

تطوير أداة تقويم القواعد النحوية في ضوء مدخل الاختبار التشخيصي بأربعة مستويات (Four Tier Diagnostic Test) لتحديد المفاهيم الخاطئة عند طلبة المرحلة الثانوية

Mohammad Fahmi Bajuri<sup>1</sup>, Afif Kholisun Nashoih<sup>2</sup>

Universitas KH. A. Wahab Hasbullah Jombang, Indonesia<sup>1,2</sup>

Email: [fahmibajuri12@gmail.com](mailto:fahmibajuri12@gmail.com)<sup>1</sup>, [afifkholis@unwaha.ac.id](mailto:afifkholis@unwaha.ac.id)<sup>2</sup>

#### ABSTRACT

Misconceptions are one of the problems that often occur in the learning process. Misconceptions occur when students form incorrect or incomplete understandings of basic concepts, which can potentially hinder their learning process in the future. This study aims to develop a four-tier diagnostic test instrument to identify students' misconceptions related to qawaid nahwiyah material at MTs Mu'allimin Mu'allimat Denanyar Jombang. The study uses the Research and Development (R&D) method based on Sugiyono's development model. The small-scale pilot test subjects were students in class IX A, while the large-scale pilot test involved students in classes IX B and IX C. Data collection techniques included tests, questionnaires, and interviews. The instrument consists of a test blueprint, instructions, test questions, answer keys, scoring guidelines, and result interpretation guidelines. The validation results by experts indicated that the developed instrument was valid. The reliability of the instrument reached a score of 0.860, indicating a very high level of consistency. Item analysis showed that the instrument consisted of 22 questions with moderate difficulty and 3 questions with high difficulty. Based on discriminative power, 3 items are categorized as excellent, 12 items are categorized as good, and 12 items are categorized as adequate. The misconception profile of students shows that there is 25% misconception in CDQ answers, 33% in CDQ reasons, and 25% in the combination of CDQ answers and reasons. The highest misconception was in analyzing errors in sentences at 49%, and the lowest misconception was in analyzing the position of a word (tarkib) in a sentence at 10.5%.

Keywords: Four-Tier Diagnostic Test, Misconceptions, Qawaid Nahwiyah.

#### ملخص

المفاهيم الخاطئة هي إحدى المشكلات الشائعة في عملية التعلم، وتحدث المفاهيم الخاطئة عندما يشكل الطلاب فهمًا خاطئًا أو غير كامل للمفاهيم الأساسية، مما قد يعوق عملية تعلمهم في المستقبل. تهدف هذه الدراسة إلى تطوير أداة تقويم القواعد النحوية في ضوء

مدخل الاختبار التشخيصي بأربعة مستويات (Four Tier Diagnostic Test) لتحديد المفاهيم الخاطئة عند طلبة فيما يتعلق بمواد القواعد النحوية في مدرسة متوسطة المعلمين المعلمين دينانيار جومبانج. تستخدم الدراسة طريقة البحث والتطوير (R&D) بناءً على نموذج التطوير الذي وضعه سوجيونو. الاختبار التجريبي الصغير شمل طلاب الصف التاسع أ، بينما شمل الاختبار التجريبي الواسع طلاب الصف التاسع ب والصف التاسع ج. وتضمنت تقنيات جمع البيانات الاختبارات والاستبيانات والمقابلات. وتتكون الأداة من مخطط الاختبار، وإرشادات الإجابة، وأسئلة الاختبار، ومفتاح الإجابة، ودليل التقييم، ودليل تفسير النتائج. نتائج التحقق من قبل الخبراء أظهرت أن الأدوات التي تم تطويرها صالحة. موثوقية الأدوات بلغت 0.860، مما يدل على مستوى عالٍ جدًا من الاتساق. تحليل بنود الأسئلة أظهر أن الأدوات تتكون من 22 سؤالاً بمستوى صعوبة متوسط و3 أسئلة بمستوى صعوبة عالٍ. بناءً على القدرة التمييزية، تدرج 3 أسئلة في فئة جيد جدًا، و12 سؤالاً في فئة جيد، و12 سؤالاً في فئة كافٍ. يُظهر ملف المفاهيم الخاطئة لدى الطلاب أن هناك 25% من المفاهيم الخاطئة في إجابات CDQ، و33% في أسباب CDQ، و25% في مزيج إجابات CDQ والأسباب. أعلى نسبة من المفاهيم الخاطئة كانت في تحليل الأخطاء في الجملة بنسبة 49%، وأدنى نسبة كانت في تحليل موقع كلمة (تركيب) في الجملة بنسبة 10.5%.

