



**Taṭwīr Adawāt Taqwīm Ta'līm al-Balāghah 'abra Minaṣṣah
Wordwall fī Ḍauī Madkhal al-Ikhtibār at-Tasykhīsyī li
Arba'ati Mustawayāt**

تطوير أدوات تقويم تعليم البلاغة عبر منصة وورد وول في ضوء مدخل
الاختبار التشخيصي لأربعة مستويات

Lailatul Mathoriyah

Universitas KH. A. Wahab Hasbullah Jombang, Indonesia
lelymathoriyah1982@gmail.com

Afif Kholisun Nashoih

Universitas KH. A. Wahab Hasbullah Jombang, Indonesia
afifkholis@unwaha.ac.id

ENGLISH ABSTRACT

This study aims to develop an evaluation instrument based on the Four Tier Diagnostic Test on Qawā'id Balāghah material by utilizing an information and communication technology (ICT) platform to diagnose students' understanding in depth. This study uses a research and development (R&D) method based on the Borg & Gall model, which consists of ten steps. Data collection was carried out through tests and questionnaires. The test consists of 25 multiple-choice questions based on the Four Tier Diagnostic Test with a total of 100 questions divided into four levels: the first level measures cognitive ability, the second level determines the level of confidence in the answer, the third level explores the reasons behind the answer, and the fourth level measures confidence in the reasoning used. The validation results by material experts showed that this instrument was valid with a percentage of 79.5%, although some aspects such as the suitability of the questions to student competencies only reached 70%. The depth of the material and the ability of the questions to measure misconceptions scored 80%. Validation by media experts showed that this tool was valid with an average score of 82.5%. The validity test of the question items showed that all items were valid with a significance value of <math><0.05</math> and $R\text{-count}> R\text{-table}$ (0.361). The reliability of the instrument based on Cronbach's Alpha of 0.884 indicates that this instrument is very reliable. Analysis of the level of difficulty shows that the questions are in the moderate category (0.3-0.73), with good discriminatory power (0.33-0.73). Thus, this instrument is declared feasible to be used as an evaluation tool in Balaghah learning and is effective in diagnosing misconceptions and students' level of understanding.

Keywords: Evaluation, Balāghah, Four Tier Diagnostic Test, Wordwall

INDONESIAN ABSTRACT

Penelitian ini bertujuan untuk mengembangkan instrumen evaluasi berbasis Four Tier Diagnostic Test pada materi Qawā'id Balāghah dengan memanfaatkan platform teknologi informasi dan komunikasi (ICT) untuk mendiagnosis pemahaman mahasiswa secara mendalam. Penelitian ini menggunakan metode penelitian pengembangan (R&D) berdasarkan model Borg & Gall, yang terdiri dari sepuluh langkah. Pengumpulan data dilakukan melalui tes dan kuesioner. Tes terdiri dari 25 soal pilihan ganda berbasis Four Tier Diagnostic Test dengan total 100 soal yang dibagi menjadi empat tingkatan: tingkatan pertama mengukur kemampuan kognitif, tingkatan kedua menentukan tingkat keyakinan terhadap jawaban, tingkatan ketiga mengeksplorasi alasan di balik jawaban, dan tingkatan keempat mengukur keyakinan terhadap penalaran yang digunakan. Hasil validasi oleh ahli materi menunjukkan bahwa instrumen ini valid dengan persentase 79,5%, meskipun beberapa aspek seperti kesesuaian soal dengan kompetensi siswa hanya mencapai 70%. Kedalaman materi dan kemampuan soal dalam mengukur miskonsepsi mendapat skor 80%. Validasi oleh ahli media menunjukkan bahwa alat ini valid dengan skor rata-rata 82,5%. Uji validitas butir soal menunjukkan semua butir valid dengan nilai signifikansi $< 0,05$ dan $R\text{-hitung} > R\text{-tabel}$ (0,361). Reliabilitas instrumen berdasarkan Cronbach's Alpha sebesar 0,884 menunjukkan bahwa instrumen ini sangat reliabel. Analisis tingkat kesulitan menunjukkan soal berada pada kategori sedang (0,3-0,73), dengan daya pembeda yang baik (0,33-0,73). Dengan demikian, instrumen ini dinyatakan layak untuk digunakan sebagai alat evaluasi dalam pembelajaran Balaghah, serta efektif dalam mendiagnosis miskonsepsi dan tingkat pemahaman mahasiswa.

Kata Kunci: Evaluasi, Balāghah, Four Tier Diagnostic Test, Wordwall

المقدمة

إن قسم تعليم اللغة العربية في أية جامعات له مسؤولية كبيرة في صناعة بيئة تعليمية ذات صلة ومناسبة لاحتياجات الطلبة، وخاصة في إتقان المواد الأساسية للغة العربية مثل القواعد البلاغية والصرفية والنحوية. وهذه التخصصات الثلاثة هي الركائز الأساسية لعلوم اللغة العربية التي تشكل أساس الكفاءة للطلاب الدارسين في هذا القسم. ولذلك فإن تطوير المناهج الفعالة يجب أن يعتمد على التحديد المتعمق للمفاهيم الخاطئة التي تحدث غالباً بين الطلاب فيما يتعلق بهذه المواد. إن تحديد هذه المفاهيم الخاطئة وتشخيصها ليس فقط مهماً في مجال التعلم النظري، بل أيضاً في ممارسة اللغة (Boori dkk., 2024). لأن هذه الثلاثة تلعب دوراً عظيماً في تكوين تفهيم النصوص العربية وتحليل عناصر قواعدها وإنتاج المهارات اللغوية السليمة والبليلة.

ولا شك في أن فهم مادة قواعد البلاغة هو من أشياء حاسمة في تعليم اللغة العربية، لأن فهم هذه القواعد تشكل أساساً قوياً لتنمية الكفاءة اللغوية بشكل فعال. وإتقان البلاغة لا يقتصر على

فهم القواعد اللغوية فحسب، بل يتعلق أيضًا بكيفية استخدام اللغة العربية وفقًا لسياق المواقف والأحوال المختلفة، وكيفية تجميل التعبيرات في التواصل (Nuryani dkk., 2023). ومن المؤكد أن ضعف التمكن من قواعد اللغة، بما في ذلك البلاغة، سيكون له تأثير على انخفاض جودة الاستخدام العملي للغة العربية، مثل كيفية تكييف الكلام العربي حسب السياق والمقتضى الحال، وكيفية التحدث باللغة العربية الفصيحة، وكيفية تجميل الكلام العربي (Nashoih & Darmawan, 2019). ولذلك فإن القدرة على فهم البلاغة تلعب دورًا هامًا جدًا في تحسين مهارات الاتصال، فضلًا عن إثراء جمال اللغة العربية، سواء كان شفويًا أو كتابيًا.

كما هو المعلوم أن البلاغة في العام هي دراسة أسلوب اللغة، بما في ذلك كيفية ترتيب الكلمات والجمل بشكل فعال وجمالي لتحقيق المعنى والسياق المطلوب (I. S. Huda, 2011). ومن خلال أفكار السكاكي ينقسم البلاغة إلى ثلاثة فروع من المعرفة، وهي المعاني والبيان والبديع (Mahliatussikah, 2021; Muttaqin, 2017). تحتوي الثلاثة جميعًا على مواد معقدة غالبًا ما تكون منطقة تحدث فيها مفاهيم خاطئة. تعتبر هذه المادة صعبة لأنها تتضمن فهمًا عميقًا للسياق والمعنى الضمني والاستخدام العربي الذي يتضمن في كثير من الأحيان الذوق أيضًا (Sokip dkk., 2023). ومن الطبيعي أن تثير مادة البلاغة مفاهيم خاطئة أو سوء فهم بين الطلاب.

في بعض الأحيان، هناك فجوات في الفهم والكفاءة اللغوية بين الطلبة. قد يكون لدى بعضهم خلفيات تعليمية مختلفة، حتى يختلف أيضًا مستوى فهمهم لمهارات اللغة العربية وإتقانهم في علوم البلاغة. ولذلك، هناك حاجة ماسة إلى أدوات تقويم يمكنها تحديد مستوى الفهم والمفاهيم الخاطئة لدى الطلبة (Pahlefi, 2022; Tomasouw & Marantika, 2023). وتم إجراء هذا التقويم لجمع بيانات تجريبية حول قدرة الطلاب على فهم قواعد اللغة وتطبيقها في التواصل الكلامي أو الكتابي (Shahmirzadi, 2023). عبر فهم مراحل كفاءة الطلبة وفهمهم لمواد معينة، يمكن للمدرسين تكييف أساليب التدريس الخاصة بهم لتلبية احتياجاتهم (Zaenudin, 2021). ولذلك لا بد من القيام بهذه السلسلة من عمليات التقويم، لأنها بمثابة مقياس يحدد جودة ونوعية التعليم (Anani Sarab & Rahmani, 2023; Fauziah dkk., 2020; Mendoza Ramos & Martinez, 2023).

يعتبر الاختبار التشخيصي ذو المستويات الأربعة منهجًا يمكن أن يوفر صورة أعمق لمعايير فهم الطلبة. يتكون الاختبار التشخيصي من أربعة مستويات: المستوى الأول يتكون من أسئلة ذات

إجابات مختارة، المستوى الثاني يطلب الثقة في اختيار إجابة المستوى الأول، المستوى الثالث يسأل عن أسباب الإجابة المختارة في المستوى الأول، والمستوى الرابع يطلب الثقة في أسباب اختيار الإجابة في المستوى الثالث (Diani dkk., 2019; Hidayah dkk., 2022). لذا يمكن القول أن هذا النوع من التقويم يركز أكثر على زيادة مستوى ثقة الطلبة في اختيار الإجابات وأسباب إجاباتهم. ويتم التعبير عن هذا المستوى من الثقة بمقياسين، وهما الإجابات المؤكدة وغير المؤكدة. ويشير تقويم الثقة هذا إلى تقويم الفرد لجودة أدائه الأكاديمي (Caleon & Subramaniam, 2010). ولذلك، فإن استخدام الاختبار التشخيصي ذا المستويات الأربعة يمكن أن يوفر صورة أكثر تفصيلاً عن مدى فهم الطلبة للمادة ما، وكذلك عن تعيين المواد الدراسية التي يجب التركيز عليها (Fariyani dkk., 2015; Widiyanto dkk., 2018). ويتضمن ذلك المعرفة الأساسية، وفهم المفاهيم، والقدرة على تطبيق تلك المعرفة، والقدرة على التقويم. يوفر هذا النهج فرصاً للمحاضرين لتعيين أساليب التدريس المناسبة، والمواد التعليمية، ونماذج التقويم، من أجل تحسين جودة وأهداف التعلم (Hamid dkk., 2022; Putra dkk., 2018).

