Arabiyatuna
Jurnal Bahasa Arab

Vol. 10. No. 2, November 2026, 753-778
P-ISSN: 2580-5045, E-ISSN: 2580-5053
DOI: http://dx.doi.org/10.29240/jba.v10i2.16889

تحديد المفاهيم الخاطئة لدى الطلبة في مادة التراكيب النحوية عبر تصميم أداة التقويم القائمة على اختبار التشخيص رباعي المستويات

Identifying University Students' Misconceptions about Arabic Syntactic Structures Using an Assessment Design Based on a Four-Tier Diagnostic Test

Afif Kholisun Nashoih[1*], Hanik Mahliatussikah[2]
Universitas Kh. A. Wahab Hasbullah Jombang, Indonesia[1]
Universitas Negeri Malang, Indonesia[2]

[*]Corresponding Author: *afifkholis@unwaha.ac.id*

Cite this article:
Nashoih, A. K., & Mahliatussikah, H. (2026). Identifying University Students' Misconceptions about Arabic Syntactic Structures Using an Assessment Design Based on a Four-Tier Diagnostic Test. Arabiyatuna: Jurnal Bahasa Arab, 10(2), 753–778. https://doi.org/10.29240/jba.v10i2.16889

Received: 20-06-2026	Revised: 10-09-2026	Accepted: 02-11-2026

Abstract

In higher education, students may select correct answers to Arabic grammar (*nahw*) questions despite gaps in their conceptual understanding. Scoring answers alone can obscure these gaps. This study aimed to develop a four-tier diagnostic test (FTDT) to identify students' misconceptions about Arabic syntactic structures (*tarākīb naḥwiyyah*). The study adapted Borg and Gall's research and development model. The instrument comprised 25 items with four components each, yielding 100 components in total. Tier 1 posed five-option multiple-choice questions on grammatical concepts. Tier 2 recorded confidence in the answer, Tier 3 elicited the reasoning behind it, and Tier 4 recorded confidence in that reasoning. The overall mean expert rating was 4.38, corresponding to the study's highly valid category. Diagnostic relevance received the highest mean rating (M = 4.75). All 25 items met the empirical validity criterion in both Tiers 1 and 3 ($r > .361$), and the reliability coefficient was .884. Overall, 56% of items were moderately difficult, 28% were difficult, and 16% were easy. In addition, 68% showed good or excellent discrimination. The item-level diagnostic classifications indicated conceptual understanding (52%), lack of

Copyright © 2026 **Author'(s)**
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License

conceptual understanding (20%), misconceptions (16%), and partial misconceptions (12%). Conceptual errors accompanied by high confidence were most evident in Items 10 and 20. False-positive response patterns involving correct answers paired with incorrect reasoning appeared in Items 15, 19, and 23. These findings support the use of the FTDT as a valid, reliable, and effective tool for diagnosing university students' misconceptions about Arabic grammar.

Keywords: Assessment Instrument; Four-Tier Diagnostic Test; Misconceptions; Arabic syntactic.

المقدمة

من الجدير بالعلم أن اللغة العربية مكانةً بارزةً في مؤسسات التعليم العالي الإسلامية، لما لها من صلة بفهم النصوص الدينية وممارسة الشعائر في الحياة اليومية. ويرمي تعليمها كذلك إلى تمكين الطلبة من التواصل بالعربية على نحو سليم. وفي جامعة الشيخ الحاج عبد الوهاب حسب الله، تظهر في كتابات طلبة قسم تعليم اللغة العربية صعوباتٌ في تطبيق القواعد النحوية (Nashoih et al., 2023).

وكما هو المعلوم أن اللغة العربية جزء من أجزاء اللغات السامية، وتتميّز ببنية نحوية وتركيبية بالغة التعقيد (Asiah et al., 2022; Nurbayan et al., 2020). ومن أشد جوانبها تحديًا قواعد النحو؛ ذلك العلم الضابط للتراكيب النحوية في الجملة العربية (Nashoih, 2019). ولا يقتصر هذا التعقيد على كثرة القواعد وتشعّبها، بل يمتدّ إلى أسسها الفلسفية المنطقية التي كثيرًا ما تطغى على الجانب التربوي وتتقدم (Ismail et al., 2023; Rezi & Amrina, 2019). كما أن طابعه المنهجي والفلسفي وما يزخر به من ضوابط صارمة كثيرا ما يعسر على الطلبة مسيرة تعلمه (Luthfi, 2018)، ولا عجب أن يشعر كثير منهم بالإرهاق في مواجهة النحوي (Nashoih & Darmawan, 2019).

غير أنّ النحو يظل مادة لا يمكن الاستغناء عنها في تعليم اللغة العربية؛ إذ يُعدّ من أهم مكونات الإتقان اللغوي، وهو الركيزة التي يتوقف عليها إتقان الطالب للعربية الصحيحة (Murad & Khalil, 2015). ومع ذلك، يُفضي غياب الفهم السليم في أحيان كثيرة إلى تكوين مفاهيم خاطئة سابقة تعرقل التحصيل (Abdurrohman &

(Ulinnuhaa, 2024; Anggara, 2024). وتتشكّل هذه المفاهيم الخاطئة نتيجة عوامل متشعّبة، أبرزها قصور أساليب التدريس، وعدم نضج الفهم المفاهيمي، وشح أدوات التقويم القادرة على رصد مواطن الخلل المعرفي بدقة (Rusilowati et al., 2024).

إنّ عملية التحديد هذه تُسهم في تصحيح المفاهيم المغلوطة، وتوطيد الأسس المعرفية، وتحفيز التعلم العميق، مما يُفضي إلى بيئة تعليمية استجابية تنبثق من الحاجات الفردية للطلبة وتُنمي تحصيلهم الأكاديمي وملكتهم النقدية (Tomasouw & Marantika, 2023; Zaenudin, 2021). لذا يغدو من الضرورة بمكان أن يعمل المعلمون على صياغة استراتيجيات تعليمية تكيّفية تستجيب لاحتياجات الطلبة، وأن يبدؤوا بتحديد المفاهيم الخاطئة في تعلّم النحو كي يتسنّى التدخل بالع تستهدف هذه الدراسةلاج الملائم (Hamid et al., 2022).

ولدعم استراتيجية التحديد، يُقترح اختبار التشخيص رباعي المستويات (Four Tier Diagnostic Test) أداة للتقويم تبلغ من الفعالية حدا بعيدا في تحليل المفاهيم الخاطئة التي يحملها الطلبة، بما فيها مفاهيم النحو (Diani et al., 2019). ويتألف هذا الاختبار في بنيته العامة من أربعة مستويات متدرجة (Hidayah et al., 2022; Nasyidiah et al., 2020). يقيس للطالب المستوى الأول المعرفة الأساسية المتمركزة حول الحقائق والمفاهيم الجوهرية. ويطلب المستوى الثاني من الطالب تقديم المسوّغ أو التعليل لإجابته المختارة. ويُقيس المستوى الثالث درجة ثقة الطالب بإجابته المعطاة. أما المستوى الرابع فيختبر قدرة الطالب على التقييم والتأمل في مسار تفكيره الذاتي (Laila & Nashoih, 2024). ولا يقتصر هذا الاختبار على قياس النتائج النهائية للفهم، بل يُسلّط الضوء على المسار المعرفي الكامن وراء اختيارات الطالب (Boori et al., 2024).

فضلًا عن ذلك، تُعين درجة الثقة على تفسير الاستجابة عند قراءتها مع التعليل. فقد يقترن الخطأ بتردّدٍ في تطبيق القاعدة، وقد يقترن بتعليلٍ خاطئ يثق الطالب بصحّته. ويفيد فحص الثقة في سياق الاستجابة في تحديد الحالات التي تستدعي مزيدًا من التحقّق من فهم الطالب (Istiyono et al., 2023). وممّا سبق بيانه، تبرز الحاجة إلى تقويمٍ تشخيصيٍّ يتناول الأساس الذي يستند إليه الطالب في الحكم

على التراكيب النحوية. وتوفّر الدراسات التي تناولت أخطاء الطلبة في الكتابة وطرائق تقويم فهمهم منطلقًا لتحديد ما ينبغي فحصه في هذا السياق الجامعي.