**الكلمات المفتاحية:** اختبار التشخيصي بأربعة مستويات، المفاهيم الخاطئة، قواعد النحوية.

## المقدمة

إن اللغة العربية تلعب دورًا استراتيجيًا في مجال التعليم الإسلامي. فبالإضافة إلى كونها وسيلة تواصل، تُعدّ اللغة العربية أيضًا مفتاحًا أساسيًا لفهم مصادر التعاليم الدينية،<sup>1</sup> كالقرآن الكريم والحديث

---

<sup>1</sup> Nuryanto, Ahmad Rifa'i, and Abbas Sofwan Matlail Fajar, "The Development of Nahwu Teaching Materials Based on Mind Maps Using the ADDIE Model" 1, no. 1 (2024): 35–47, <https://creativecommons.org/licenses/by-sa/4.0/>.

الشريف وغيرهما من المؤلفات الإسلامية الكلاسيكية. وفي مستوى المدارس المتوسطة والثانوية، يُعدّ إتقان اللغة العربية من أهم المهارات التي يجب على الطلاب إتقانها،<sup>2</sup> لا سيما لدعم فهم النصوص الدينية. ومن العناصر الأساسية في تعليم اللغة العربية علم النحو، وهو علم ينظم بناء الجملة والعلاقات بين عناصرها. ويؤثر إتقان النحو بشكل كبير على قدرة الطلاب على فهم الجمل العربية وتكوينها بشكل صحيح.

ولكن في الواقع، من المعلوم أن تعليم النحو لا يزال يواجه تحديات جسيمة.<sup>3</sup> في مختلف المدارس الدينية، بما في ذلك مدرسة "المعلمين المعلمات دينانيار جومبانج"، يواجه العديد من الطلاب صعوبة في فهم المفاهيم الأساسية للنحو. ويعود ذلك إلى أن مفاهيم النحو غالبًا ما تكون مجردة وذات قواعد معقدة.<sup>4</sup> ويميل العديد من الطلاب إلى ارتكاب أخطاء متكررة، أو حتى لا يدركون وجود أخطاء في الفهم أو مفاهيم خاطئة لديهم.<sup>5</sup>

إن المفاهيم الخاطئة أو سوء الفهم (misconceptions) هو شكل من أشكال سوء تفسير المفاهيم العلمية التي تنحرف عن الحقيقة الأكاديمية،<sup>6</sup> وتحدث المفاهيم الخاطئة عندما يكون المتعلمون فهمًا غير صحيح أو غير مكتمل للمفاهيم الأساسية، مما قد يعيق عملية التعلم في المستقبل.<sup>7</sup> في سياق تعلم

---

<sup>2</sup> Afif Kholisun Nashoih, "HOTS-Based Qiro'ah Textbook Development to Enhance Arabic Learners' Extensive Reading Skills," *DINAMIKA : Jurnal Kajian Pendidikan Dan Keislaman* 09, no. 02 (2024): 24–42, <https://doi.org/10.32764/dinamika.v9i2.4751>.

<sup>3</sup> Fina Nur Azizah and Afif Kholisun Nashoih, "Development of Functional Arabic Grammar (Nahwu) Textbooks Based on Local Wisdom for Arabic Language Students," *Arabiyatuna: Jurnal Bahasa Arab* 8, no. 2 (2024): 703–30.

<sup>4</sup> Afif Kholisun Nashoih and M Faridl Darmawan, "Pengembangan Bahan Ajar Nahwu Berbasis Kontrastif Untuk Mengatasi Interferensi Bahasa Indonesia Terhadap Bahasa Arab," *Arabiyatuna: Jurnal Bahasa Arab* 3, no. 2 (2019): 335–50, <https://doi.org/10.29240/jba.v3i2.1008>; Dedeng Rosyidin Nurbayan, Yayan Nurbayan, and Kurnia Nurul Falah, "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, no. 2 (2020): 274–90, <https://doi.org/10.18860/ijazarabi.v3i2.8293>.