يمكن القول إن الأبحاث حول تقويم تعليم اللغة العربية يتم إجراؤها بشكل متكرر، مثل الأبحاث التي أجراها حميد وآخرون (Hamid dkk., 2022)، وبهالوان وآخرون (Bahalwan & Wahidah, 2021)، وياني وآخرون (Yani dkk., 2021)، الذين ساهموا أيضاً في تطوير أدوات التقويم. ومع ذلك، بناءً على أبحاث الباحثين، لم يتم العثور على أبحاث حول تطوير أدوات التقويم على أساس الاختبار التشخيصي المكون من المستويات الأربعة (*four tier diagnostic test*) في مجال تعليم اللغة العربية على وجه عام، وخاصة مادة البلاغة. يتم تنفيذ معظم تطوير أسئلة التقويم باستخدام هذا النهج من قبل الباحثين في مجال العلوم، مثل الأبحاث التي أجراها روج وآخرون، فيما يتعلق بتطوير اختبار تشخيصي من أربعة مستويات لتحديد فهم طلاب المدرسة الثانوية الحكومية في باندونج للأدوات البصرية (Rawh dkk., 2020). كما تم إجراء أبحاث أخرى في مجال الفيزياء بواسطة ويدياتنو وآخرون بعنوان "تحليل فهم الطلاب للمفاهيم باستخدام أداة الاختبار التشخيصي ذات المستويات الأربعة على مادة الموجات الميكانيكية" (Widiyanto dkk., 2018).

إن عدم وجود الأبحاث المكافئة في مجال تعلم اللغة العربية، وخاصة مادة البلاغة، يؤكد أهمية تطوير أدوات التقويم باستخدام أساس الاختبار التشخيصي ذو المستويات الأربعة في هذا المجال.

استناداً إلى الأبحاث السابقة التي ذكرت، أثبت مدخل الاختبار التشخيصي ذو المستويات الأربعة فعال في استكشاف المفاهيم الخاطئة بشكل متعمق وشامل. وفي سياق تعلم اللغة، وخاصة البلاغة التي تجمع بين عناصر المنطق والجمال والمعنى، يمكن لهذا المدخل أن يوفر رؤية أعمق لفهم الطلبة والأخطاء المفاهيمية.

بضوء ما سبق ذكره، تم إجراء هذا البحث كمحاولة للتعرف على المفاهيم الخاطئة لدى الطلاب من خلال تطوير أدوات تقييم تعتمد على الاختبار التشخيصي ذي المستويات الأربعة في مادة البلاغة. يتم بعد ذلك تجميع هذه الأداة باستخدام منصة وورد وول (Wordwall) كتطبيق قائم على تكنولوجيا المعلومات والاتصالات (ICT). يهدف استخدام وورد وول كأساس لإعداد أدوات التقييم إلى توفير الانتعاش في تطبيق تقييم التعليم. بحيث ما يمكن التقليل من الانطباع بوجود اختبار مخيف، وسيكون تنفيذ التقييم أكثر تشويقاً وإمتاعاً (Rasydiana dkk., 2019). بالإضافة إلى ذلك، يتيح استخدام هذه المنصة جمع البيانات تلقائياً وتحليلها بشكل أسرع. وبصرف النظر عن ذلك، يمكن حفظ نتائج الاختبار رقمياً مباشرة. علاوة على ذلك، فإن الأدوات التي تحتوي على هذه المنصة لديها القدرة على التنفيذ على نطاق أوسع ويمكن إعادة إنتاجها بسهولة. وهذا يسمح بإجراء البحث على عينة أكبر.

على وجه خاص، يهدف هذا البحث التطويري إلى استكشاف المفاهيم الخاطئة لدى الطلبة في قسم تعليم اللغة العربية بجامعة كياهي الحاج عبد الوهاب حسب الله جومبانج بخصوص مادة البلاغة. وفي هذا الجهد، يكون الهدف الأول هو التعرف على مدى فهم الطلبة للمفاهيم الأساسية، ثم الوصف بعمق الصعوبات أو العقبات التي يواجهها الطلاب. وتصبح هذه عملية التقييم أساساً لتطوير النماذج والاستراتيجيات والأساليب ووسائل التعليم الأكثر فعالية وتناسب احتياجات الطلبة (Fauzi dkk., 2020; Mohammadkarimi, 2022). ويهدف هذا البحث أيضاً إلى تقييم طرق التقييم الحالية في قياس فهم الطلبة. لذلك من الممكن أن يقال أن هذا البحث يسعى إلى تقديم تقويمات بديلة أكثر شمولية وغنية بالمعلومات، ويمكن أن توفر رؤية أعمق لفهم الطلبة.

منهجية البحث

اعتمد هذا البحث تصميم البحث التطويري (R&D) على أساس النموذج الذي طوره بروج وجول (Borg and Gall)، وهو يتكون من عشر خطوات (W. R. & M. D, 2003). وتحتوي عملية تطوير المنتج على عدة مراحل مهمة، تبدأ بتحليل أدوات التعليم والمواد الدراسية المستخدمة في عمليات التعليم والتعلم في الفصل، ثم تم إعداد شبكة أسئلة التقييم، وكتابة عناصر الأسئلة، واختبار الصحة والصلاحية لمحتوى الأسئلة أو ملاءمتها، يليها إجراء التصحيح على الأسئلة المطورة.

أما القيام بتقويم الصحة والصلاحية لمضمون الأسئلة، فقد تم التقويم والتفتيش من قبل خبيرين، خبير المواد وخبير الوسائل. بعد هذه المرحلة، قام الباحثان بعدة التجارب التي يستفادها لاختبار مدى ملاءمة الأسئلة عند تطبيقها في الميدان. وفي هذه الحالة، تتألف موضوعات التجارب في تطوير أدوات التقييم من ٣٠ طالبًا، وهم لم يزالوا يجلسون في الفصل الدراسي الرابع من قسم تعليم اللغة العربية بجامعة كياهي الحاج عبد الوهاب حسب الله.

وأدوات البحث المستخدمة لجمع البيانات تتكون من نوعين، وهما الاختبار والاستبانة. استخدام أداة الاختبار يهدف إلى قياس نتائج التعليم التي حققها الطلاب، وذلك بورقة اختبار تتكون من ٢٥ سؤالاً على مدخل الاختبار التشخيصي المكون من المستويات الأربعة في شكل المتعدد اختياراً على أساس تكنولوجيا المعلومات والاتصالات. يهدف هذا الاختبار إلى التعرف على مستوى المفاهيم الخاطئة لدى الطلاب حول مادة قواعد البلاغة، بمجموع ١٠٠ سؤال موزعة على أربعة مستويات. وأما أدوات الاستبيان، فهي تستعمل لاستخراج المعلومات حول احتياجات الطلبة لأدوات تقويم البلاغة، وأيضا لجمع المعلومات المتعلقة بدقة وصلاحية مكونات محتوى أداة الاختبار التشخيصي المكون من المستويات الأربعة.

تم إجراء تحليل البيانات من خلال تقويم الصدق والثبات ودرجة الصعوبة وقوة التمييز وتحليل الاستبيان. تم تحليل جميع البيانات التي تم جمعها باستخدام برنامج SPSS، باستثناء تحليل الاستبيانات، سواء أكان استبيانات الصدق من الخبيرين أم استبيانات استجابة الطلبة. التقنية المستخدمة لتحليل بيانات صدق المنتج الخبراء هي مقياس ليكرت (Likert). ولتحديد أرقام العناصر الصحيحة، قام الباحثان بتأكيدهما في جدول Product-moment R. إذا كان $R\text{-count} > R\text{-table}$ ،

فإن عنصر الاستبيان يكون صالحًا، وإذا كان $R\text{-table} > R\text{-count}$ ، فإن عنصر الاستبيان غير صالح. بينما استخدم اختبار الثبات في هذا البحث كرونباخ ألفا بمساعدة برنامج SPSS. يعد معامل ألفا كرونباخ (Cronbach Alpha) معيارًا يستخدمه الباحث لوصف الارتباط أو العلاقة بين المقياس الذي تم إنشاؤه وجميع المقاييس المتغيرة الموجودة. يُقال إن الأداة المستخدمة في هذا المتغير موثوقة إذا كان معامل ألفا كرونباخ أكثر من ٠,٦٠.

مواصفات أدوات تقويم البلاغة المطورة

تم تطوير أدوات التقويم على أساس تكنولوجيا المعلومات والاتصالات (ICT) بناءً على نتائج تحليل الاحتياجات التي أظهرت الحاجة إلى أدوات تقويم أكثر تفصيلاً وتفاعلية، لمساعدة الطلاب على تحديد وتصحيح مفاهيمهم الخاطئة حول مادة البلاغة. تشتمل الأداة التي تم تطويرها على ميزات رقمية متنوعة، مثل الأسئلة المتدرجة المصممة باستخدام تنسيق الاختبار التشخيصي المكون من أربعة المستويات. يتكون كل سؤال من أربع طبقات من الأسئلة التي تسمح باختبار مستوى فهم الطلاب من زوايا مختلفة، بدءًا من المعرفة الأساسية وحتى الأسباب الكامنة وراء الإجابات المختارة.