كشفت الدراسات التي تناولت كتابات طلبة الجامعات عن مواضع يتكرّر فيها الخطأ عند تطبيق قواعد النحو. فقد رصدت إحداها أثر اللغة الإندونيسية في كتابات طلبة برنامج تعليم اللغة العربية (Faridl et al., 2026). وسجّلت دراسة أخرى أخطاءً في الإعراب وتراكيب الإضافة وتحديد أنواع الجمل (Nurbayan et al., 2020). كما أظهر تحليل خمسة وعشرين ملخّصًا لبحوث التخرّج في جامعة الشيخ الحاج عبد الوهاب حسب الله (UNWAHA) أخطاءً في علامات الإعراب وتركيب الإضافة والمطابقة بين النعت والمنعوت (Nashoih et al., 2023).

وفي مجال التقويم التشخيصي للنحو، طوّرت دراسةٌ اختبارًا ثنائيَّ المستوى تناول الفعل وأدوات الاستفهام والظرف لدى طلبة الصف العاشر في مدرسة ثانوية إسلامية (Maula et al., 2026). وقدّم تطبيق الاختبار وصفًا أوليًّا لأنماط الاستجابة، شمل الفهم والتخمين والمفاهيم الخاطئة ونقص المعرفة .وتتمثّل الفجوة البحثية الأولى في كيفية الكشف عن المفاهيم النحوية الخاطئة لدى طلبة الجامعات من خلال تحليل العلاقة بين إجابة الطالب وتعليلها ودرجة ثقته بكلٍّ منهما.

يجمع الاختبار التشخيصي رباعي المستويات (*Four-Tier Diagnostic Test*) في كل بند أربعة مكوّنات هي الإجابة، والثقة بالإجابة، والتعليل، والثقة بالتعليل .وفي دراسة عن الموجات الميكانيكية، تبيّن أن متوسط درجات الإجابة ومتوسط الثقة بها كانا أعلى من نظيريهما في التعليل والثقة به (Caleon & Subramaniam, 2010). وتبرز هذه النتيجة أهمية قياس الثقة بالإجابة والثقة بالتعليل كلًّا على حدة. وفي تعليم اللغة العربية، وظّفت إحدى الدراسات هذا الاختبار في تقويم تعلّم البلاغة عبر منصّة وورد وول (Laila & Nashoih, 2024).

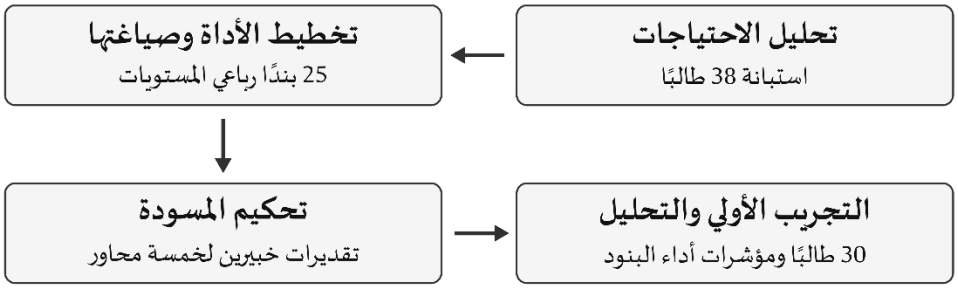
وتتمثّل الفجوة البحثية الثانية في كيفية تكييف الاختبار رباعي المستويات بحيث تكشف التعليلات التي يختارها الطلبة عن طريقة تحليلهم لبنية الجملة العربية، ويقتضي هذا التكييف أن يُصاغ سؤال الإجابة وبدائل التعليل بوصفهما جزأين

مترابطين من البند نفسه. فيُطلب من الطالب أن يحدّد الوظيفة النحوية للكلمة أو علامة إعرابها، ثم يُفحَص من خلال اختياره للتعليل مدى استناده إلى القرينة المناسبة في الجملة. ويمكن بناء البدائل المشتِّتة على استدلالات قد تبدو مقنعةً للطالب، كأن يحدّد وظيفة الكلمة اعتمادًا على موقعها وحده، أو يفترض أن علامةً إعرابيةً واحدةً تدلّ دائمًا على الوظيفة النحوية نفسها.

وفي ضوء هاتين الفجوتين، تهدف هذه الدراسة إلى تطوير أداة تقويم قائمة على الاختبار التشخيصي رباعي المستويات لرصد مؤشرات المفاهيم الخاطئة في التراكيب النحوية لدى طلبة قسم تعليم اللغة العربية بجامعة الشيخ الحاج عبد الوهاب حسب الله. وتفحص الدراسة صدق الأداة وثباتها، وتُحلّل أداء بنودها في تجربة أولية. ويمكن الإفادة من أنماط استجابات الطلبة في تحديد مواطن التعثّر وتوجيه المعالجة التعليمية.

منهج البحث

اعتمدت الدراسة منهج البحث والتطوير بتكييف المراحل الأولى من نموذج بورغ وغال . وُظّف النموذج لتوجيه تطوير نسخة أولية وفحصها في سياق واحد. بدأ الإجراء بتحليل الحاجة إلى تقويم يضم التعليل والثقة، ثم تحديد مجالات النحو التي تغطيها الأداة وصياغة البنود رباعية المستويات، تلا ذلك تحكيم المسودة وتجريبها أوليًا وتحليل مؤشرات البنود. ويبيّن الشكل 1 المراحل التي تعرض الدراسة نتائجها. ويتصل نطاق البحث بفحص المنتج الأولي. أما التجريب الميداني الرئيس والتجريب التشغيلي والمراجعة النهائية والتعميم فتندرج ضمن مراحل تطوير لاحقة (Borg & Gall, 2003).



الصورة ١. إجراءات تطوير أداة تقويم التراكيب النحوية

أجريت الدراسة في برنامج تعليم اللغة العربية بجامعة الشيخ الحاج عبد الوهاب حسب الله في جومبانغ بإندونيسيا. وأجاب ٣٨ طالبًا عن استبانة تحليل الاحتياجات المكوّنة من ١٧ عبارة. وحكّم الأداة خبيران في النحو وتقويم تعليم اللغة العربية باستخدام استمارة تضمّ ٢٠ مؤشرًا موزّعةً على خمسة محاور. ثم شارك ٣٠ طالبًا في التجريب الأولي للأداة.

تكوّن الاختبار من ٢٥ بندًا تناولت الوظائف النحوية وعلامات الإعراب وأنواع الجمل وعلاقة العامل بالمعمول وتحليل التراكيب في سياقها. وبدأ إعداد كل بند بتحديد المؤشر واختيار جملة تتضمّن القرينة اللازمة للتحليل. ثم صيغ سؤال الإجابة في المستوى الأول وسؤال التعليل في المستوى الثالث، ولكلٍّ منهما خمسة بدائل ومفتاح واحد. وسجلت الثقة بالإجابة في المستوى الثاني وبالتعليل في المستوى الرابع باستخدام خياري "متأكد" و"غير متأكد".

لخصت استجابات الاستبانة بمتوسطات العبارات، وتقديرات الخبيرين بمتوسطات محاور التحكيم. وحُلّل أداء البنود باستخدام ارتباط بيرسون بين درجة البند والدرجة الكلية عند مستوى دلالة ٠,٠٥ في اختبار ثنائي الطرف وبـ٢٨ درجة حرية. وقُدّر الاتساق الداخلي لدرجات البنود الخمسة والعشرين بمعامل ألفا كرونباخ. وحسبت صعوبة البند بنسبة الإجابات الصحيحة. وصُنّفت البنود إلى صعبة عند نسبة تقلّ عن ٠,٣٠، ومتوسطة بين ٠,٣٠ و٠,٧٠، وسهلة عند نسبة تتجاوز ٠,٧٠ (Arikunto, 2012). وحسب التمييز بالفرق بين نسبتي الإجابة الصحيحة في المجموعتين العليا والدنيا، وفُسّر وفق تصنيف إبل وفريسبي (Ebel & Frisbie, 1991). أعطيت

الإجابة والتعليل درجة ١ للصواب و٠ للخطأ. وتُصنّف استجابة الطالب لكل بند بعد اكتمال مستوياته الأربعة وفق القواعد المكيّفة لهذه الأداة (Caleon & Subramaniam, 2010).