<sup>5</sup> Nurul Mukhlisa, "Miskonsepsi Pada Peserta Didik," *SPEED Journal: Journal of Special Education* 4, no. 2 (2021): 66–76, <https://doi.org/10.31537/speed.v4i2.403>; Sanaa Jauza Roghdah, Muhammad Zammi, and Julia Mardhiya, "Pengembangan Four-Tier Multiple Choice Diagnostic Test Untuk Mengetahui Tingkat Pemahaman Konsep Peserta Didik Pada Materi Termokimia," *Phenomenon: Jurnal Pendidikan MIPA* 11, no. 1 (2021): 57–74, <https://doi.org/10.21580/phen.2021.11.1.8573>.

<sup>6</sup> Nafisa Meila, Fakhruddin, and Islami Nur, "Pengembangan Tes Diagnostik Four Tier Untuk Menganalisis Konsepsi Peserta Didik SMP Pada Materi Listrik Dinamis Berdasarkan Rasch Model," *Jurnal Pendidikan Dan Pengajaran* 2, no. 3 (2024): 331–37.

<sup>7</sup> S. N. Azizah et al., "Analysis of College Students Misconceptions in Astronomy Using Four-Tier Test," *Journal of Physics: Conference Series* 2165, no. 1 (2022), <https://doi.org/10.1088/1742-6596/2165/1/012004>; Sri Jumini, Banar Dwi Retyanto, and Vivi Noviyanti, "Identifikasi Miskonsepsi Fisika Menggunakan Three-Tier Diagnostic Test Pada Pokok Bahasan Kinematika Gerak," *SPEKTRA: Jurnal Kajian Pendidikan Sains* 3, no. 2 (2017): 196, <https://doi.org/10.32699/spektra.v3i2.38>.

القواعد النحوية، يصبح هذا الأمر معقدًا بشكل متزايد لأن العديد من المفاهيم في النحو مجردة ومنهجية ومليئة بالقواعد الصارمة، وبالتالي لديها القدرة على التسبب في مفاهيم خاطئة لا يمكن اكتشافها بسهولة.<sup>8</sup> لذلك، هناك حاجة إلى أداة تشخيصية فعالة لتحديد المفاهيم الخاطئة التي تحدث لدى المتعلمين في تعلم القواعد النحوية.

للأسف الشديد، يُعدّ أسلوب التقييم المستخدم في تعلم النحو أسلوبًا تجميعيًا فقط، ويقتصر على قياس النتائج النهائية في صورة إجابات صحيحة أو خاطئة، دون دراسة عملية تفكير الطلاب بعمق.<sup>9</sup> هذا النوع من الأدوات أقل قدرة على كشف المفاهيم الخاطئة بدقة، مما لا يسمح للمعلمين بالحصول على بيانات كافية لتحسين استراتيجيات التعلم بشكل مُستهدف.<sup>10</sup> لذلك، ثمة حاجة إلى أسلوب تقييم تشخيصي يقدر على كشف فهم الطلاب العميق، بما في ذلك أسبابهم ومعتقداتهم حول الإجابات المقدمة.

أحد الأساليب المبتكرة ذات الصلة لمعالجة هذه المشكلة هو اختبار التشخيص ذي المستويات الأربعة، وهو اختبار تشخيصي من أربع مستويات مصمم لقياس ليس فقط الجوانب المعرفية، ولكن أيضًا معتقدات الطلاب وأسباب إجاباتهم على الأسئلة. تشمل المستويات الأربعة: (1) إجابات الأسئلة المفاهيمية، (2) مستوى الثقة، (3) الأسباب وراء اختيار الإجابة، و(4) مستوى الثقة في الأسباب المقدمة.<sup>11</sup> يسمح هذا النهج للمعلمين ليس فقط باكتشاف أخطاء الطلاب، ولكن أيضًا

---

<sup>8</sup> Shabrina Khairunnisa Aripiani et al., "Diagnostic Instruments of Four-Tier Test Work and Energy (FORTUNE) to Identify The Level of Students' Conceptions," *Tadris: Jurnal Keguruan Dan Ilmu Tarbiyah* 8, no. 1 (2023): 19–32, <https://doi.org/10.24042/tadris.v8i1.13524>.