على عكس التقويمات التقليدية التي تستخدم الوسائط الورقية، يتم إعداد هذه أدوات التقويم مكونة ومصنوعة رقميًا باستخدام منصة وورد وول. تقدم هذه المنصة مجموعة متنوعة من القوالب التي يمكن استخدامها لإنشاء أنشطة تعليمية أكثر تشويقًا وتفاعلية وديناميكية (Khairunnisa, 2024). إن استخدام هذه التكنولوجيا لا يجعل عملية طرح الأسئلة أسهل فحسب، بل يوفر أيضًا تجربة تعليمية أكثر متعة للطلاب.

ونواحي التفاعل التي يقدمها منصة وورد وول يمكن أن يساعد الطلاب في التغلب على الخوف والقلق الذي غالبًا ما ينشأ عند مواجهة تقويمات التعليم (Dewaele & Al-Saraj, 2015; Papi & Khajavy, 2023). عبر أحوال أكثر متعة، يُتوقع من الطلاب أن يكونوا أكبر تحفيزًا وثقة في العمل على الأسئلة، بحيث تعكس نتائج التقويم التي تم الحصول عليها قدراتهم بشكل أكثر دقة (Yani dkk., 2021). إضافة إلى ذلك، فإن استخدام هذه المنصة القائمة على التكنولوجيا يسمح أيضًا للمعلمين بتقديم تعليقات أسرع وأكثر دقة، مما يدعم عملية تعليمية أكثر فعالية.

تتكون أدوات تقويم البلاغة التي تم تطويرها من ١٠٠ سؤال، وهي مصممة لقياس فهم الطلاب بشكل شامل باستخدام منهج الاختبار التشخيصي المكون من أربعة المستويات. وأما المواصفات التفصيلية لتلك الأدوات فهي كما يلي:

في المستوى الأول هناك ٢٥ سؤالاً تركز على اختبار المعرفة الأساسية لدى الطلاب في قواعد البلاغة. تم تجهيز كل سؤال بخمسة خيارات مختلفة للإجابة، تهدف إلى تحديد الفهم والمفاهيم الخاطئة المحتملة التي قد تكون لدى الطلاب.

في المستوى الثاني هناك ٢٥ سؤالاً مصممة لتقويم ثقة الطلاب في الإجابات التي اختاروها في المستوى الأول. نظرًا لأن خيارات الإجابة هي "بالتأكيد" أو "غير تأكيد"، فإن هذا القسم مهم لتقويم مدى ثقة الطلاب في فهمهم، وما إذا كانت المفاهيم الخاطئة التي تنشأ مصحوبة بمعتقدات خاطئة.

في المستوى الثالث تم إعداد ٢٥ سؤالاً لاستكشاف أسباب الإجابات التي اختارها الطلاب في المستوى الأول. تم تصميم هذه الأسئلة لتكشف عن فهم أعمق أو مفاهيم خاطئة ربما لم تكن واضحة في المراحل السابقة.

في المستوى الرابع ٢٥ سؤالاً تركز على ثقة الطلاب في الأسباب التي قدموها في المستوى الثالث. تقيم هذه الأسئلة مدى ثقة الطلاب في التفسيرات والحجج التي يقدمونها فيما يتعلق بقواعد البلاغة، مما يوفر مزيدًا من المعرفة لفهمهم.

كما ذكر سابقاً، فإن مادة البلاغة التي هي محور التركيز الرئيسي في أدوات التقويم تشمل المسند إليه، كلام الخبر والإنشاء، القصر، الوصل والفصل، والإيجاز والإطناب والمساواة. تعتبر هذه المواد عناصر أساسية في علم المعاني، وهو أحد الفروع الرئيسية لعلم البلاغة التي تلعب دوراً هاماً في فهم كيفية تأثير تركيب الجمل واختيار الكلمات على فعالية التواصل المناسب للسياق الكلامي أو لمقتضى الحال، سواء كان كتابياً أم شفهيًا (Akkawiy, 1996; an-Nur, 1979; Hasan, 2000).

ولذلك فإن الأسئلة في هذه الأداة تركز أكثر على تطبيق تحليل النصوص العربية، وليس على الأسئلة التعريفية التي تتطلب مجرد الحفظ. فإن هذا المدخل يهدف إلى اختبار قدرة الطلاب على فهم البلاغة وتحليل استخدام عناصرها في سياقات حقيقية. لذا، تركز هذه الأداة على المهارات العملية

والتطبيقية في تحديد العناصر البلاغية في النصوص العربية، وتحليلها وتقويمها، حتى ينال الطلاب فهماً عميقاً لوظيفة علوم المعاني في البلاغة، وتطبيقها في التواصل العربي.

لاختبار فهم الطلاب الأساسي لقواعد علوم المعاني، يعتبر نوع LOTS (مهارات التفكير الأدنى) أكثر ملاءمة. في هذا المستوى، يُطلب من الطلاب تحديد القواعد الأساسية والتعرف عليها وفهمها، على الرغم من أنه في هذه المرحلة يقوم الطلاب أيضاً بإجراء تحليل متعمق لتعيين عناصر البلاغة ونوعها في النصوص العربية. وهذا أمر مهم كأساس لضمان حصول الطلاب على أول الفهم القوي قبل الانتقال إلى المستويات العليا.

أما أسئلة المستوى الثالث في المستويات الأربعة فقد تم ترتيبها في المستوى المعرفي الرابع (C4). في هذا المستوى، يُطلب من الطلاب تقديم التعليل أو الأسباب وراء الإجابات التي اختاروها في المستوى الأول. وكل الأسئلة في هذا المستوى مصنوعة لاختبار مهارات التفكير المتوسطة أو ما يسمى بـ MOTS، حيث لا يفهم الطلاب القواعد فحسب، بل يكونون قادرين أيضاً على تحليل عميق فيما يتعلق بمواد علوم المعاني. في هذا المستوى المعرفي الرابع، يُطلب أيضاً من الطلاب تحليل مفهوم علوم المعاني، مثل تحديد عناصر المسند ومسند إليه، وبنية القصر، وما إلى ذلك في النص، بالإضافة إلى تحليل كيفية تفاعل هذه العناصر لتكوين المعنى.

تتمتع أدوات تقويم البلاغة التي تم تطويرها بمزايا في عدة جوانب، خاصة فيما يتعلق بالتصميم المرئي. يمكن تغيير مظهر الأدوات في منصة وورد وول وفقاً لرغبات المطور. بالإضافة إلى ذلك، يمكن تقديم كل سؤال وإجابة في شكل صورة. من ناحية سهولة الاستعمال، تدعم هذه الأداة العديد من الأجهزة، على سبيل المثال الحاسوب أو الهاتف المحمول، حتى يسهل على الطلاب الوصول إلى الاختبار في أي مكان وفي أي وقت بدون حد ما.

٢٠,١ وما توفيقى إلا بالله عليه توكلت وإليه انيب.
عين عناصر القصر وطريقه في قول الله تعالى!

الحرف	المقصور	المقصور عليه	طريق القصر
أ	الله - عليه وإليه	توفيقى - توكلت - أنيب	إنما - وتقديم المفعول به على الفعل والفاعل
ب	توفيقى - توكلت - أنيب	الله - عليه وإليه	النفي والاستثناء - وتقديم ما حقه التأخير
ج	حرف النفي وأداة الاستثناء - عليه وإليه	توفيقى - توكلت - أنيب	العطف - تقديم المفعول به
د	الله - توكلت - أنيب	توفيقى - اسما الضمير المجرور بالجر	النفي - العطف
هـ	توفيقى - حرفا الجر على وإلى	توكلت - أنيب	العطف - تقديم ما حقه التأخير

٢٠,٢ هل أنت تتيقن بإجابة سؤال ٢٠,١؟
أ. نعم، أنا أتيقن
ب. لا، أنا لا أتيقن

٢٠,٣ لماذا احترت تلك الإجابة لسؤال ٢٠,١؟

أ. المقصور عليه في النفي الإستثناء مذكور بعد أداة الاستثناء، والمقصور عليه في تقديم ما حقه التأخير هو المذكور المتقدم

ب. المقصور عليه مع أداة إنما مذكور بعدها، والمقصور عليه في تقديم ما حقه التأخير هو مذكور في الأخير.

ج. المقصور عليه مع لا العاطفة هو المذكور قبلها، والمقصور عليه في تقديم المفعول به هو الفاعل من الفعل

د. والمقصور عليه مع ما النافية مذكور بعدها، المقصور عليه مع بل لكن العاطفة هو المذكور بعدها

هـ. والمقصور عليه مع إلا النافية مذكور قبلها، المقصور عليه مع تقديم ما حقه التأخير من الجملة الاسمية هو المذكور المتقدم

٢٠,٤ هل أنت تتيقن بإجابة سؤال ٢٠,٣؟
أ. نعم، أنا أتيقن
ب. لا، أنا لا أتيقن

صورة ١: نموذج سؤال التقييم لمادة البلاغة على أساس FTDT



صورة ٢: المظهر الأول لأدوات التقييم من خلال منصة Wordwall

كما هو موضح في الصورة ٢ أعلاه، يقال إن المظهر الأول لأدوات التقييم في وورد وول بسيط بشكل عام. ويهدف هذا إلى ضمان عدم شعور المستخدمين، وخاصة الطلاب، بعبء الكثير من المعلومات في أول مرة حين الوصول إليها. وهذا المظهر البسيط من الممكن للمستخدمين، أن يقوموا بالتركيز على محتوى المادة دون تشتيت انتباههم بالعناصر المرئية المفرطة. ينصب التركيز الرئيسي لهذا العرض الأول على سهولة الاستخدام وإمكانية الوصول، بحيث يتمكن الطلاب من فهم ما يجب فعله على الفور دون الحاجة إلى الكثير من الشرح الإضافي.