جدول ١. تصنيف الاستجابات المكتملة لكل بند

التصنيف	الثقة بالإجابة والتعليل	الإجابة والتعليل
فهم المفهوم	متأكد في كليهما	كلاهما صحيح
مؤشر مفهوم خاطئ	متأكد في كليهما	كلاهما خاطئ
إيجابية زائفة	متأكد في كليهما	الإجابة صحيحة والتعليل خاطئ
سلبية زائفة	متأكد في كليهما	الإجابة خاطئة والتعليل صحيح
فهم غير مؤكّد	غير متأكد في أحدهما على الأقل	كلاهما صحيح
معرفة غير كافية	غير متأكد في أحدهما على الأقل	خطأ في أحدهما على الأقل

تُحسب نسبة كل فئة من مجموع الاستجابات المكتملة للبند نفسه. واقتصر عرض نتائج التجريب الأولي في هذه الدراسة على مؤشرات أداء البنود الإجمالية.

نتائج البحث ومناقشاتها

تعليم النحو في قسم تعليم اللغة العربية وتحليل الاحتياجات

يكشف تعليم النحو في قسم تعليم اللغة العربية بجامعة الشيخ الحاج عبدوهاب حسب الله أنّ مسيرة التعليم لا تزال راسخةً في قوالب الطريقة التقليدية بصورة لافتة (Nashoih & Darmawan, 2019). ويُدرَّس النحو على مراحل متتالية تمتد من الفصل الثالث حتى الخامس، غير أنّ النهج المتبع يغلب عليه الطابع الإلقائي: يشرح الأستاذ القاعدة، ويضرب الأمثلة، ثم يُسجّل الطلبة ويحفظون . ولا يبتعد هذا النمط كثيرًا عن تقاليد التعليم في المدارس الدينية الداخلية، نظرًا لارتباط الجامعة بمحيط معهد بحر العلوم (Azizah & Nashoih, 2024). بيد أنّ هذا النهج في سياق التعليم

العالي يُبدي محدوديةً واضحة، إذ لم يُفضِ بعدُ إلى تنمية ملكة الطلبة التحليلية في بنية التراكيب النحوية بالعمق المطلوب.

إنّ رحابة مادة النحو وتعقيداتها وما تزخر به من مصطلحات متخصصة تُلقي بأعباء معرفية ثقيلة على كاهل الطلبة، لا سيما حين يخلو التدريس من المقاربة السياقية والتطبيقية. ونتيجةً لذلك، كثيرًا ما يستوعب الطلبة القواعد استيعابًا مجتزءًا، ثم يعترضهم العجز حين يُطلب منهم توظيفها في تحليل التراكيب النحوية.

جدول ٢. نتائج استبانة تحليل الاحتياجات

الناحية	عدد البنود	مجموع الدرجة	المتوسط	المعيار
فهم النحو	٦	٧٤٧	٣,٢٨	متوسط
تعليم النحو	٥	٥٩٨	٣,١٥	متوسط
تقييم تعليم النحو	٦	٨٠٤	٣,٥٣	مرتفع نسبيا
العدد الإجمالي	١٧	**٢١٤٩**	**٣,٣٣**	متوسط

وتُعزّز هذه النتائج استبانةُ تحليل الاحتياجات التي أُجريت على ٣٨ طالبًا، وهي موجزة في الجدول ٠١. ويُظهر المتوسط الكلي البالغ ٣,٣٣ أنّ واقع تعليم النحو يقع في خانة المتوسط، ممّا يُشير إلى أنّ مسيرة التعليم لم تبلغ حدّ الكمال وتحتاج إلى تقوية في جوانب رئيسية عديدة.

فعلى صعيد فهم النحو، بلغ المتوسط ٣,٢٨، ما يدلّ على أنّ الطلبة يمتلكون فهمًا معقولًا لكنّه لم يرسخ بعمق. ويتجلى ذلك في ارتفاع درجات بنود صعوبة تحليل التراكيب النحوية بدرجة ٣,٨٤ والتردد في الإجابة بدرجة ٣,٧٤، في حين كانت درجات فهم المفاهيم الجوهرية كالعامل والمعمول أدنى نسبيًا بدرجة ٢,٨٤. يُنبئ هذا النمط باختلال واضح بين الإتقان المفاهيمي والقدرة التطبيقية، مما يعني أنّ البنية المعرفية لدى الطلبة لم تتكامل وترسّخ بعد.

وعلى صعيد تعليم النحو، جاء المتوسط ٣,١٥، وهو في دائرة المتوسط. فرغم إقرار الطلبة بتنظيم المادة بدرجة ٣,٢١ وسهولتها النسبية بدرجة ٣,٠٣، ثمة ميل جليّ نحو التنظير الخالص بدرجة ٣,٦٨، كما تكشف الدرجات المنخفضة في بنود التمرين

بدرجة ٢,٩٢ والتغذية الراجعة بدرجة ٢,٨٩ أنّ التعليم لم يُتح بعدُ مساحةً كافيةً لتنمية المهارات التحليلية، فغلب على النشاط التعليمي طابع التلقّي السلبي بدلًا من البناء المعرفي الفاعل.

أما محور تقويم التعليم فقد سجّل أعلى متوسط بدرجة ٣,٥٣، غير أنّه بالمقابل يكشف عن خلل جوهري في منظومة التقويم؛ إذ تدلّ الدرجة المرتفعة على بند عجز التقويم عن استجلاء الفهم بعمق للحصول على درجة ٣,٩٢ على أنّ الأدوات المستعملة تظل في دائرة القياس السطحي. ويزيد على ذلك معاناة الطلبة في تعليل إجاباتهم بدرجة ٣,٧٤ وميلهم إلى الإجابة الصحيحة دون فهم مفاهيمي حقيقي للحصول على ٣,٨٤. وتُلمح هذه الوقائع إلى ظاهرة المفاهيم الخاطئة الكامنة، حيث لا تعكس الإجابة الصحيحة بالضرورة فهمًا مفاهيميًا راسخًا.

وتكشف نتائج تحليل الاحتياجات أنّ إشكالية تعليم النحو في قسم تعليم اللغة العربية بجامعة الشيخ الحاج عبد الوهاب حسب الله ذات طابع منظومي تتشعّب أبعادها في ثلاثة محاور: قصور الفهم المفاهيمي العميق، وشُح الممارسة التحليلية في التعليم، وغياب البُعد التشخيصي في التقويم. وهذا يُقرّر قصور التقويم التقليدي عن الإحاطة بالبنية المعرفية لدى الطلبة إحاطةً شاملة. وعليه، لا بدّ من تطوير أداة تقويم لا تقيس المخرجات النهائية فحسب، بل تُحدّد مسيرة التفكير ومستوى اليقين والمفاهيم الخاطئة بصورة أدق وأوفى. وفي هذا السياق، يغدو FTDT خيارًا موفّقًا وبديلًا واعدًا يدعم إصلاح تعليم النحو نحو مزيد من التحليل والتأمل والعمق المفاهيمي.

مواصفات أداة تقويم النحو القائمة على FTDT

ينبثق تطوير أداة تقويم النحو في هذا البحث من نتائج تحليل الاحتياجات التي أثبتت أنّ تعليم النحو في قسم تعليم اللغة العربية بجامعة الشيخ الحاج عبج الوهاب حسب الله يرزح تحت إشكالية محورية تتجلّى في العجز عن تحديد أشكال المفاهيم الخاطئة لدى الطلبة بدقّة. فالتقويم المعتمد حاليًا ذو طابع تقليدي يقتصر على قياس المخرجات النهائية (صواب/خطأ)، دون أن يُعنى بالكشف عن مسيرة التفكير ودرجة اليقين والمسوّغات المفاهيمية الكامنة وراء إجابات الطلبة. من هنا تبرز الحاجة الملحّة

إلى أداة تقويم لا تؤدّي وظيفة ختامية فحسب، بل تأخذ بزمام الوظيفة التشخيصية أيضًا.

وقد صُمّمت هذه الأداة وفق مقاربة FTDT، وهو نموذج تقويمي تشخيصي متعدد الطبقات يتألف من أربعة مستويات متتالية. وقد وجد هذا النموذج قبولًا واسعا في تحديد المفاهيم الخاطئة، لقدرته على التمييز بين الجهل وسوء الفهم والاعتقاد الخاطئ بصواب مفهوم ما (Gurel et al., 2015; Treagust, 2006). وفي سياق تعليم النحو تحديدًا، تتجلى أهمية هذه المقاربة لأنّ أخطاء الطلبة لا تكمن فحسب في النتيجة النهائية لتحليل الإعراب، بل كثيرًا ما تتعمّق في طريقة فهم المفاهيم الأساسية كالعلاقات التركيبية في الجملة العربية.