<sup>9</sup> Nur Fadilah Amin, "Konsep Dasar Evaluasi Pembelajaran Bahasa Arab," *Matluba: Journal of Arabic Language and Education* 1, no. 1 (2023): 65–76, <https://doi.org/10.57215/matluba.v1i1.320>; Jumadi Jumadi, Moh Irma Sukarelawan, and Heru Kuswanto, "An Investigation of Item Bias in the Four-Tier Diagnostic Test Using Rasch Model," *International Journal of Evaluation and Research in Education* 12, no. 2 (2023): 622–29, <https://doi.org/10.11591/ijere.v12i2.22845>.

<sup>10</sup> Mojtaba Mohammadi, Maryam Zarrabi, and Jaber Kamali, "Formative Assessment Feedback to Enhance the Writing Performance of Iranian IELTS Candidates: Blending Teacher and Automated Writing Evaluation," *International Journal of Language Testing* 13, no. 1 (2023): 206–24, <https://doi.org/10.22034/ijlt.2022.364072.1201>; Nur Qomari et al., "Evaluation of Concept Maps Strategy in Learning Al-Qowa'id Al-Nahwiyyah I Taqwim Istirajiyah Kharaiih Al-Mafahim Fi Ta'lim Al-Qawa'id Al-Nahwiyyah," *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab* 16, no. 1 (2024): 125, <https://doi.org/10.24042/albayan.v16i1.19286>.

<sup>11</sup> Lailatul Mathoriyah and Afif Kholisun Nashoih, "Taṭwīr Adawāt Taqwīm Ta'lim Al Balāgh 'Abra Minaṣṣah Wordwall Fī Ḍauī Madkhal Al-Ikhtibār at-Tasykhiṣī Li Arba'Ati Mustawayāt," *Journal of Arabic Education and Literature* 8, no. 2 (2024): 227–56; Funky Iqlima Nasyidiah, Parsaoran Siahaan, and Dedi Sasmita, "Pengembangan Instrumen Four-Tier Diagnostic Test Untuk Mendeteksi Miskonsepsi Siswa Kelas X Pada Materi Impuls," *WaPFI (Wahana Pendidikan Fisika)* 5, no. 2 (2020): 31–40, <https://doi.org/10.17509/wapfi.v5i2.27156>; Zaleha Zaleha, Achmad Samsudin, and Muhamad Gina Nugraha,

بفهم مستوى ثقة الطلاب في إجاباتهم، بحيث يمكن للمعلمين تحديد ما إذا كان الخطأ ناتجًا عن نقص المعرفة أو سوء الفهم.<sup>12</sup>

تختلف هذه الأداة عن الاختبارات التقليدية لأنها تحتوي على أربع مراحل وهي الأسئلة الموضوعية والثقة في الإجابة والأسباب وراء الإجابة والثقة في الأسباب.<sup>13</sup> وهذا يسمح للمعلمين ليس فقط بتحديد الإجابات غير الصحيحة ولكن أيضًا بفهم كيفية توصل الطلاب إلى هذه الاستنتاجات ومدى ثقتهم.<sup>14</sup> لذا، ليس من المبالغة القول إن تطوير هذا النوع من الأدوات أمر بالغ الأهمية، لأنه من خلال اختبار التشخيص ذي المستويات الأربعة، يمكن للمعلمين توفير تركيز أكثر استهدافًا على مواد النحو التي غالبًا ما يكون لديها القدرة على التسبب في مفاهيم خاطئة.