بعد الضغط على رمز "البدأ"، تظهر مباشرة للطلاب مجموعة من الأسئلة التي لازم عليهم أن يعملوها تدريجياً من السؤال الأول إلى الأخير. وعلى الرغم من أن منصة وورد وول توفر الراحة والبساطة في استخدامها، إلا أن هناك عقبات تواجه الباحثين، وهي عدم توفر مساحة خاصة لإدخال التعليمات الخاصة بالعمل على الأسئلة.

ويشكل غياب هذه الميزة تحدياً، لأن تعليمات العمل تلعب دوراً هاماً في مساعدة الطلاب على فهم كيفية الإجابة على كل سؤال، فضلاً عن تقديم نظرة عامة على شكل التقييم المستخدم. في أدوات التقييم القائمة على تكنولوجيا المعلومات والاتصالات مثل هذه، هناك حاجة إلى تعليمات واضحة حتى يتمكن الطلاب من التنقل والعمل على الأسئلة بشكل صحيح وفقاً لمستوى الأداة التي تم تصميمها.

دلائل إكمال أسئلة إختبار البلاغة الأولى

١. هذه الأسئلة مترتبة من ٤ طبقات.
٢. الطبقة الأولى تتكون من الأسئلة باختيار خمس إجابات بين أ - ب - ج - د - هـ.
٣. الطبقة الثانية هي السؤال عن تصديق اليقين على الإجابة في الطبقة الأولى.
٤. الطبقة الثالثة هي السؤال عن الحجة أو التعليل من سؤال الطبقة الأولى.
٥. الطبقة الرابعة هي السؤال عن تصديق اليقين على الإجابة في الطبقة الثالثة.

صورة ٣: دلائل عمل الأسئلة البلاغية

يتطلب غياب عمود التعليمات من الباحثين البحث عن حلول بديلة، مثل تقديم التعليمات بشكل منفصل قبل أن يبدأ الطلاب التقويم، أو من خلال الوسائل الداعمة الأخرى، كما في الشكل ٢ أعلاه. ومن المتوقع أن يؤدي ذلك إلى تقليل الارتباك والتأكد من أن الطلاب لديهم فهم مناسب لكيفية الإجابة على الأسئلة، على الرغم من استمرار القيود المفروضة على المنصة.



صورة ٤ : صفحة السؤال للمستوى الأول

يعرض الصورة ٤ صفحة أسئلة المستوى الأول والتي تركز على مادة الكلام في علم المعاني. تم تنظيم هذا السؤال بتنسيق متعدد الاختيارات يتكون من خمسة اختيارات للإجابة. يتضمن كل خيار إجابة واحدة صحيحة والعديد من عوامل تشتيت الانتباه، والتي تم تصميمها لاختبار دقة الطلاب وفهمهم المتعمق للمفاهيم التي يتم اختبارها. تتكون خيارات الإجابة من إجابة واحدة صحيحة والعديد من إجابات تشتيت الانتباه المصممة لاختبار دقة الطلاب. تهدف هذه الإجابة المشتتة إلى قياس مدى فهم الطلاب للقواعد الأساسية، وكذلك معرفة قدرتهم على التمييز بين الإجابات الصحيحة والاختيارات التي تبدو صحيحة ولكنها في الواقع خاطئة، وكذلك تقويم قدرة الطلاب على التمييز بين المفاهيم المتشابهة لكنها مختلفة إلى حد كبير. وبالتالي، فإن هذا السؤال لا يختبر الفهم النظري للطلاب فحسب، بل يختبر أيضاً قدراتهم التحليلية في تطبيق قواعد علوم المعاني في سياق أوسع. ويضمن هذا الأسلوب ألا يقوم الطلاب بالتخمين فحسب، بل يفكرون فعلياً بشكل نقدي حول المواد التي يتم اختبارها.



صورة ٥: صفحة السؤال للمستوى الثاني

يعرض الصورة ٥ صفحة أسئلة المستوى الثاني لأدوات التقويم، والتي تعمل على تقويم ثقة الطلاب في الإجابات التي يقدمونها على أسئلة المستوى الأول. في هذا المستوى، لا يُطلب من الطلاب الإجابة على أسئلة جديدة، ولكن يُطلب منهم تقويم مدى ثقتهم في خيارات الإجابة التي قاموا بها سابقًا.

ثم تصميم صفحة أسئلة المستوى الثاني يظل بسيطًا، مع نص واضح والتركيز على اختيارات المعتقدات. لا توجد أسئلة إضافية، ولكن فقط تعليمات للطلاب لتقويم ثقتهم في الإجابات التي اختاروها. سيُطلب من الطلاب الاختيار بين خيارات مثل "متأكد" أو "غير متأكد"، والذي يهدف إلى قياس مستوى ثقتهم في فهم الأسئلة. وبالتالي، فإن هذا المستوى الثاني لا يقيس المعرفة الواقعية للطلاب فحسب، بل يقيم أيضًا مستوى ثقتهم في فهم المادة. وهذا مهم للحصول على صورة أكثر شمولًا لفهمهم، سواء من حيث إتقان المفاهيم أو الثقة في الإجابة على الأسئلة.



صورة ٦. صفحة السؤال للمستوى الثالث

تعرض الصورة ٦ صفحة الأسئلة في المستوى الثالث لهذه الأداة التقييمية. وكل الأسئلة لهذا القسم تركز على اختبار أكثر تعمقاً لفهم الطلاب من خلال مطالبتهم بتقديم الأسباب وراء الإجابات التي اختاروها على أسئلة المستوى الأول. تتطلب الأسئلة في هذا المستوى من الطلاب توضيح سبب اختيارهم لإجابات معينة، وبالتالي تقويم قدراتهم الاستدلالية في فهم قواعد علم البلاغة. هناك ٥ خيارات بين الألف والباء والجيم والداد والهاء، وهي تقدم أسباباً مختلفة مختلفة تتعلق بالمفهوم الذي تم اختباره في السؤال الأول. قد يتضمن كل خيارات شرحاً مفاهيمياً، أو تطبيقاً للنظرية، أو أسباب عملية أكثر لدعم الإجابة المحددة. تمثل هذه خيارات الإجابة الخمسة تحدياً للطلاب لإظهار فهم النظريات الأعمق باختيار الأسباب التي تتوافق مع الأشياء المتعلقة بالأسئلة التي كانت في المستوى الأول. تقيس هذه العملية مدى قدرتهم على ربط الإجابات بالمبادئ الأساسية في علوم المعاني.



صورة ٧: صفحة السؤال للمستوى الرابع

وكما هو يكون بأسئلة الثقة في المستوى الثاني، فإن الصورة ٧ سابقاً تحتوي أيضاً على أسئلة ثقة من السؤال السابق. تم تصميم هذه الصفحة لتقويم مدى ثقة الطلاب في أسباب اختيارهم لأسئلة المستوى الثالث. تتكون خيارات الإجابة في هذه الصفحة من إجابتين، "متأكد" و"غير متأكد". تم تحديد هذين الخيارين للتأكد من أن الطلاب يقدمون صورة أكثر تحديداً عن مدى ثقتهم في الأسباب التي اختاروها.

تعمل صفحة تقويم الثقة هذه أيضاً بمثابة انعكاس للطلاب في تقويم فهمهم للمفهوم الذي يتم اختباره. ومن خلال توفير خيارين فقط للإجابة، وهما "المتأكد" و"غير متأكد"، من المتوقع أن

يتمكن الطلاب من تحديد ما إذا كانوا يفهمون المفهوم حقًا أم أنهم ما زالوا غير متأكدين بشكل أكثر وضوحًا. ويشير خيار "أكيد" إلى أن الطالب لديه ثقة كاملة في الإجابة والسبب، بينما يعكس خيار "غير متأكد" الشك أو عدم اليقين في اختيار السبب.

يعد هذا المدخل مهمًا في عملية التعليم لأنه يسمح للطلاب بتقويم فهمهم بشكل مستقل ويصبحوا على دراية بالمجالات التي قد تحتاج إلى تحسين. بالإضافة إلى ذلك، يمكن أن توفر البيانات المستمدة من تقويم الثقة هذا أيضًا معلومات قيمة للمعلمين في تقويم فعالية طرق التدريس والمواد المقدمة. إذا كان معظم الطلاب يشعرون "بعدم التأكد" بإجاباتهم، فقد يفكر المعلم في مراجعة المواد التي يتم تدريسها أو تقديم تفسيرات إضافية.

البيان	النسبة المئوية	الخبير		المعايير المقدره	الرقم
		الثاني	الأول		
صالح	٪٨٠	٤	٤	عمق مادة السؤال يمكن أن يقيس المفاهيم الخاطئة لدى الطلاب	١
صالح	٪٧٠	٤	٣	أن تكون جودة الأسئلة مناسبة لمستوى الطالب وكفاءته	٢
صالح	٪٨٠	٤	٤	منهجية طرح الأسئلة المنطقية	٣
صالح	٪٨٠	٤	٤	وتكون تغطية السؤال متوافقة مع المادة العلمية المعاني في البلاغة	٤
صالح	٪٨٠	٤	٤	يتم طرح الأسئلة بإيجاز ووضوح وحزم	٥
صالح	٪٨٠	٤	٤	أنواع الأسئلة المقدمة لها مستويات مختلفة من الصعوبة	٦
صالح	٪٨٠	٤	٤	يمكن لمواد الأسئلة قياس قدرات التفكير لدى الطلاب	٧
صالح	٪٨٠	٤	٤	توافر إجابات تشثيت الانتباه	٨
صالح جدا	٪٩٠	٤	٥	السؤال لا يوفر مفتاح الإجابة	٩
صالح	٪٨٠	٤	٤	من الممكن أن يتم حل المشكلة	١٠
صالح	٪٨٠	٤	٤	طول صياغة السؤال هو نفسه نسبيًا	١١