وتتكون الأداة في مجملها من ١٠٠ بند موزّعة على أربعة مستويات: المستوى الأول يضمّ ٢٥ سؤالًا لقياس الفهم المفاهيمي الأساسي في مادة النحو، وهي أسئلة اختيار من متعدد بخمسة بدائل، تُقدّم إلى جانب الإجابة الصحيحة مشتّتات (distractors)تُجسّد المفاهيم الخاطئة المحتملة. والمستوى الثاني من ٢٥ بندًا تقيس ثقة الطالب بإجابته في المستوى الأول، حيث يختار بين "متأكد" أو "غير متأكد"، فيُعرَف ما إذا كانت إجابته مبنيةً على فهم متين أم مجرد تخمين.

والمستوى الثالث من ٢٥ سؤالًا يستقصي المسوّغات الكامنة وراء إجابة الطالب في المستوى الأول، وهي أيضًا اختيار من متعدد بخمسة بدائل تعبر عن تنويعات الاستدلال المفاهيمي: الصحيح منه وذلك الحامل للمفاهيم الخاطئة. وبذلك يُصبح المستوى الثالث المحطةَ الأكثر أهمية في الكشف عن عمق فهم الطالب للنحو. أما المستوى الرابع فيضمّ ٢٥ سؤالًا تقيس ثقة الطالب بالمسوّغ الذي اختاره في المستوى الثالث، ويتكوّن من خيارَين فحسب: "متأكد" و"غير متأكد". ويُمكّن الجمع بين هذه المستويات الأربعة من تصنيف فهم الطالب في فئات: فاهم للمفهوم، وغير فاهم للمفهوم، ومُبتلى بمفهوم خاطئ، ومُبتلى بمفهوم خاطئ مع درجة عالية من اليقين.



اختبار تشخيصي رباعي المستويات في التراكيب النحوية

لطلبة قسم تعليم اللغة العربية بجامعة كياهي الحاج عبد الوهاب حسب الله في جومبانغ

اقرأ الجملة ثم اختر بديلاً واحداً في كل مستوى، واكتب حرفه في خانة الاختيار.

مقياس الثقة للمستويين ٢ و٤	أ. واثق جداً	ب. واثق	ج. غير واثق	د. غير واثق جداً

السؤال ١	ذهب الطالب إلى المكتبة	
المستوى	المطلوب والبدائل	الاختيار
١.١ الإجابة	**ما إعراب كلمة «الطالب»؟** أ. مفعول به منصوب ب. فاعل مرفوع ج. مبتدأ مرفوع د. اسم مجرور هـ. نائب فاعل مرفوع	☐
١.٢ الثقة بالإجابة	ما مدى ثقتك بإجابتك؟	☐
١.٣ التعليل	**لماذا اخترت هذه الإجابة؟** أ. لأن كل اسم يأتي بعد الفعل مباشرةً يكون فاعلاً. ب. لأن الطالب هو الذي قام بفعل الذهاب. ج. لأن الكلمة مسبوقة بحرف جر. د. لأن الجملة اسمية. هـ. لأن الاسم بعد الفعل المبني للمعلوم يُعرب نائب فاعل.	☐
١.٤ الثقة بالتعليل	ما مدى ثقتك بالتعليل الذي اخترته؟	☐

الصورة ٢. نموذج السؤال للتراكيب النحوية القائمة على FTDT.

وقد وزِّعت الأسئلة توزيعًا متناسبًا ليشمل جميع هذه المؤشرات وتقيس الأداة فهم الطالب قياسًا شاملًا. أما مادة النحو المُختبَرة فتشمل: مفهوم الإعراب، والوظائف التركيبية للكلمة في الجملة، وأنواع الجمل، ومفهوم العامل والمعمول، وتحليل التراكيب النحوية في صور متنوعة من الجملة العربية.

ومن حيث المستوى المعرفي، تقع أسئلة المستوى الأول في الغالب عند درجة C2 (الفهم) في تصنيف بلوم المُراجَع، لكنّ طابعها التحليلي يجعل بعضها يستلزم كذلك مهارات تحليل أولية. وتُصمَّم أسئلة المستوى الثالث على درجة C4 (التحليل)، لأنّ الطالب لا يكتفي بتحديد الإجابة بل يعلل الأسباب المفاهيمية التي تقف وراءها. وفي بعض البنود التي تستدعي إعادة بناء الجملة أو تغيير علاماتها الإعرابية، يُوظَّف مستوى C5 (التقييم أو الإنتاج) الذي يستدعي مهارات التفكير العليا.

وتتمحور خصيصة هذه الأداة حول صياغة أسئلتها وفق البنية الاستنطاقية لأنماط النص العربي؛ إذ تصاغ كل بند في صورة فقرة أو جملة مطولة ومعقدة تستدعي من الطالب قراءة دقيقة وتعريفا للبنية التركيبية وتحديدا للعلاقات بين عناصر الجملة، وهو ما يميزها جوهريا عن التقويم التقليدي الذي يركز على حفظ القواعد. فتُعلي الأداة من شأن المهارة التطبيقية والتحليلية للطالب في فهم النحو فهمًا سياقيًا.

وقد صممت المشتتات في كل سؤال بصورة منهجية مستندةً إلى الأخطاء الشائعة التي يُبديها الطلبة في فهم النحو، إذ لم تُختَر عشوائيا بل انبثقت من نتائج تحليل الاحتياجات والمعطيات التجريبية الميدانية، فأصبحت قادرةً على تمثيل أنماط المفاهيم الخاطئة الفعلية. وهذا يتّسق مع مبادئ تطوير الاختبارات التشخيصية التي تُنبّه إلى أهمية المشتّتات المبنية على المفاهيم الخاطئة (Haladyna et al., 2002).

ويرتكز تطوير هذه الأداة على مبدأ أنّ التقويم يجب أن يرسم صورةً متكاملةً عن مسيرة التعلم ونتائجه.(Astuti et al., 2023; Kiray & Simsek, 2021) وفي سياق تعليم النحو ذي الطابع المعقّد والمجرّد، يكتسب النهج التشخيصي أهمية قصوى للتثبّت من أنّ الطالب لا يكتفي بالإجابة الصحيحة بل يحمل فهمًا مفاهيميًا عميقًا ومحصّنًا من المفاهيم الخاطئة الراسخة (Pulu & Amahoru, 2023).

إن الاختبارات التشخيصية المتعددة الطبقات توفّر فرصةً أوسع لاستجلاء البنية المعرفية للمتعلم مقارنةً بالاختبارات التقليدية.(Boori et al., 2024) ويُعزز ذلك ما أثبتته الدراسات من أنّ FTDT قادر على تحديد المفاهيم الخاطئة بدقة أكبر لتضمينه أبعاد الإجابة والمسوّغ والثقة.(Putri et al., 2021) ولذا يُؤمَل أن يُشكّل توظيف هذا النموذج في تقويم تعليم النحو ابتكارًا جوهريًا يرفع سوية التعلم، خاصةً في مواجهة إشكالية المفاهيم الخاطئة التي طالما اعترضت مسيرة الطلبة.

صلاحية أداة تقويم النحو القائمة على اختبار التشخيص رباعي المستويات (FTDT)

أُجري اختبار صدق الأداة عبر التحكيم الخبراتي بمشاركة خبيرَين متخصصَين في النحو وتقويم تعليم اللغة العربية. وقد استُخدمت في التقييم استبانةٌ للتحقق من صدق المادة وُضعت بصورة منهجية، شملت خمسة محاور رئيسية: ملاءمة المادة للمنهج، ودقة المفاهيم النحوية وصوابها، وعمق المادة وشمولها، وملاءمة الأداة لخصائص اختبار التشخيص رباعي المستويات، وارتباطها بالأهداف التشخيصية. واختيار هذه المحاور الخمسة استند إلى كونها تُمثّل أبعادًا جوهريةً في تطوير أداة التقويم التشخيصي عالية الجودة.