أظهرت دراسات سابقة أن أدوات مثل اختبار التشخيص رباعي المستويات فعالة للغاية في تحديد المفاهيم الخاطئة، وخاصةً في العلوم والرياضيات.<sup>15</sup> ومع ذلك، لا يزال استخدامها في مجال اللغة، وخاصةً العربية، جديدًا نسبيًا ونادرًا ما يُستخدم. في الواقع، لا تقل أهمية الحاجة إلى التقييم التشخيصي في تعلم اللغة العربية، وخاصةً النحو. باستخدام هذه الأداة، يمكن للمعلمين تصميم

---

“Pengembangan Instrumen Tes Diagnostik VCCI Bentuk Four-Tier Test Pada Konsep Getaran,” *Jurnal Pendidikan Fisika Dan Keilmuan (JPFK)* 3, no. 1 (2017): 36, <https://doi.org/10.25273/jpfk.v3i1.980>.

<sup>12</sup> Zahra Javidanmehr and Mohammad Reza Anani Sarab, “Cognitive Diagnostic Assessment: Issues and Considerations,” *International Journal of Language Testing* 7, no. 2 (2017): 73–98; R. Diani et al., “Four-Tier Diagnostic Test with Certainty of Response Index on the Concepts of Fluid,” *Journal of Physics: Conference Series* 1155, no. 1 (2019), <https://doi.org/10.1088/1742-6596/1155/1/012078>; Defrina Natalia, Jeffry Handhika, and Farida Huriawati, “Pengembangan Instrumen Tes Diagnosis Konsep IPA Fisika,” *Momentum: Physics Education Journal* 1, no. 2 (2017): 103, <https://doi.org/10.21067/mpej.v1i2.1966>.

<sup>13</sup> Pangoloan Soleman Ritonga and Arif Yasthophi, “Pengembangan Instrumen Test Diagnostik Multiple Choice Four Tier Pada Materi Ikatan Kimia,” *Konfigurasi: Jurnal Pendidikan Kimia Dan Terapan* 3, no. 1 (2019): 23–31, <https://doi.org/10.24014/konfigurasi.v3i1.6797>.

<sup>14</sup> Dedah Siti Jubaedah et al., “Pengembangan Tes Diagnostik Berformat Four-Tier Untuk Mengidentifikasi Miskonsepsi Siswa Pada Topik Usaha Dan Energi” (Pendidikan Fisika dan Fisika FMIPA UNJ, 2017), SNF2017-RND-35-SNF2017-RND-40, <https://doi.org/10.21009/03.SNF2017.01.RND.06>; Sharareh Sadat Sarsarabi and Zeinab Sazegar, “Investigating Different Kinds of Stems in Multiple-Choice Tests: Interruptive Vs. Cumulative,” *International Journal of Language Testing* 13, no. 2 (2023): 170–87, <https://doi.org/10.22034/IJLT.2023.392953.1250>.

<sup>15</sup> Sylvi Aidiya Febriyana, Winny Liliawati, and Ida Kaniawati, “Identifikasi Miskonsepsi Dan Penyebabnya Pada Materi Gelombang Stasioner Kelas Xi,” *Jurnal Fisika Dan Pendidikan Fisika* 5, no. 2 (2020): 42–51; Maison Maison, Dwi Agus Kurniawan, and Retno Sari Widowati, “The Quality of Four-Tier Diagnostic Test Misconception Instrument for Parabolic Motion,” *Jurnal Pendidikan Dan Pengajaran* 54, no. 2 (2021): 359, <https://doi.org/10.23887/jpp.v54i2.35261>; Yuberti Yuberti, Yani Suryani, and Indah Kurniawati, “Four Tier Diagnostic Test With Certainty of Response Index to Identify Misconception in Physics,” *Indonesian Journal of Science and Mathematics Education* 3, no. 2 (2020): 245–53, <https://doi.org/10.24042/ijmsme.v3i2.6061>.

تدخلات تعليمية أكثر استهدافًا، إما بإعادة ترتيب المواد التعليمية، أو اختيار الأساليب المناسبة، أو تطوير كتب مدرسية سياقية.<sup>16</sup>

بناءً على الملاحظات الأولية التي قام به الباحثان في مدرسة المعلمين المعلمات دينيار جومبانج، وُجدت اختلافات كبيرة في مستوى فهم الطلاب لمادة القواعد النحوية، وخاصةً لدى طلاب الصف التاسع. واجه بعض الطلاب صعوبة في التمييز بين فئات الكلمات كاسم وفعل وحرف، وفي فهم تطبيق الإعراب في الجمل الاسمية والفعلية. لذا، إن هذه المشكلات في حاجة ماسة إلى الحل عبر تطوير أداة التقييم القادرة على اكتشاف المفاهيم الخاطئة لدى الطلاب بدقة ومنظمة.