١٢	يوجد مفتاح إجابة واحد صحيح فقط	٤	٣	٧٠٪	صالح
١٣	وضوح تعليمات العمل على الأسئلة	٤	٤	٨٠٪	صالح
١٤	توافر المبادئ التوجيهية للتسجيل	٤	٤	٨٠٪	صالح
١٥	لا تعتمد الأسئلة المعرفية على إجابات الأسئلة التي تسبقها و/أو بعدها	٤	٤	٨٠٪	صالح
١٦	اللغة المستخدمة سهلة الفهم	٤	٤	٨٠٪	صالح
١٧	اللغة المستخدمة متوافقة مع القواعد العربية	٤	٤	٨٠٪	صالح
١٨	اللغة المستخدمة هي التواصل	٤	٤	٨٠٪	صالح
١٩	لا تستخدم الكلمات/العبارات التي تؤدي إلى تفسيرات متعددة	٤	٤	٨٠٪	صالح
٢٠	وعدم عناصر لا تتميز على أساس العرق أو الدين أو العرق أو الطبقة الاجتماعية (ethnicity, religion, race, or social class)	٤	٤	٨٠٪	صالح

جدول ١: نتيجة صلاحية المواد من الخبراء

استنادًا إلى جدول نتائج الصحة أعلاه، يُظهر تقويم أسئلة الأدوات التشخيصية التي تم قياسها من قبل الخبراء، نتائج جيدة جدًا بمتوسط نسبة مئوية تبلغ ٧٩,٥٪، وهي ضمن المعايير الصالحة والصحيحة. تغطي هذه الصلاحية جوانب مهمة مختلفة في تطوير الأسئلة لقياس المفاهيم الخاطئة والقدرات المعرفية لدى الطلاب في مادة علوم المعاني في البلاغة. وحصلت النتيجة على العديد من الجوانب التي تم تقويمها، مثل عمق مادة السؤال وقدرة الأسئلة على قياس المفاهيم الخاطئة، على نسبة ٨٠٪، مما يشير إلى أن الأسئلة كانت قادرة على تحقيق الأغراض التشخيصية وكانت عميقة بما يكفي لتقويم الفهم المفاهيمي لدى الطلاب. عدا عن ذلك، تم تقويم جودة الأسئلة على أساس ملاءمتها لمستوى الطالب وكفاءته، حيث حصل على نسبة ٧٠٪، وهي تعتبر صحيحة أيضاً، رغم أنها تحتاج إلى تحسين بسيط لتكون أكثر ملاءمة لمستوى قدرة الطالب.

ومن ناحية العرض المنهجي للأسئلة والتغطية المادية ووضوح اللغة، وعدم عناصر لا تتميز على أساس العرق أو الدين أو العرق أو الطبقة الاجتماعية (ethnicity, religion, race, or social class) على نسبة ٨٠٪، مما يدل على أن الأسئلة تم إعدادها بشكل منطقي ومختصر وواضح، مناسب بالمواد التدريسية. وهذا يوضح أن الأسئلة بشكل عام تلي المتطلبات القياسية لصنع أدوات تقييم جيدة. أحد الجوانب التي تبرز بدرجة ٩٠٪ هو أن الأسئلة لا تقدم أدلة للإجابة الصحيحة، وذلك يشير إلى أن الأسئلة مصممة بشكل جيد لتجنب الغش أو الإجابات غير الصحيحة. وبشكل عام، على الرغم من أن تقييم الصحة يظهر نتائج جيدة، إلا أن العديد من الجوانب مثل توفر مفتاح الإجابة الصحيحة وملاءمة الأسئلة لكفايات الطلاب تحتاج إلى مزيد من الاهتمام، بنسبة ٧٠٪. وهذا يدل على أن هناك مجالاً للتحسين في عدة مجالات بحيث يمكن تحسين جودة الأسئلة لتكون أكثر مثالية.

الرقم	المعايير المقدره	النتيجة	المائوية	البيان
١	سهولة استخدام المنتج المطور	٣	٧٥٪	صالح
٢	المنتج موجز وبسيط	٣	٧٥٪	صالح
٣	المنتج لا يتضمن الكثير من مجموعات الحروف	٣	٧٥٪	صالح
٤	وضوح الكتابة على الصور (نوع الخط وحجمه)	٤	١٠٠٪	صالح جداً
٥	تصميم عرض الوسائط بسيط أو سهل الفهم	٣	٧٥٪	صالح
٦	يعد اختيار ألوان الوسائط جيداً بشكل عام	٤	١٠٠٪	صالح جداً
٧	تصميم غلاف إعلامي جذاب	٣	٧٥٪	صالح
٨	تتنغم عناصر التخطيط الكاملة للأسئلة والخيارات المتعددة	٣	٧٥٪	صالح
٩	جاذبية الرسم التوضيحي	٤	١٠٠٪	صالح جداً
١٠	وضوح الرسم التوضيحي	٣	٧٥٪	صالح
	النتيجة المتوسطة	٣,٢	٨٢,٥٪	صالح

جدول ٢: نتيجة صلاحية التصميم من الخبر

تتضمن نتائج تقييم صحة خبراء الإعلام الموضحة في الجدول أعلاه تقويماً لعشرة جوانب مهمة في تطوير الوسائط التعليمية. وبشكل عام، حصلت هذه الوسائط على متوسط درجات ٣,٢ مع نسبة إنجاز بلغت ٨٢,٥٪، وهي مدرجة في فئة "صالحة". وهذا يعني أن هذه الوسائل التعليمية تعتبر جيدة جداً ومناسبة للاستخدام في عملية التعليم. ومع ذلك، وعلى الرغم من صحته، إلا أن هناك العديد من الجوانب التي لا تزال بحاجة إلى مزيد من التحسين من أجل تحسين جودة وفعالية وسائل الإعلام. يوضح هذا التحقق أن الوسائط بشكل عام تلي المعايير المطلوبة، ولكن مع العديد من التحسينات في جوانب معينة، يمكن تحسين فعاليتها وكفاءتها بشكل أكبر.

ومن حيث سهولة تشغيل الوسائط، والتصميم البسيط، وعدم الإفراط في استخدام مجموعات الحروف، حصلت هذه الوسائط على درجة ٣ مع نسبة إنجاز ٧٥٪. ويشير هذا إلى أن وسائل التعليم تعتبر صالحة، ولكن لا يزال هناك مجال للتحسين. تعد سهولة تشغيل الوسائل عنصراً أساسياً في دعم عملية تعليمية أكثر كفاءة، خاصة للمستخدمين الذين قد لا يكونون على دراية كبيرة بالتكنولوجيا. هناك حاجة إلى تصميم بسيط بحيث يكون من السهل فهم الوسائط والوصول إليها من قبل مجموعات مختلفة من المستخدمين. ومع ذلك، يجب أن تكون هذه البساطة متوازنة مع الجاذبية البصرية حتى يظل المستخدمون مهتمين باستخدام الوسائط بشكل مستمر. في جانب استخدام مجموعات الحروف، من المهم الحفاظ على اتساق النص ووضوحه، بحيث يمكن نقل المعلومات المقدمة بشكل فعال دون التسبب في ارتباك للمستخدمين.

وتحصل تصميمات العرض الجذابة وغطاء الوسيلة على نسبة التقييم تبلغ ٧٥٪. على الرغم من أن هذه النتائج تبدو أن الوسيلة جيدة جداً من الناحية المرئية، إلا أن هناك إمكانية لتحسين العناصر المرئية لجعلها أكثر جاذبية للمستخدمين. تصميم الغلاف الجيد ليس جمالياً فحسب، بل يجب أيضاً أن يكون قادراً على وصف محتوى الوسيلة بوضوح وجذب اهتمام القراء أو المستخدمين من النظرة الأولى. بالإضافة إلى ذلك، يجب تصميم التخطيط وتصميم العرض العام بطريقة تجعل المستخدمين يشعرون بالراحة ولا يثقلون بشاشات العرض المعقدة أو الكثيفة للغاية. في سياق التعليم، ستساعد شاشة العرض جيدة التنظيم للمستخدمين، وخاصة الطلاب، على التنقل في المحتوى بشكل

أكثر فعالية، حتى يتمكنوا من التركيز على المواد المقدمة دون تشتيت انتباههم بعناصر مرئية لا تلائم بالمواد.

ومن الجوانب المتعددة التي حصلت على التقويم الأعلى بنسبة إلى ١٠٠٪، هي وضوح الكتابة على الصورة، واختيار اللون، وجاذبية الرسم التوضيحي. يدل هذا التقويم أن الوسائل تتمتع بقوة في العناصر المرئية المستخدمة. يعد وضوح الكتابة المصاحبة للصور أمرًا مهمًا للغاية لضمان سهولة فهم المستخدمين للمعلومات المرئية. وبصرف النظر عن ذلك، فإن اختيار اللون المناسب لا يؤدي إلى تحميل المظهر فحسب، بل يلعب أيضًا دورًا في تحسين إمكانية القراءة وراحة العين عند استخدام الوسيلة. وتزداد الرسوم التوضيحية الجذابة قيمة أكبر إلى الوسيلة التعليمية، لأن الرسوم التوضيحية الجيدة يمكن أن تساعد في توضيح المفاهيم والمواد المقدمة، فضلاً عن زيادة اهتمام المستخدم بمواصلة استخدام الوسيلة. وبالتالي، فإن هذه الوسيلة قادرة على تقديم تجربة بصرية عالية الجودة، تدعم عملية التعليم بطريقة أكثر تفاعلية ومتعة.

بشكل عام، إن هذه الوسيلة التعليمية قد تم استفاؤها بمعايير الوسيلة الفعالة والصالحة بناءً على تقويمات الخبراء. بالرغم من أن هناك العديد من النواحي التي تحتاج إلى اهتمام خاص لتحسين جودة الوسيلة، سواء من حيث سهولة الاستخدام، والتصميم المرئي، والترتيب الأمثل للمحتوى. وبالتحسينات في العديد من الجوانب، من المرجو أن يكون لهذه الوسيلة التعليمية تأثير أكثر إيجابية في دعم عملية التدريس والتعليم، وخاصة في تسهيل فهم الطلاب للمواد المقدمة.