جدول ٣. نتائج التحقق من صدق الأداة

محور التقييم	عدد البنود	مجموع الدرجات	المتوسط	الفئة
ملاءمة المادة للمنهج	٤	٣٥	4.38	صادق جدًا
دقة المفاهيم النحوية وصوابها	٤	٣٤	٤,٢٥	صادق جدًا
عمق المادة وشمولها	٤	٣٥	٤,٣٨	صادق جدًا
ملاءمة الأداة لخصائص FTDT	٤	٣٢	٤,٠٠	صادق
الارتباط بالأهداف التشخيصية	٤	٣٨	٤,٧٥	صادق جدًا
العدد الإجمالي	٢٠	١٧٥	٤,٣٨	صادق جدًا

بيّنت نتائج تحكيم الخبيرَين المعروضة في الجدول ٢ أنّ أداة تقويم النحو القائمة على اختبار التشخيص رباعي المستويات حصلت على متوسط إجمالي بلغ ٤,٣٨، وهو ضمن فئة "صادق جدًا"، ممّا يدل على أنّ الأداة استوفت معايير صدق المضمون والبنية وتصلح لقياس فهم مفاهيم النحو تشخيصيًا (Javidanmehr & Sarab, 2017).

وبالتحديد، نال محور الارتباط بالأهداف التشخيصية أعلى درجة(M = 4.75) ، ممّا يؤكد قدرة الأداة على تحديد المفاهيم الخاطئة والتمييز بين مستويات الفهم وكشف المسوّغات الكامنة وراء الإجابات. وتتّسق هذه الخصائص مع مزايا اختبار التشخيص رباعي المستويات المُصمَّم لاستكشاف البنية المعرفية بعمق يفوق ما تبلغه الاختبارات التقليدية (Gurel et al., 2015).

وسجّل محورا ملاءمة المادة للمنهج وعمق المادة متوسطًا قدره ٤,٣٨ لكل منهما، ما يُفيد بأنّ البنود متسقة مع مخرجات التعلم وتغطي تنوّعًا في المستويات المعرفية. وأحرز محور دقة المفاهيم النحوية ٤,٢٥، ويعني ذلك أنّ الأداة صِيغت بالضبط النحوي الصحيح والمصطلحات المناسبة، وهو ضروري للحفاظ على صدق البنية وضمان خلوّ القياس من التحيّز المفاهيمي (Haladyna et al., 2002).

وجاء محور ملاءمة الأداة لخصائص FTDT بأدنى درجة (M=4,00)، وإن بقي ضمن فئة "صادق". وتدلّ هذه النتيجة على أنّه رغم الصدق العام، ثمة حاجة إلى تحسين صياغة بعض البنود وتطوير المشتّتات لتكون أكثر تمثيلًا للمفاهيم الخاطئة. ولا تخفى أهمية جودة المشتّتات في أداة التقويم التشخيصي؛ إذ تُمثّل مفتاح الكشف عن أنماط الخلل المفاهيمي (Gierl et al., 2017).

وخلاصة القول، تُثبت هذه النتائج أنّ أداة اختبار التشخيص رباعي المستويات المطوَّرة ذات صدق متين وحساسية عالية لرصد التباين في فهم الطلبة. وعليه، لا تصلح الأداة للاستخدام التقويمي فحسب، بل تُعدّ كذلك وسيلةً تشخيصيةً فعّالة تُعين على تحسين استراتيجيات تدريس النحو بصورة أكثر توجيهًا (Caleon & Subramaniam, 2010).

الصدق والثبات لأداة تقويم النحو القائمة على FTDT

يُمثّل الصدق والثبات ركيزتَين أساسيتَين في تطوير أدوات التقويم التربوي. فالصدق يُحدد مدى قياس الأداة للبنية التي تستهدفها بالفعل، بينما يرتبط الثبات بمستوى اتساق نتائج القياس عند استخدام الأداة في ظروف متشابهة نسبيًا (Laila & Nashoih, 2024). وفي سياق أداة التقويم القائمة على اختبار التشخيص رباعي المستويات، لا يقتصر الصدق على الصدق المضموني. بل يشمل أيضًا صدق البنية والصدق التجريبي الموثّق بالتحليل الإحصائي للبنود (Prabowo et al., 2023).

وبعد استكمال مرحلتَي تحليل الاحتياجات وتحكيم الخبراء، جُرّبت الأداة على عيّنة قوامها ٣٠ طالبًا كعيّنة تجريب تجريبية للتحقق من الصدق والثبات، وذلك للتأكد من أنّ كل بند يتمتع بجودة قياسية كافية قبل استخدامه على نطاق أوسع.

وتُعرض نتائج اختبار الصدق التجريبي لـ٢٥ بندًا في المستوى الأول (Tier 1) والمستوى الثالث (Tier 3) في الجدول ٤.

جدول ٤. نتائج اختبار الصدق استنادًا إلى معامل ارتباط بيرسون

Statistic	Value
Number of items	25
rmin	0.472
rmax	0.622
rtable	0.361
Significance level	< .05
Decision	All items valid

يُبيّن الجدول ٣ أنّ جميع البنود حصلت على معاملات ارتباط موجبة (r-count) تفوق r-table بدرجة ٠,٣٦١ عند مستوى دلالة ٥% بدرجات الحرية 30 = (df). كما أنّ جميع البنود بلغت مستوى الدلالة أقلّ من ٠,٠٥، ممّا يكشف عن علاقة دالة بين درجة البند والدرجة الكلية.(Maison et al., 2021) وبذلك يُعلَن أنّ جميع بنود هذه الأداة صادقة تجريبيًا.

وتدلّ هذه النتيجة على أنّ كل بند يُسهم إسهامًا ذا صلة في قياس البنية المنشودة، أي فهم مفاهيم النحو والقدرة على تحليل تراكيب الجملة العربية. ومن منظور التقويم التشخيصي، يكتسب هذا الصدق أهمية بالغة، لأنّ الأداة لا تقيس تحصيل نتائج التعلم فحسب، بل تُحدّد كذلك المفاهيم الخاطئة لدى الطلبة بعمق أكبر.(Irawan et al., 2012) وللتحقق من جاهزية جميع البيانات لاختبار الثبات، أُجري تحليل ملخّص معالجة البيانات (case processing summary) كما في الجدول ٥.

جدول ٥. ملخص عدد بيانات المستجيبين

Category	N	%
Valid	30	100%
Excluded	0	0%

يتضح من الجدول ٤ أنّ جميع بيانات المستجيبين قابلة للتحليل دون استبعاد أي منها، ممّا يعني أنّ البيانات مكتملة وتستوفي شروط التحليل المتقدم. وقد أُجري اختبار الثبات باستخدام معامل ألفا كرونباخ للكشف عن الاتساق الداخلي بين البنود، وهي موضّحة في الجدول ٦.

جدول ٦. قيم إحصاءات ثبات أداة تقويم النحو القائمة على FTDT

Cronbach's Alpha	Number of Items
0.884	25

بلغ معامل ألفا كرونباخ ٠,٨٨٤ لـ٢٥ بندًا وفق الجدول ٥، وهو يتجاوز الحدّ الأدنى المقبول في البحوث التربوية (٠,٧٠)، مما يُصنّفه في خانة الثبات العالي جدًا. ويُثبت ذلك أنّ أداة تقويم النحو القائمة على اختبار التشخيص رباعي المستويات تتمتع باتساق داخلي ممتاز.

ويكشف هذا المستوى المرتفع من الثبات أنّ كل بند يقيس البنية ذاتها بصورة متسقة. فالطلبة ذوو الفهم الراسخ للنحو يميلون إلى الإجابة الصحيحة باتساق، بينما يُظهر أصحاب المفاهيم الخاطئة أنماط خطأ مستقرة نسبيًا. وهذا الاتساق بالغ الأهمية في التقويم التشخيصي إذ يُمكّن من تحديد أنماط المفاهيم الخاطئة بصورة منهجية ودقيقة.(Gierl et al., 2017)

وفي سياق اختبار التشخيص رباعي المستويات تحديدًا، لا يعكس الثبات الاتساقَ في إجابات المستوى الأول (Tier 1) فحسب، بل يشمل أيضًا الترابط المنطقي بين الإجابات والمسوّغات (Tier 3) ودرجات يقين الطلبة Tier 2) و .(4 Tierوهذا ما يمنح أداة اختبار التشخيص رباعي المستويات ميزةً تفوق بها الأدوات التقليدية، لكونها تُقدّم صورةً أشمل للبنية المعرفية بما فيها مناطق الشكّ والمفاهيم الخاطئة المستترة .(Kirbulut & Geban, 2014)

وبدمج نتائج صدق المضمون الخبراتي والصدق التجريبي وارتفاع معامل الثبات، يخلص البحث إلى أنّ أداة تقويم النحو القائمة على اختبار التشخيص رباعي المستويات تستوفي معايير كونها أداةً صادقة وثابتة. فهي قادرة على قياس مستوى الإتقان النحوي بدقة، وفعّالة في تحديد المفاهيم الخاطئة بعمق.