بضوء ما قد ذكر سابقاً من خلفية البحث، تهدف هذه الدراسة إلى تطوير أداة اختبار تشخيصي لأربعة مستويات في تعليم القواعد النحوية في مدرسة المعلمين دينيار جومبانج، لتكون أداة تقييم صالحة وموثوقة، ويمكن تطبيقها بشكل جيد. ومن المتوقع أن تكون هذه الأداة قادرة على تحديد المفاهيم الخاطئة لدى الطلاب بدقة، وأن تكون أيضاً أساساً لتطوير أساليب واستراتيجيات تعلم أكثر فعالية واستجابة لاحتياجات الطلاب.<sup>17</sup> وبالتالي، فإن هذه الدراسة لا تساهم فقط في تطوير أدوات تقييم صالحة ومبتكرة، بل تلعب أيضاً دوراً في تحسين جودة تعليم اللغة العربية تعلمها في المدارس على نطاق أوسع.

هذا البحث هو نوع من البحوث ذات منهج البحث والتطوير (R&D) الذي يهدف إلى تطوير أداة اختبار على شكل اختبار رباعي المستويات لتحديد مدى فهم الطلاب لمفهوم مادة القواعد النحوية.<sup>18</sup> ووفقاً لسوجيونو، فإن البحث والتطوير (R&D) هو أسلوب بحثي يهدف إلى إنتاج منتج ما

---

<sup>16</sup> Qisthi Fariyani, "Pengembangan Four-Tier Diagnostic Test Untuk Mengungkap Miskonsepsi Fisika Siswa Sma Kelas X," *Journal of Innovative Science Education* 4, no. 2 (2015): 41–49; Dewi Yulianawati, Asih Wahyuningsih, and Nur Aisa Pebriana, "Pengembangan Instrumen 4TSDT (Four Tier – Science Diagnostic Test) Untuk Mengidentifikasi Level Konsepsi Calon Guru Sekolah Dasar," *Jurnal Basicedu* 6, no. 6 (2022): 9483–90, <https://doi.org/10.31004/basicedu.v6i6.4117>.

<sup>17</sup> Pujia Rawh, Achmad Samsudin, and Muhamad Gina Nugraha, "Pengembangan Four-Tier Diagnostic Test Untuk Mengidentifikasi Profil Konsepsi Siswa Pada Materi Alat-Alat Optik," *WaPFI (Wahana Pendidikan Fisika)* 5, no. 1 (2020): 84–89, <https://doi.org/10.17509/wapfi.v5i1.22888>; Shanthi Tiruchttampalam and Alistair Ross, "Validation of a Screening Tool for Dyslexia in English among Arabic-Speaking University Students in the Gulf Area," *International Journal of Arabic-English Studies* 24, no. 2 (2024): 231–48, <https://doi.org/10.33806/ijaes.v24i2.662>.

<sup>18</sup> Torang Siregar, "Stages of Research and Development Model Research and Development (R&D)," *DIROSAT: Journal of Education, Social Sciences & Humanities* 1, no. 4 (2023): 142–58, <https://doi.org/10.58355/dirosat.v1i4.48>.

بالإضافة إلى اختبار فعالية المنتج أو نجاحه.<sup>19</sup> وبناءً على هذا التفسير، يهدف هذا البحث التطويري إلى إنتاج منتج يتم التحقق من صلاحيته من قبل الخبراء واختباره على الطلاب. المنتج الذي تم تطويره في هذه الدراسة هو أداة اختبار تشخيصية من أربعة مستويات تُستخدم لتحديد المفاهيم الخاطئة لدى الطلاب في مادة القواعد النحوية.