الصدق والثبات لأدوات تقويم البلاغة

إن اختبار صحة وموثوقية أدوات التقويم يصبح عنصرًا مهمًا للتأكد من أن الأداة قادرة على إنتاج بيانات صحيحة ومتسقة، بحيث يمكن الاعتماد عليها على مدى فترة زمنية مستدامة. يتم اختبار الموثوقية في سياق هذا التقويم من خلال قياس ما إذا كانت إجابات المشاركين على نفس الاختبارات أو اختبارات مماثلة في أوقات مختلفة تظهر الاتساق. وهذا يعني أن الأداة يجب أن توفر نتائج مستقرة وألا تتغير بشكل كبير عند إعادة اختبارها في ظل ظروف مماثلة. ومن ناحية أخرى، يستخدم اختبار الصدق لتقويم مدى قياس هذه الأداة فعليًا للقدرات التي تريد تقويمها، وهي فهم الطلاب وإتقانهم لمفاهيم البلاغة.

وبناء على نتائج اختبار صدق المحتوى على فقرات اختبار الاختيار من متعدد، فمن المعروف أن قيمة الدلالة للأسئلة من ١ إلى ٢٥ أقل من ٠,٠٥. وهذا يوضح أن كل بند من بنود السؤال له مستوى مناسب من الأهمية. وبغض النظر عن ذلك، فإن قيمة R المحسوبة (ارتباط بيرسون) للأسئلة من ١ إلى ٢٥ تكون موجبة وأكبر من جدول R بدرجات حرية (DF) قدرها ٣٠، أي ٠,٣٦١. وبالتالي، يمكن أن نستنتج أن جميع أسئلة الاختيار من متعدد تستوفي معايير الصلاحية ويتم الإعلان عن صحتها.

فلا شك في أن اختبار الصلاحية هي خطوة مهمة في ضمان قدرة أداة التقويم على قياس ما ينبغي قياسه بشكل مناسب. توضح القيمة المحسوبة لـ R والتي تكون أعلى من جدول R أن هناك ارتباطاً كبيراً بين كل عنصر سؤال والنتيجة الإجمالية، مما يشير إلى أن كل عنصر يقدم مساهمة ذات صلة بالأداة الإجمالية. قيمة دلالة أقل من ٠,٠٥ تعزز هذه النتائج، مما يشير إلى أن العلاقة التي تم العثور عليها لم تحدث عن طريق الصدفة.

Case Processing Summary			
		N	%
Cases	Valid	30	100.0
	Excluded ^a	0	0.0
	Total	30	100.0
Reliability Statistics			
Cronbach's Alpha	N of Items		
.884	25		

جدول ٣: نتائج اختبار الثبات على أسئلة اختبار الاختيار من متعدد

وبناء على نتائج اختبار الثبات على أسئلة اختبار الاختيار من متعدد، فإن قيمة ألفا كرونباخ (Cronbach's Alpha) التي تم الحصول عليها كانت ٠,٨٨٤. وهذا يدل على أن الأداة تتمتع بمستوى عالٍ من الثبات وثباتها في قياس قدرات الطلاب. بشكل عام، تعتبر قيم ألفا كرونباخ الأعلى من ٠,٧٠ جيدة في البحث التربوي، لكن في بعض السياقات تكون القيم الأعلى من ٠,٦١ مقبولة. في هذه الحالة، بقيمة ٠,٨٨٤، يمكن اعتبار الأداة موثوقة جداً ومناسبة للاستخدام في مزيد

من البحث. وهذا يعني أن الأداة قادرة على تحقيق نتائج متسقة عند اختبارها مرة أخرى على نفس المجموعة في ظل ظروف مماثلة.

تعمل الثبات، والتي يتم قياسها باستخدام ألفا كرونباخ، على تقييم الاتساق الداخلي لعناصر الاختبار، بحيث يتوقع من كل عنصر أن يقيس نفس المفهوم بطريقة مستقرة. تشير القيمة ٠,٨٨٤ إلى أن العناصر تتمتع بمستوى جيد جدًا من الاتساق. لذلك يمكن تفسير أن المشاركين الذين لديهم فهم عالٍ لمفهوم البلاغة سوف يميلون إلى إعطاء إجابات صحيحة على عناصر الأسئلة المختلفة، في حين أن المشاركين ذوي الفهم الأقل سيظهرون نمطًا ثابتًا في الإجابة على هذه العناصر.

درجة صعوبة أدوات تقييم البلاغة وتمييزها

وتعد درجة صعوبة أدوات تقييم البلاغة جانبًا مهمًا في تطوير الأداة للتأكد من أن الأسئلة أو الأسئلة المقدمة يمكن أن تفرق بين الطلاب الذين لديهم فهم جيد وأولئك الذين ما زالوا يواجهون صعوبة في المادة. في أداة قائمة على الاختبار التشخيصي مكونة من أربعة مستويات، يلعب مستوى الصعوبة دورًا رئيسيًا في تقييم قدرات الطلاب المعرفية والمفاهيمية فيما يتعلق بمواد البلاغة، مثل المسند ومسند إليه، وكلام خبر وإنسيا، والقشر، والإعجاز، والإذنب، و مساواة. تم تصميم كل درجة صعوبة خصيصًا لاختبار عمق فهم الطالب.

ودرجة صعوبة الأسئلة يتم تحليلها مستندا على بيانات من تجارب الأدوات على مجموعات من الطلاب تتوافق مع خصائص السكان الذين يتم قياسهم. وفي هذا البحث يتم تحديد مستوى الصعوبة من خلال النسبة بين عدد الطلاب الذين أجابوا بشكل صحيح والعدد الإجمالي للطلاب الذين أدوا الاختبار. تصف هذه النسبة نسبة الطلاب القادرين على الإجابة على كل سؤال بشكل صحيح، وبالتالي تقديم فكرة عن درجة صعوبة كل سؤال. الأسئلة ذات المستوى المثالي من الصعوبة هي الأسئلة التي ليست سهلة للغاية ولا صعبة للغاية، وذلك بهدف التأكد من قدرة الأداة على التمييز بين الطلاب الذين لديهم فهم جيد وأولئك الذين ليس لديهم فهم كافٍ. ويمكن ملاحظة درجة صعوبة أدوات التقييم التي تم تطويرها في الجدول التالي.

الرقم	الصحيح	الخطأ	درجة الصعوبة	النتيجة
١	١٣	١٧	٠،٤	متوسط
٢	١٢	١٨	٠،٤	متوسط
٣	١١	١٩	٠،٣٧	متوسط
٤	١٨	١٢	٠،٦	متوسط
٥	٩	٢١	٠،٣	صعب
٦	١٤	١٦	٠،٤٧	متوسط
٧	١٢	١٨	٠،٤	متوسط
٨	١٥	١٥	٠،٥	متوسط
٩	١٦	١٤	٠،٥٣	متوسط
١٠	١٨	١٢	٠،٦	متوسط
١١	٢٠	١٠	٠،٦٧	متوسط
١٢	١٩	١١	٠،٦٣	متوسط
١٣	١٤	١٦	٠،٤٧	متوسط
١٤	١٤	١٦	٠،٤٧	متوسط
١٥	١٥	١٥	٠،٥	متوسط
١٦	١٧	١٣	٠،٥٧	متوسط
١٧	١٤	١٦	٠،٤٧	متوسط
١٨	١٩	١١	٠،٦٣	متوسط
١٩	١٣	١٧	٠،٤٣	متوسط
٢٠	١٤	١٦	٠،٤٧	متوسط
٢١	١٨	١٢	٠،٦	متوسط
٢٢	٢٢	٨	٠،٧٣	سهل
٢٣	١٠	٢٠	٠،٣٣	متوسط
٢٤	١٨	١٢	٠،٦	متوسط
٢٥	١٧	١٣	٠،٥٧	متوسط

جدول ٤: درجة التمييز من أدوات تقويم البلاغة

وبناء على البيانات الخاصة بدرجة صعوبة الأسئلة، يتبين أن أكثر أسئلة الاختبار تقع في الفئة "المتوسطة" بدرجة صعوبة يتراوح بين ٠,٣ إلى ٠,٧٣. من مجموع ٢٥ سؤالاً، تم إدراج سؤال واحد فقط في الفئة "الصعبة" (السؤال ٥، بدرجة صعوبة ٠,٣)، في حين تم تصنيف سؤال آخر على أنه "سهل" (السؤال ٢٢، بدرجة صعوبة ٠,٧٣). تتراوح درجة صعوبة معظم الأسئلة في حدود ٠,٤ إلى ٠,٦، مما يدل على أن الأسئلة ذات مستوى متوسط من الصعوبة ومناسبة لقياس قدرات الطلاب بشكل فعال.

والأسئلة ذات المستوى المتوسط من الصعوبة، مثل السؤال ١ (٠,٤٣)، والسؤال ٤ (٠,٦)، والسؤال ١٢ (٠,٦٣)، كلها تظهر أن هذه الأداة قادرة تمامًا على تمييز الطلاب الذين لديهم فهم أفضل عن أولئك الذين لديهم فهم أقل. وبعبارة أخرى، فإن توزيع مستويات الصعوبة المتفاوتة في الفئة المتوسطة يظهر أيضاً أن هناك توازناً بين الأسئلة الأسهل والأسئلة الأكثر تحدياً، بحيث يمكن قياس قدرات الطلاب بشكل أكثر شمولاً. توفر الأسئلة ذات مستوى الصعوبة الأقل (السؤال ٢٢) فرصة للطلاب ذوي القدرات الأساسية للإجابة بشكل صحيح، في حين أن الأسئلة الأكثر صعوبة (السؤال ٥) يمكن أن تختبر الطلاب بفهم أعمق. بشكل عام، تتمتع هذه الأداة بمستوى متوازن من الصعوبة ويمكن استخدامها بفعالية في عملية التقويم.