وتُفيد هذه النتائج بأن توظيف أداة التقييم في تعليم النحو يُمكن أن يمثل بديلا تقويميا ذا طابع تشخيصي ومعنى عميق. فمن خلال أداة اختُبر صدقها وثباتها، يستطيع الأستاذ الحصول على صورة أكثر شمولا عن واقع فهم الطلبة، وصياغة

استراتيجيات تعليمية أكثر دقة وفاعلية، لا سيما في التصدّي للصعوبات والمفاهيم الخاطئة التي طالما شكّلت العقبة الكبرى في تعليم النحو في قسم تعليم اللغة العربية بجامعة الشيخ الحاج عبد الوهاب حسب الله.

مستوى الصعوبة والقدرة التمييزية

أُجري تحليل جودة بنود أداة تقويم النحو القائمة على الاختبار التشخيصي رباعي المستويات من خلال اختبار مستوى الصعوبة والقدرة التمييزية. ويُعدّ هذان المعطيان مؤشّرَين جوهريَين في الحكم على جدارة الأداة، لأنّهما يعكسان ليس صعوبة السؤال وحسب، بل أيضًا قدرته على التمييز بين مستويات إتقان المفاهيم لدى الطلبة.

جدول ٧. توزيع مؤشر مستوى الصعوبة

الفئة	نطاق P	عدد البنود	النسبة (%)
سهل	$P > 0.70$	٤	%١٦
متوسط	$0.30 \leq P \leq 0.70$	١٤	%٥٦
صعب	$P < 0.30$	٧	%٢٨
الإجمالي	—	٢٥	%١٠٠

وفقًا لنتائج تحليل مستوى الصعوبة، هيمنت الفئة المتوسطة على توزيع البنود بواقع ١٤ بندًا من ٢٥ أي ما نسبته ٥٦%. وجاءت ٧ بنود بنسبة ٢٨% في فئة الصعب، و٤ بنود بنسبة ١٦% في فئة السهل. ويُشير هذا التوزيع إلى توازن نسبي معقول بين درجات الصعوبة، وإن كان التوجه نحو الفئة المتوسطة واضحًا.

ويُعدّ هيمنة الأسئلة المتوسطة الصعوبة سمةً مرغوبة في أدوات التقويم التربوي، إذ تُعدّ أكثر فعالية في قياس قدرات الطلبة. والأسئلة السهلة جدًا تفشل في كشف التباين في القدرات، في حين قد تُفضي الصعبة جدًا إلى الإحباط وتُشوّه المعلومات حول مستوى الإتقان. بيد أنّ وجود أسئلة صعبة بنسبة معقولة (٢٨%) يُعدّ ميزةً في سياق التقويم التشخيصي لقدرته على كشف المفاهيم الخاطئة الأكثر عمقًا، ولا سيما في المفاهيم النحوية الأكثر تعقيدًا.

جدول ٨. توزيع مؤشر القدرة التمييزية

الفئة	نطاق D	عدد البنود	النسبة (%)
ممتاز	$D \geq 0.50$	٥	%٢٠
جيد	$0.40 \leq D < 0.50$	١٢	%٤٨
مقبول	$0.20 \leq D < 0.40$	٨	%٣٢
ضعيف	$D < 0.20$	٠	%٠
الإجمالي	-	٢٥	%١٠٠

وتكشف نتائج تحليل القدرة التمييزية أنّ معظم البنود تتمتع بجودة تمييزية جيدة؛ إذ إنّ ١٢ بندًا بنسبة ٤٨% تقع في فئة الجيد، و٥ بنود بنسبة ٢٠% في فئة الممتاز، و٨ بنود بنسبة ٣٢% في فئة المقبول، ولم يُعثر على أيّ بند ذي قدرة تمييزية ضعيفة. وتدلّ هذه النتائج على أنّ الأداة تمتلك قدرة كافية للتمييز بين الطلبة ذوي مستوى الإتقان المرتفع وأولئك المتدنّيه.

بلغت نسبة البنود المصنفة ضمن فئتَي التمييز الجيد والممتاز ٦٨%، مما يشير إلى أن معظم البنود تميز بين مستويات أداء الطلبة بصورة ملائمة. أما البنود ذات التمييز المقبول، فيمكن الاحتفاظ بها بعد مراجعتها في ضوء أهداف الاختبار التشخيصي رباعي المستويات، الذي يهتم باستكشاف أنماط التفكير والمفاهيم الخاطئة.

ويشير تكامل نتائج تحليلي الصعوبة والتمييز إلى مؤشرات إيجابية لجودة الأداة. فغلبة البنود متوسطة الصعوبة، إلى جانب ارتفاع نسبة البنود ذات التمييز الجيد والممتاز، تدعم تقويم مستويات متفاوتة من إتقان المفاهيم النحوية، والتمييز بين مستويات أداء الطلبة.

وتتضح القيمة التشخيصية عند ربط هذه المؤشرات بتحليل الإجابات ومبرراتها ودرجات الثقة بهما، بدلًا من الاكتفاء بالحكم على صحة الإجابة أو خطئها وحدها. فالبنود الصعبة ذات التمييز الملائم قد تسهم في تحديد مواضع التعثر، بينما يتيح

تحليل الاستجابات عبر المستويات الأربعة فهمًا أعمق للبنية المفاهيمية، وتوجيه التدخلات التعليمية لمعالجة المفاهيم الخاطئة.

تحليل المفاهيم الخاطئة لدى الطلبة استنادًا إلى اختبار التشخيص

يستعين تحليل المفاهيم الخاطئة في هذا البحث بمقاربة اختبار التشخيص رباعي المستويات التي تُدمج أربعة مكوّنات رئيسية: الإجابة (Tier 1)، ومستوى الثقة بالإجابة (Tier 2)، والمسوّغ (Tier 3) ، ومستوى الثقة بالمسوّغ .(Tier 4) وبما أنّ البيانات المجمّعة ذات طابع إجمالي، فقد اعتُمد في تصنيف الاستجابات أسلوب التوزيع النسبي المستند إلى توزيع التكرارات في كل مستوى. ويُمكّن هذا الأسلوب من تحديد أنماط الفهم تشخيصيًا دون الحاجة إلى تتبّع الأفراد بصورة منفردة، وهو ما توصي به دراسات اختبار التشخيص رباعي المستويات.

جدول ٩. توزيع تصنيفات مستوى فهم الطلبة استنادًا إلى FTDT

الفئة	عدد البنود	النسبة (%)	المعيار
فاهم للمفهوم	١٣	%٥٢	فهم مفاهيمي متين
مفهوم جزئي	٣	%١٢	فهم غير مستقر
غير عارف للمفهوم	٥	%٢٠	ضعف في الإتقان المفاهيمي
مفهوم خاطئ	٤	%١٦	خطأ مفاهيمي مؤكَّد
العدد الإجمالي	٢٥	%١٠٠	-

تُظهر نتائج التحليل أنّ فئة "فاهم للمفهوم" سيطرت على معظم البنود، لا سيما ذات الصعوبة المنخفضة إلى المتوسطة والقدرة التمييزية الجيدة. ويُنبئ هذا النمط بأنّ الطلبة لم يكتفوا بالإجابة الصحيحة، بل قدّموا مسوّغات وافيةً مع مستوى ثقة مناسب، ممّا يُرسّخ صدق البنية المفاهيمية للأداة.

غير أنّ النتيجة الأبرز في هذا البحث هي الحضور اللافت لفئة المفهوم الخاطئ. لا سيما في البنود ذات الصعوبة المرتفعة. وتحديدًا، أبدت البنودان رقم ١٠ و٢٠ نمطَ "مفهوم خاطئ راسخ" يتجلى في تدنّي نسبة الإجابات الصحيحة مع ارتفاع درجة ثقة الطلبة بأنفسهم. ويُشير هذا النمط "خاطئ لكن واثق" إلى بناء معرفي مُغلوط ومستقر،

وهو وفق الأدب التربوي أشدّ أنواع المفاهيم الخاطئة مقاومةً للتغيير (Jannah et al., 2022; Rohmanasari & Ermawati, 2020). وبذلك لا يكون خطأ الطالب ناجمًا عن الجهل فحسب، بل عن فهم خاطئ بناؤه المعرفي متماسك.