لتطوير أداة اختبار تشخيصية من أربعة مستويات تكون فعّالة في تحديد المفاهيم الخاطئة في مادة القواعد النحوية؛ يتبع هذا البحث إجراء البحث والتطوير (R&D) وفقاً لسوجيونو. تم تصميم هذا الإجراء لإنتاج منتجات صالحة وموثوقة من خلال سلسلة من المراحل المنهجية والمنظمة. وتشمل هذه المراحل تحديد المشكلة، وتصميم المنتج، والتحقق من صحة المنتج، والتحقق من صحة الخبراء، والتجربة، والمراجعة، بهدف ضمان جودة وفعالية الأداة في التعلم.

تمثلت الخطوة الأولى في تحديد المشاكل المحتملة والقائمة في هذا المجال والمتعلقة بالحاجة إلى أدوات تشخيصية حول مادة القواعد النحوية. وعلاوة على ذلك، تم جمع البيانات من خلال دراسة الأدبيات وتحليل المناهج الدراسية لتعزيز أساس تطوير المنتج. وتتمثل المرحلة التالية في تصميم منتج على شكل أداة اختبار تشخيصية من أربعة مستويات، تتضمن شبكة من الأسئلة والبنود وتعليمات العمل.

بعد الانتهاء من تصميم المنتج، تم التحقق من صحته من قبل خبراء لضمان جودة محتواه وبنائه. وبناءً على الملاحظات التي تم تلقيها، تم تنقيح المنتج بعد ذلك. تم اختبار الأداة المنقحة على مجموعة صغيرة من المتعلمين لقياس فعاليتها. سيتم استخدام نتائج هذه التجربة المحدودة لإجراء المزيد من التحسينات. بعد ذلك، تم اختبار المنتج على نطاق أوسع على الطلاب في الصفين التاسع ب والتاسع ج للحصول على بيانات أكثر تمثيلاً. وبعد إجراء التجربة على نطاق واسع، خضع المنتج لمراجعة نهائية للتأكد من أن الأداة الناتجة جاهزة للاستخدام في التعلم.

تم تنفيذ تقنيات جمع البيانات في هذه الدراسة من خلال ثلاث طرق، وهي الاختبارات والاستبيانات والمقابلات. وأجري تحليل المنتج في هذه الدراسة من خلال المنهجين الكيفي والكمي. في التحليل النوعي، تم إجراء التقييم من قبل خبراء، وهم خبراء المواد وخبراء التقييم. وقاموا بتقييم كل

---

<sup>19</sup> Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, Dan R&D*, 19th ed., vol. 7 (Bandung: CV. Alfabeta, 2013).

سؤال بناءً على عدة جوانب، مثل مدى ملاءمة محتوى المادة، ودقة ترتيب الأسئلة (البناء)، ووضوح اللغة، ودقة مفاتيح الإجابة، واكتمال إرشادات تسجيل الدرجات. تم استخدام مدخلات هؤلاء الخبراء لتحسين الأداة وتنقيحها قبل استخدامها في الاختبار التجريبي.

أما في التحليل الكمي، فقد تم تحليل البيانات التي تم الحصول عليها من نتائج تجربة الأداة لحساب صحة البنود، وموثوقية الأداة، ومستوى صعوبة الأسئلة، وقوة التفريق، وأداء المدققين. تم حساب اختبار الموثوقية باستخدام معادلة كرونباخ ألفا للتأكد من اتساق الأداة. بالإضافة إلى ذلك، تم تحليل المفاهيم الخاطئة لدى الطلاب باستخدام حساب حاصل تمييز الثقة (حاصل تمييز الثقة)، باتباع الطريقة التي استخدمها كاليون وسوبرامانيام في تطوير الاختبار التشخيصي رباعي المستويات.<sup>20</sup>

## النتائج والمناقشة

### أ. تطوير أداة الاختبار التشخيصية رباعية المستويات

إن أداة الاختبار التشخيصي رباعي المستويات التي تم تطويرها في هذه الدراسة هي شكل من أشكال الاختبار المصمم للكشف عن المفاهيم الخاطئة لدى الطلاب. وتركز المادة المستخدمة في إعداد هذه الأداة على قواعد اللغة العربية. ويتألف المنتج الأولي لهذه الأداة من 25 بندًا تهدف إلى قياس فهم المتعلمين لمفاهيم قواعد اللغة العربية النحوية مع تحديد المفاهيم الخاطئة أو التصورات الخاطئة الشائعة. وبالإضافة إلى السؤال الرئيسي للاختبار التشخيصي رباعي المستويات، فإن المنتج المطور مزود أيضًا بشبكات الأسئلة، وإرشادات للعمل على الأسئلة، ونصوص الأسئلة، ومفاتيح الإجابة، وإرشادات لتحديد الدرجات، وإرشادات لتفسير النتائج.