من خلال تحليل مستوى الصعوبة هذا، يمكن لمطوري الأدوات التأكد من أن تقويمات البلاغة تشخيصية ويمكنها تحديد الطلاب الذين يحتاجون إلى مزيد من الفهم، بالإضافة إلى تمييزهم عن الطلاب الذين حققوا مستويات أعلى من الإتقان. بمعنى آخر، الأسئلة السهلة جداً أو الصعبة جداً لن تقدم صورة دقيقة عن قدرات الطالب. ولذلك، هناك حاجة إلى توزيع متوازن لمستويات الصعوبة حتى تكون الأداة فعالة في تقويم الجوانب المختلفة لفهم الطلاب لمادة البلاغة.

رقم السؤال	المجموعة العليا	المجموعة السفلى	عدد مجموعة العليا	عدد مجموعة السفلى	قوة التمييز	النتيجة
١	١٠	٣	١٥	١٥	٠,٤٧	جيد
٢	١٠	٢	١٥	١٥	٠,٥٣	جيد
٣	١٠	١	١٥	١٥	٠,٦	جيد

جيد	٠,٤	١٥	١٥	٦	١٢	٤
جيد	٠,٤٧	١٥	١٥	١	٨	٥
جيد	٠,٤	١٥	١٥	٤	١٠	٦
جيد	٠,٤	١٥	١٥	٣	٩	٧
جيد	٠,٤٧	١٥	١٥	٤	١١	٨
جيد	٠,٤	١٥	١٥	٥	١١	٩
جيد	٠,٤	١٥	١٥	٦	١٢	١٠
جيد	٠,٥٣	١٥	١٥	٦	١٤	١١
مقبول	٠,٥٣	١٥	١٥	٧	١٢	١٢
جيد	٠,٥٣	١٥	١٥	٣	١١	١٣
جيد	٠,٤	١٥	١٥	٤	١٠	١٤
مقبول	٠,٣٣	١٥	١٥	٥	١٠	١٥
جيد	٠,٤٧	١٥	١٥	٥	١٢	١٦
جيد	٠,٤	١٥	١٥	٤	١٠	١٧
جيد	٠,٤٧	١٥	١٥	٦	١٣	١٨
جيد جدا	٠,٧٣	١٥	١٥	١	١٢	١٩
جيد	٠,٥٣	١٥	١٥	٣	١١	٢٠
جيد	٠,٥٣	١٥	١٥	٥	١٣	٢١
جيد	٠,٤	١٥	١٥	٨	١٤	٢٢
جيد	٠,٤	١٥	١٥	٢	٨	٢٣
جيد	٠,٦٧	١٥	١٥	٤	١٤	٢٤
مقبول	٠,٣٣	١٥	١٥	٦	١١	٢٥

جدول ٥: قوة التمييز من أدوات التقييم المطورة

تعتبر قوة التمييز جانبا هاما في تقييم فقرات الاختبار، لأنها تظهر قدرة الأسئلة على التمييز بين المشاركين ذوي القدرات العالية والمنخفضة. استناداً إلى الجدول المعروض، يمكن ملاحظة أن قيمة القدرة التمييزية تتراوح من ٠,٣٣ إلى ٠,٧٣. الأسئلة ذات القوة التمييزية العالية، مثل السؤال رقم ١٩

بقيمة ٠,٧٣، تظهر أداءً ممتازاً في التمييز بين المشاركين. تم تصنيف هذا السؤال على أنه "جيد جداً" لأنه يتمتع بالقدرة المثلى على تحديد الفروق في قدرات المشاركين.

معظم الأسئلة في هذا الجدول لها قوة تمييز بين ٠,٤٧ و ٠,٦. تقع هذه الأسئلة، مثل الأسئلة رقم ١ و ٢ و ٣ و ٥ وغيرها، ضمن الفئة "الجيدة". وهذا بمعنى أن هذه الأسئلة فعالة في الفصل بين المشاركين ذوي القدرات المختلفة، وإن لم تكن بقوة الأسئلة ذات القوة التمييزية الأعلى. وهذا يدل على أن هذه الأسئلة قادرة على تقديم صورة دقيقة عن الاختلافات في أداء الممتحن.

بالإضافة إلى ذلك، هناك أيضاً أسئلة ذات قوة تمييزية تبلغ حوالي ٠,٤، والتي على الرغم من أنها لا تزال مصنفة على أنها "جيدة"، إلا أنها تتمتع بقوة تمييزية أقل من الأسئلة الأخرى. لا تزال هذه الأسئلة -مثل الأسئلة رقم ٤ و ٦ و ٧ والعديد من الأسئلة الأخرى- تعتبر كافية، ولكنها ليست فعالة مثل الأسئلة ذات القوة التمييزية الأعلى. ومن ناحية أخرى، فإن الأسئلة ذات القوة التمييزية ٠,٣٣، مثل تلك الموجودة في الأسئلة رقم ١٢ و ١٥ و ٢٥، يتم تصنيفها على أنها "جيدة إلى حد ما". تتمتع هذه الأسئلة بقدرة تمييزية أضعف، لذلك في سياق التقييم، قد تتطلب مراجعة لتكون أكثر انتقائية في قياس الاختلافات في قدرة المشاركين.

وعلى وجه عام، تظهر معظم الأسئلة في هذا الجدول جودة جيدة من حيث القوة التمييزية. تعتبر الأسئلة ذات القوة التمييزية العالية أكثر موثوقية في التقييمات، في حين أن الأسئلة ذات القوة التمييزية المنخفضة قد تتطلب مراجعة لتحسين وظيفتها في قياس الاختلافات في قدرات المشاركين. وبالتالي، يوضح هذا التحليل أنه بشكل عام، تم تصميم الأسئلة بشكل جيد لأغراض التقييم، على الرغم من وجود بعض الأسئلة التي تحتاج إلى تحسين.

الخلاصة

من الممكن الاستخلاص أن هذا البحث قد أنتج منتجاً بشكل أدوات تقييم البلاغة بناءً على الاختبار التشخيصي المكون من أربعة مستويات في تعليم البلاغة. تتكون هذه الأداة من أربع طبقات مصممة لتشخيص فهم الطلاب بعمق. تقيس الطبقة الأولى قدرات الطلاب المعرفية المتعلقة بمواد البلاغة، وتحدد الطبقة الثانية ثقة الطلاب في الإجابات المقدمة، وتستكشف الطبقة الثالثة

الأسباب الكامنة وراء إجابات الطلاب، وتقيس الطبقة الرابعة مستوى ثقة الطلاب في الاستدلال المستخدم.

وإطلاعا على نتائج تقويم صحة الأداة، يتضح أن المنتج الذي تم تطويره يتمتع بمستوى جيد إلى حد ما من الصلاحية. وبناءً على تقويم مدقق المادة، حصلت الأداة التشخيصية على ٧٩,٥٪، وهي تدخل ضمن الفئة الصالحة. ومع ذلك، هناك العديد من الجوانب التي لا تزال بحاجة إلى التحسين، مثل ملاءمة الأسئلة لكفاءة الطالب والتي حصلت على نسبة ٧٠٪ فقط. بالإضافة إلى ذلك، حصل عمق المادة وقدرة الأسئلة على قياس المفاهيم الخاطئة على درجة ٨٠٪، مما يدل على أن هذه الأداة قد حققت معظم أهداف التطوير. فيما يتعلق بالتحقق من صحة الوسائط، تم اعتبار المنتج الذي تم تطويره صالحًا أيضًا بمتوسط نسبة ٨٢,٥٪.

يوضح اختبار الصلاحية على أسئلة الاختيار من متعدد أن جميع عناصر الأسئلة من ١ إلى ٢٥ لها قيمة دلالة أقل من ٠,٠٥، وقيمة R-محسوبة أكبر من جدول R (0.361). لذلك يتم الإعلان عن صحتها. ويشير هذا إلى أن كل بند من بنود السؤال يساهم بشكل كبير في الأداة الشاملة ويقاس القدرات ذات الصلة. ويبين اختبار الموثوقية باستخدام ألفا كرونباخ البالغ ٠,٨٨٤، أن هذه الأداة موثوقة للغاية، مما يعكس مستوى عالٍ من الاتساق. مع قيمة موثوقية تزيد عن ٠,٧٠، تعتبر هذه الأداة مناسبة جدًا للاستخدام في مزيد من الأبحاث، لأنها قادرة على تحقيق نتائج مستقرة.

ويبين تحليل مستوى الصعوبة والقوة التمييزية لأدوات تقويم البلاغة أن غالبية الأسئلة تقع في الفئة "المتوسطة"، حيث تتراوح مستويات الصعوبة من ٠,٣ إلى ٠,٧٣، مما يعكس التوازن بين الأسئلة السهلة والصعبة. يمكن لمعظم الأسئلة أن تميز الطلاب ذوي الفهم الجيد عن الطلاب ذوي الفهم الأقل. وتراوحت القوة التمييزية للعناصر من ٠,٣٣ إلى ٠,٧٣، وكانت غالبية العناصر تتمتع بقوة تمييزية "جيدة"، مما يعني أن الأداة كانت فعالة في فصل الطلاب ذوي القدرات المختلفة.