كذلك رُصدت فئة المفهوم الجزئي. في بعض البنود كـ١٥ و١٩ و٢٣، وتتسم بصحة الإجابة من دون مسوّغ سليم أو العكس. وتُلمح هذه الظاهرة إلى ما يُسمى "الفهم الزائف" (false positive understanding)، إذ يبدو الطالب مُصيبًا في الظاهر لكنه يفتقر في الجوهر إلى الفهم المفاهيمي الحقيقي. وقد كشفت دراسات سابقة أنّ هذه الحالة تنبثق في الغالب من التعلم القائم على الحفظ وتنطوي على مخاطر عالية لأخطاء نقل المفاهيم (Safitri et al., 2024).

وعلى صعيد آخر، رُصدت فئة "غير عارف للمفهوم"، وتتسم بالخطأ وانخفاض مستوى الثقة. وخلافًا للمفهوم الخاطئ، تكشف هذه الفئة عن وعي الطالب بمحدودية معرفته، وهو من الناحية التربوية أيسر علاجًا بالتدخل التعليمي الملائم. وتكشف نتائج تحليل اختبار التشخيص رباعي المستويات أنّ فهم الطلبة للتراكيب النحوية يمتد على طيف واسع يشمل: الفهم المفاهيمي، والمفهوم الخاطئ الراسخ، والمفهوم الجزئي، وانعدام معرفة المفهوم. وتُؤكد هذه النتائج أنّ التقويم القائم على اختبار التشخيص رباعي المستويات لا يقتصر على تحديد الصواب والخطأ، بل يكشف البنية المعرفية الكامنة بما فيها بُعد الثقة الذي كثيرًا ما يُهمَل في التقويم التقليدي.

الخلاصة

أسفر هذا البحث عن تطوير أداة تقويم تشخيصية قائمة على اختبار التشخيص رباعي المستويات للكشف عن المفاهيم الخاطئة لدى طلبة برنامج تعليم اللغة العربية في مادة التراكيب النحوية. وأظهرت نتائج تحكيم الخبراء والتجريب مؤشرات إيجابية لجودة الأداة، إذ بلغ متوسط تقديرات الخبراء ($M = 4.38$)، وبلغ معامل الاتساق الداخلي ($\alpha = 0.884$). كما أظهرت الأداة توزيعًا متوازنًا لمستويات الصعوبة وقدرةً تمييزيةً ملائمة، مما يدعم استخدامها في السياق المدروس.

وكشفت النتائج التشخيصية عن تفاوت في الفهم المفاهيمي لدى الطلبة، شمل الفهم الصحيح والمفاهيم الخاطئة، ومنها إجابات خاطئة مصحوبة بثقة عالية. كما ظهرت حالات من الفهم الإيجابي الزائف، حيث لا تقترن صحة الإجابة بسلامة التفسير المفاهيمي. وبذلك تشير النتائج إلى أن الأداة المطوَّرة حقّقت، في حدود السياق المدروس، الهدف التشخيصي المحدَّد في المقدّمة، وهو الكشف عن المفاهيم الخاطئة لدى الطلبة في التراكيب النحوية. فقد أتاح تحليل الإجابة ومبرّراتها ودرجة الثقة بكلٍّ منهما التمييز بين أنماط الفهم التي قد يحجبها الاقتصار على الحكم على صحة الإجابة.

وتدعم هذه النتائج توظيف التقويم التشخيصي ضمن الممارسات التكوينية لتحديد مواضع التعثّر وتوجيه التدخّلات التعليمية نحو معالجة المفاهيم الخاطئة وتعزيز الفهم الصحيح. وفي ضوء حدود العينة والسياق المؤسسي للدراسة، يُوصى بتوسيع التحقّق من خصائص الأداة في عينات وسياقات متنوّعة، واختبار فاعلية التدخّلات العلاجية المستندة إلى نتائجها في تحسين الفهم المفاهيمي للتراكيب النحوية.

الشكروالتنويه

يتقدّم الباحثان بالشكر إلى جامعة كياهي الحاج عبد الوهاب حسب الله في جومبانغ على إتاحة إجراء هذه الدراسة فيها. كما يشكران الخبراء المحكمين في الجامعة على تحكيم أداة التقويم القائمة على اختبار التشخيص رباعي المستويات وتقديم المراجعة والملاحظة لبنودها، والطلبة في قسم تعليم اللغة العربية بالجامعة على مشاركتهم في هذه الدراسة

إفادة مساهمات الباحث

تولّى AN تنفيذ جميع مراحل البحث، بدءًا من صياغة عنوان الدراسة ، مرورًا بتطوير أداة التقويم القائمة على اختبار التشخيص رباعي المستويات، وصولًا إلى كتابة المقال. وأسهمت HM في إرشاد الباحث الأول خلال الإعداد للبحث وكتابة المقال،

وشمل توجيهها صياغة خلفية الدراسة، ووصف منهجيتها، وعرض النتائج ومناقشتها، وكتابة الخلاصة.

بيان كتابة الذكاء الاصطناعي

يُفصح الباحثان عن استعانتهما بأداة ChatGPT (OpenAI) في ترجمة عنوان البحث إلى اللغة الإنجليزية وتحرير الملخّص الإنجليزي. وقد راجعا الصياغة الإنجليزية للعنوان والملخّص ونقّحاها، ويتحمّلان المسؤولية الكاملة عن المحتوى النهائي للمخطوطة.

المراجع

Abdurrohman, M. A., & Ulinnuhaa, M. (2024). Problematika Pembelajaran Nahwu. *AL-MIKRAJ Jurnal Studi Islam Dan Humaniora (E-ISSN 2745-4584)*, *4*(02), 1768–1772. https://doi.org/10.37680/almikraj.v4i02.5350

Anggara, S. A. (2024). Multicultural Approach to Arabic Language Learning in the Special Program for Arabic Language Development. *Proceedings of International Conference on Education, Society, and Management*, *1*(1), Article 1.

Arikunto, S. (2012). *Dasar-dasar evaluasi pendidikan* (2nd edn). Bumi Aksara.

Asiah, A., Zamroni, Z., & Rijal, M. K. (2022). Problematika Pembelajaran Nahwu Dalam Meningkatkan Keterampilan Berbahasa Arab di Lembaga Pendidikan Indonesia. *Borneo Journal of Language and Education*, *2*(2), 170–185. https://doi.org/10.21093/benjole.v2i2.6104

Astuti, I. A. D., Bhakti, Y. B., Prasetya, R., & Rahmawati, Y. (2023). Four Tier-Relativity Diagnostic Test (4T-RDT) to Identify Student Misconception. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, *8*(1), 75. https://doi.org/10.26737/jipf.v8i1.3668

Azizah, F. N., & Nashoih, A. K. (2024). Development of Functional Arabic Grammar (Nahwu) Textbooks Based on Local Wisdom for Arabic Language Students. *Arabiyatuna: Jurnal Bahasa Arab*, *8*(2), Article 2. https://doi.org/10.29240/jba.v8i2.10252

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

Borg, Welter. R., & Gall, Meredith. D. (2003). *Educational Research: An Introduction* (4th edn). Longman Inc.

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

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

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th edn). Prentice Hall.

Faridl, M. F. D., Faisol, F. H., & Nashoih, A. K. (2026). Development of a four-tier gamified diagnostic test to identify students' misconceptions in Ejaan Yang Disempurnakan (Eyd) V. *GHANCARAN: Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, *8*(1), 379–407.

Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, Analyzing, and Using Distractors for Multiple-Choice Tests in Education: A Comprehensive Review. *Review of Educational Research*, *87*(6), 1082–1116. https://doi.org/10.3102/0034654317726529

Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A Review and Comparison of Diagnostic Instruments to Identify Students' Misconceptions in Science. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(5), 989–1008. https://doi.org/10.12973/eurasia.2015.1369a

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*, *15*(3), 309–333. https://doi.org/10.1207/S15324818AME1503_5

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

Hidayah, N. A. F., Priatmoko, S., & Wardani, S. (2022). *Pengembangan Tes Four Tier Multiple Choice (4TMC) untuk Mendeteksi Miskonsepsi Peserta Didik*.

