*. Cambridge University Press.
- Graves, K. (2000). *Designing language courses: A guide for teachers*. Heinle & Heinle.
- Hasanah, M., Mubaligh, A., Sari, R. R., Syarofah, A., & Prasetyo, A. (2021). Arabic performance curriculum development: Reconstruction based on ACTFL and Douglas Brown perspective. *Ijaz Arabi Journal of Arabic Learning*, 4(3), 779–801. <https://doi.org/10.18860/ijazarabi.v4i3.11900>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Ilyas, M., Rasmi, M., & Rusydi, M. (2024). Improving modern pondok students' Arabic language skills in Indonesia: Language institutions as language improvement central. *AL-ISHLAH: Jurnal Pendidikan*, 16(2). <https://doi.org/10.35445/alishlah.v16i2.5095>

- Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Wiley Blackwell.
- Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.
- Nation, I. S. P., & Macalister, J. (2010). *Language curriculum design*. Routledge.
- Nation, I. S. P., & Newton, J. (2020). *Teaching ESL/EFL listening and speaking* (2nd ed.). Routledge.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- Nunan, D. (2004). *Task-based language teaching*. Cambridge University Press.
- Richards, J. C. (2001). *Curriculum development in language teaching*. Cambridge University Press.
- Richards, J. C. (2006). *Communicative language teaching today*. Cambridge University Press.
- Sato, M., & Ballinger, S. (Eds.). (2016). *Peer interaction and second language learning: Pedagogical potential and research agenda*. John Benjamins.
- Siregar, Z. U., Fauzia, E. L., & Fathurrahman, Y. (2023). Development of Arabic language skills for students at Syamsul Ulum Muhammadiyah Bandung Islamic Boarding School. *Ta'lim al-'Arabiyyah*, 7(1). <https://doi.org/10.15575/jpba.v7i1.24453>
- Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional development for training teachers of exceptional children: A sourcebook*. Indiana University.
- Tomlinson, B. (Ed.). (2013). *Developing materials for language teaching* (2nd ed.). Bloomsbury.
- Willis, D., & Willis, J. (2007). *Doing task-based teaching*. Oxford University Press.