- Akkawiy, I. F. (1996). *Al-Mu'jam al-Mufashshal fi Ulumi al-Balaghah*. Dar al-kutub al-ilmiah.
- an-Nur, J. A. (1979). *Al-Mu'jam al-Adabiy*. Dar al-Ilmi li al-Malayin.
- Anani Sarab, M. R., & Rahmani, S. (2023). Development and Validation of a Scenario-Based Teacher Language Assessment Literacy Test. *International Journal of Language Testing*, 13(1), 67–103. <https://doi.org/10.22034/ijlt.2022.354972.1187>
- Bahalwan, A. I., & Wahidah, P. K. (2021). The Development of Additional Materials in Kahoot! And the Use to Teach Reading Skill/ تطوير المواد الإضافية ببرنامج كاهوت واستخدامها لتعليم مهارة القراءة. *Al-Arabi: Jurnal Bahasa Arab dan Pengajarannya = Al-Arabi: Journal of Teaching Arabic as a Foreign Language*, 5(1), Article 1. <https://doi.org/10.17977/um056v5i1p28-50>
- Boori, A. A., Ghazanfari, M., Ghonsooly, B., & Baghaei, P. (2024). Cognitive Diagnostic Modeling Analysis of the Reading Comprehension Section of an Iranian High-stakes Language Proficiency Test. *International Journal of Language Testing*, 14(1), 17–33. <https://doi.org/10.22034/ijlt.2023.399561.1256>
- Caleon, I. S., & Subramaniam, R. (2010). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, 40(3), 313–337. <https://doi.org/10.1007/s11165-009-9122-4>
- Dewaele, J.-M., & Al-Saraj, T. M. (2015). Foreign Language Classroom Anxiety of Arab learners of English: The effect of personality, linguistic and sociobiographical variables. *Studies in Second Language Learning and Teaching*, 5(2), Article 2. <https://doi.org/10.14746/ssl.2015.5.2.2>
- Diani, R., Alfin, J., Anggraeni, Y. M., Mustari, M., & Fujiani, D. (2019). Four-Tier Diagnostic Test With Certainty of Response Index on The Concepts of Fluid. *Journal of Physics: Conference Series*, 1155, 012078. <https://doi.org/10.1088/1742-6596/1155/1/012078>
- Fariyani, Q., Rusilowati, A., & Sugianto -. (2015). PENGEMBANGAN FOUR-TIER DIAGNOSTIC TEST UNTUK MENGUNGKAP Miskonsepsi Fisika Siswa SMA Kelas X. *Journal of Innovative Science Education*, 4(2), Article 2. <https://journal.unnes.ac.id/sju/jise/article/view/9903>
- Fauzi, Moh. F., Murdiono, M., Anindiati, I., Nada, A. L. I., Khakim, R. R., Mauludiyah, L., & Thoifah, I. (2020). Developing Arabic Language Instructional Content in Canvas LMS for the Era and Post Covid-19 Pandemic. *Izdiyar : Journal of Arabic Language Teaching, Linguistics, and Literature*, 3(3), 161–180. <https://doi.org/10.22219/jiz.v3i3.15017>

- Fauziah, I. R. N., Syihabudin, & Asep Sopian. (2020). ANALISIS KUALITAS TES BAHASA ARAB BERBASIS HIGHER ORDER THINKING SKILL (HOTS). *LISANUNA*, 10(1), 45–54.
- Hamid, M. A., Sutaman, S., Natsir, M., & Salih, I. O. M. (2022). The Development of an Evaluation Instrument for the Implementation of the Arabic Language Curriculum in Islamic High School. *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab*, 14(1), Article 1. <https://doi.org/10.24042/albayan.v14i1.10303>
- Hasan, T. (2000). *Al-Ushul Dirasah Ebistimulujiyah ki al-Fikr al-Lughawiy Inda al-Arab*. Alam al-Kutub.
- Hidayah, N. A. F., Priatmoko, S., & Wardani, S. (2022). *Pengembangan Tes Four Tier Multiple Choice (4TMC) untuk Mendeteksi Miskonsepsi Peserta Didik*.
- Huda, I. S. (2011). SEJARAH BALAGAH: ANTARA MA'RIFAH DAN SINĀ'AH. *Adabiyāt: Jurnal Bahasa dan Sastra*, 10(1), Article 1. <https://doi.org/10.14421/ajbs.2011.10102>
- Huda, M. N., & Aini, R. (2020). Developing an Applicable Balāgh Textbook for Arabic Language Education Students at IAIN Pekalongan. *Alsinatuna*, 5(2), Article 2. <https://doi.org/10.28918/alsinatuna.v5i2.3183>
- Khairunnisa, A. (2024). Pemanfaatan Aplikasi Wordwall Sebagai Media Evaluasi dalam Pembelajaran Bahasa Arab: Manfaat, Tantangan, dan Strategi Optimalisasi. *Jurnal Pendidikan Tambusai*, 8(2), 16–21.
- Mahliatussikah, H. (2021). *Development of Interactive Learning Media "Lectora Inspire" for Balaghah Learning: International Seminar on Language, Education, and Culture (ISoLEC 2021)*, Malang, Indonesia. <https://doi.org/10.2991/assehr.k.211212.024>
- Mendoza Ramos, A., & Martinez, J. A. (2023). The Retrofit of an English Language Placement Test Used for Large-scale Assessments in Higher Education. *International Journal of Language Testing*, 13(1), 139–165. <https://doi.org/10.22034/ijlt.2022.354919.1184>
- Mohammadkarimi, E. (2022). Analytic Assessment of TEFL Undergraduate Students' Writings: Diagnosing Areas of Strength and Weakness. *International Journal of Language Testing*, 12(2), 25–44. <https://doi.org/10.22034/ijlt.2022.157123>
- Muttaqin, M. I. (2017). AFKAR AMIN AL-KHULI FI TA'LIM AL-BALAGHAH AL-'ARABIYAH. *Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 4(1), Article 1. <https://doi.org/10.15408/a.v4i1.5323>
- Muttaqin, M. I. (2020). AMIN AL-KHULI: RÂID TAJDÎD AL-BALĀGHAH FĪ AL-ASHR AL-HADĪTS. *Arabiyat: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, 7(2), Article 2. <https://doi.org/10.15408/a.v7i2.17254>
- Nashoih, A. K., & Darmawan, M. F. (2019). Pengembangan Bahan Ajar Nahwu Berbasis Kontrastif Untuk Mengatasi Interferensi Bahasa Indonesia Terhadap Bahasa Arab.

- Arabiyatuna : Jurnal Bahasa Arab*, 3(2), 335.
<https://doi.org/10.29240/jba.v3i2.1008>
- Nuryani, N., Musyafaah, N., Unsi, B. T., & Nuha, M. A. U. (2023). Student Team Achievement Division (STAD) and Cooperative Learning Model in Balaghah Learning. *An Nabighoh*, 25(1), Article 1. <https://doi.org/10.32332/an-nabighoh.v25i1.6446>
- Pahlefi, M. R. (2022). Pengembangan Instrumen Penilaian Keterampilan Menyimak (Mahārah al-Istima') dalam Pembelajaran Bahasa Arab. *Uktub: Journal of Arabic Studies*, 2(2), Article 2. <https://doi.org/10.32678/uktub.v2i2.6458>
- Papi, M., & Khajavy, H. (2023). Second language anxiety: Construct, effects, and sources. *Annual Review of Applied Linguistics*, 43, 127–139. <https://doi.org/10.1017/S0267190523000028>
- Putra, I. A., Sujarwanto, E., & Pertiwi, N. A. S. (2018). Analisis pemahaman konseptual mahasiswa pada materi kinematika partikel melalui tes diagnostik. *Jurnal Riset dan Kajian Pendidikan Fisika*, 5(1), Article 1. <https://doi.org/10.12928/jrkpf.v5i1.8923>
- Rasydiana, Hanafi, Y., & Huda, I. S. (2019). Pengembangan Tes Pembelajaran Bahasa Arab Menggunakan Aplikasi Kahoot di MTsN 2 Kota Malang. Dalam *Prosiding Konferensi Nasional Bahasa Arab V* (hlm. 437–447). Jurusan Sastra Arab, Fakultas Sastra, Universitas Negeri Malang.
- Rawh, P., Samsudin, A., & Nugraha, M. G. (2020). Pengembangan Four-Tier Diagnostic Test untuk Mengidentifikasi Profil Konsepsi Siswa pada Materi Alat-alat Optik. *WaPfi (Wahana Pendidikan Fisika)*, 5(1), 84–89. <https://doi.org/10.17509/wapfi.v5i1.22888>
- Shahmirzadi, N. (2023). Validation of a Language Center Placement Test: Differential Item Functioning. *International Journal of Language Testing*, 13(1), 1–17. <https://doi.org/10.22034/ijlt.2022.336779.1151>
- Sokip, S., Nurcholis, A., Kojin, K., Faizin, M. Z., & Hanifa, A. (2023). The Contribution to Understanding the Study of Psycholinguistics in Facilitating Balaghah Learning. *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab*, 15(1), Article 1. <https://doi.org/10.24042/albayan.v15i1.14462>
- Tomasouw, J., & Marantika, J. E. R. (2023). Implementasi Pelaksanaan Entry Behavior, Matrikulasi dan Kemampuan Reseptif Mahasiswa (Suatu Studi Kasus pada Mahasiswa Program Studi Pendidikan Bahasa Jerman Tahun Ajaran 2022/2023). *J-EDu: Journal - Erfolgreicher Deutschunterricht*, 3(1), Article 1. <https://doi.org/10.30598/J-EDu.3.1.8-15>
- W. R., B., & M. D, G. (2003). *Educational Research: An Introduction* (4 ed.). Longman Inc.
- Widiyanto, A., Sujarwanto, E., & Prihaningtiyas, S. (2018). ANALISIS PEMAHAMAN KONSEP PESERTA DIDIK DENGAN INSTRUMEN FOUR TIER DIAGNOSTIC TEST

PADA MATERI GELOMBANG MEKANIK. *Prosiding Seminar Nasional Multidisiplin, 1*, 138–146.

Yani, M. Y., Koderi, K., Zulhanan, Z., Tubagus, M., & Fauzan, M. S. (2021). The Development of Arabic Learning Evaluation based on Kahoot Application in Covid Outbreak | Tatwir Taqwim Ta'lim al-Lughah al-'Arabiyyah 'ala Asas Tatbiq Kahoot Khilala Jaikhah Covid. *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab*, 13(1), Article 1. <https://doi.org/10.24042/albayan.v13i1.7219>

Zaenudin, A. (2021). *KOMPETENSI AWAL PESERTA DIDIK DAN IMPLIKASINYA DALAM PEMBELAJARAN BAHASA ARAB. 11.*