Irawan, E., Riyadi, R., & Triyanto, T. (2012). Analisis Miskonsepsi Mahasiswa Stkip Pgri Pacitan Pada Mata Kuliah Pengantar Dasar Matematika

Pokok Bahasan Logika Ditinjau Dari Gaya Kognitif Mahasiswa. *Journal of Mathematics and Mathematics Education*, *2*(1), Article 1. https://doi.org/10.20961/jmme.v2i1.9942

Ismail, Z., Rasit, N., & Supriyatno, T. (2023). Relationship Between Oral Language Anxiety and Students' Arabic Language Learning Outcomes in Malaysian Secondary Schools. *International Journal of Language Education*, *1*(1), 143–161. https://doi.org/10.26858/ijole.v1i1.37368

Istiyono, E., Dwandaru, W. S. B., Fenditasari, K., Ayub, M. R. S. S. N., & Saepuzaman, D. (2023). The Development of a Four-Tier Diagnostic Test Based on Modern Test Theory in Physics Education. *European Journal of Educational Research*, *volume–12–2023*(volume–12–issue–1–january–2023), 371–385. https://doi.org/10.12973/eu-jer.12.1.371

Jannah, M., Erniwati, E., Mongkito, V. H. R., & Yuris, M. (2022). Analisis Miskonsepsi Siswa Menggunakan Four Tier Diagnostic Test pada Pokok Bahasan Fluida Statis Siswa SMKN 2 Kendari. *Jurnal Penelitian Pendidikan Fisika*, *7*(1), Article 1. https://doi.org/10.36709/jipfi.v7i1.23084

Javidanmehr, Z., & Sarab, M. R. A. (2017). *Cognitive Diagnostic Assessment: Issues and Considerations*. *7*(2).

Kiray, S. A., & Simsek, S. (2021). Determination and Evaluation of the Science Teacher Candidates' Misconceptions About Density by Using Four-Tier Diagnostic Test. *International Journal of Science and Mathematics Education*, *19*(5), 935–955. https://doi.org/10.1007/s10763-020-10087-5

Kirbulut, Z. D., & Geban, O. (2014). Using Three-Tier Diagnostic Test to Assess Students' Misconceptions of States of Matter. *Eurasia Journal of Mathematics, Science and Technology Education*, *10*(5), 509–521. https://doi.org/10.12973/eurasia.2014.1128a

Laila, L. M., & Nashoih, A. K. (2024). Tatwir Adawat Taqwim Ta'lim al-Balagah 'abra Minassah Wordwall fi Dhaui Madkhal al-Ikhtibar at-Tasykhisyi li Arba'ati Mustawayat. *LISANIA: Journal of Arabic Education and Literature*, *8*(2), Article 2. https://doi.org/10.18326/lisania.v8i2.227-256

Luthfi, K. M. (2018). Epistemologi Nahw Ta'Limi Dalam Perspektif Linguis Arab Kontemporer. *Arabiyat : Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, *5*(2), Article 2. https://doi.org/10.15408/a.v5i2.7959

Maison, M., Kurniawan, D. A., & Widowati, R. S. (2021). The Quality of Four-Tier Diagnostic Test Misconception Instrument for Parabolic Motion. *Jurnal Pendidikan Dan Pengajaran*, *54*(2), 359–369. https://doi.org/10.23887/jpp.v54i2.35261

Maula, S. Z. H., Ahsanuddin, M., Maziyah, L., Zawawi, M., & Ula, M. B. (2026). Development and Preliminary Validation of a Two-Tier Diagnostic Test for Identifying Misconceptions in Arabic Grammar Learning. *International Journal of Arabic Language Teaching*, *8*(01), 178–198. https://doi.org/10.32332/ijalt.v8i01.11810

Murad, T. M., & Khalil, M. H. (2015). Analysis of Errors in English Writings Committed by Arab First-year College Students of EFL in Israel. *Journal of Language Teaching and Research*, *6*(3), 475. https://doi.org/10.17507/jltr.0603.02

Nashoih, A. K. (2019). *Nahwu Kontrastif: Mengkaji Sintaksis Bahasa Arab Melalui Perbandingannya dengan Bahasa Indonesia*. Erhaka Utama.

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

Nashoih, A. K., Mathoriyah, L., Qosim, N., & Sholahuddin, M. F. (2023). Analisis Kesalahan Nahwu dalam Penulisan Abstrak Skripsi Mahasiswa Program Studi Pendidikan Bahasa Arab. *Al-Lahjah : Jurnal Pendidikan, Bahasa Arab, Dan Kajian Linguistik Arab*, *6*(2), Article 2. https://doi.org/10.32764/allahjah.v6i2.3858

Nasyidiah, F. I., Siahaan, P., & Sasmita, D. (2020). Pengembangan Instrumen Four-Tier Diagnostic Test Untuk Mendeteksi Miskonsepsi Siswa Kelas X Pada Materi Impuls. *WaPFi (Wahana Pendidikan Fisika)*, *5*(2), 31–40. https://doi.org/10.17509/wapfi.v5i2.27156

Nurbayan, D. R., Nurbayan, Y., & Falah, K. N. (2020). Grammatical Error of Arabic Language in Student Thesis Department of Education Arabic Language FBPS UPI/ Kesalahan Nahwu Bahasa Arab Dalam Skripsi Mahasiswa Departemen Pendidikan Bahasa Arab FBPS UPI. *Ijaz Arabi Journal of Arabic Learning*, *3*(2). https://doi.org/10.18860/ijazarabi.v3i2.8293

Prabowo, A., Herman, T., & Fatimah, S. (2023). *The Characteristics of Four-Tier Diagnostic Test: Classical Test Theory Perspective*. *7*(1).

Pulu, S. R., & Amahoru, A. H. (2023). Analisis Miskonsepsi Mahasiswa pada Pembelajaran IPA menggunakan Tes Diagnostik Multiple Choice Berbantuan CRI (Certainty of Response Index). *JURNAL PENDIDIKAN MIPA*, *13*(2), Article 2. https://doi.org/10.37630/jpm.v13i2.1039

Putri, W. O., Maria, H. T., & Hamdani, H. (2021). Analisis Miskonsepsi Menggunakan Four Tier Diagnostic Test Berbantuan Google Form pada

Materi Tekanan Zat. *Edukatif : Jurnal Ilmu Pendidikan*, *3*(6), 4738–4749. https://doi.org/10.31004/edukatif.v3i6.1445

Rezi, M., & Amrina, A. (2019). SEMIT: Asal Muasal Bahasa Arab. *Lughawiyah: Journal of Arabic Education and Linguistics*, *1*(2). https://doi.org/10.31958/lughawiyah.v1i2.1524

Rohmanasari, F., & Ermawati, F. (2020). Using Four-Tier Multiple Choice Diagnostic Test to Identify Misconception Profile of 12th Grade Students In Optical Instrument Concepts. *Journal of Physics: Conference Series*, *1491*(1), 012011. https://doi.org/10.1088/1742-6596/1491/1/012011

Rusilowati, A., Darsono, T., Sekarningtias, F. O., & Fariyani, Q. (2024). Development of A Four-Tier E-Diagnostic Test on The Topic of Momentum to Measure and Reduce Student Misconception. *Jurnal Pendidikan Fisika Indonesia*, *20*(2), Article 2. https://doi.org/10.15294/jpfi.v20i2.1297

Safitri, M., Chotimah, C., Mufidah, Z., & Putra, I. A. (2024). Students' Concept Understanding Through Four Tier Diagnostic Test In Fiqih Subject Of Class XI Madrasah Aliyah. *APPLICATION: Applied Science in Learning Research*, *3*(3), Article 3. https://doi.org/10.32764/application.v3i3.4610

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

Treagust, D. F. (2006). Diagnostic assessment in science as a means to improving teaching, learning and retention. *Proceedings of The Australian Conference on Science and Mathematics Education*. https://openjournals.library.sydney.edu.au/IISME/article/view/6375

Zaenudin, Achmad. "Kompetensi Awal Peserta Didik dan Implikasinya dalam Pembelajaran Bahasa Arab." *Madaniyah,* vol. 11, no. 1, 31 Jan. 2021, pp. 33-48.