تم إعداد شبكة الاختبار التشخيصي لأربعة المستويات بناءً على ستة عناصر رأسية، وهي: مؤشرات تحصيل الكفاءة، والمواد الفرعية، ومؤشرات الأسئلة، ومستويات الأسئلة، وأنواع الأسئلة، وأعداد الأسئلة، وقد صُممت مؤشرات تحصيل الكفاءة ومؤشرات الأسئلة لتكون مرجعًا في إعداد الأسئلة. تعمل الشبكة على تجميع الأسئلة وفقًا للمعدل التراكمي ومؤشرات الأسئلة، بالإضافة إلى

---

<sup>20</sup> Imelda S. Caleon and R. Subramaniam, "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, no. 3 (2010): 313–37, <https://doi.org/10.1007/s11165-009-9122-4>.

إظهار المادة الفرعية التي تم اختبارها، ومستوى الأسئلة، ونوع الأسئلة، وعدد الأسئلة. كانت جميع الأسئلة متعددة الاختيارات وتم تطويرها بالرجوع إلى تصنيف بلوم، الذي يغطي أربعة مجالات معرفية: التذكر (C1)، والفهم (C2)، والتطبيق (C3)، والتحليل (C4).<sup>21</sup>

### الجدول 1. جدول تحديد الأسئلة

الرقم	الكفاءة الأساسية	المادة	مؤشر السؤال	المستوى المعرفي	رقم السؤال
1	تحليل موقع لفظ (تركيب) في الجملة.	الموقع في الجملة	يستطيع المتعلم تحديد موقع اللفظ (التركيب) كمبتدأ، خبر، فاعل، نائب فاعل، اسم كان وأخواتها، خبر إن وأخواتها، أو بدل في الجملة.	C4	١، ٤
2	تحديد الإعراب للفظ في الجملة.	الإعراب في الجملة	يستطيع المتعلم تحديد نوع الإعراب (رفع، نصب، جر، أو جزم) للفظ معين في الجملة.	C3	٣، ٥، ٧، ١٠، ١٣، ١٧، ١٩، ٢٠، ٢١
3	التعرف على القواعد النحوية	قواعد اللغة العربية	يستطيع المتعلم اختيار بنية الجملة المناسبة وفقاً لقواعد اللغة العربية.	C4	٢، ٩، ١٤، ١٨، ٢٢
4	تحليل أنواع الكلمات (اسم، فعل، حرف).	أنواع الألفاظ	يستطيع المتعلم التمييز بين الاسم، الفعل، والحرف في الجملة.	C2	١٢، ١٥، ١٦

<sup>21</sup> Arif Yasthophi and Pangoloan Soleman Ritonga, "Pengembangan Instrumen Test Diagnostik Multiple Choice Four Tier Pada Materi Ikatan Kimia," *Konfigurasi: Jurnal Pendidikan Kimia Dan Terapan* 3, no. 1 (February 2019): 23, <https://doi.org/10.24014/konfigurasi.v3i1.6797>.

١١، ٨، ٦	C4	يستطيع المتعلم تحديد نوع الاسم كنكرة، معرفة، مذكر، مؤنث، أو غيرها في الجملة.	تحليل الأخطاء في الجملة	استخلاص الخطأ النحوي.	5
----------	----	--	-------------------------------	--------------------------	---

تعمل التعليمات كدليل إرشادي للطلاب أثناء أداء الاختبار. ويتضمن هذا الدليل القواعد التي يجب الالتزام بها، وإجراءات العمل على الأسئلة، وعدد الأسئلة التي يجب الإجابة عليها، ومدة زمن المعالجة. هناك خمس نقاط مهمة في هذه التعليمات يجب أن يفهمها الطلاب حتى تسير عملية العمل على الاختبار التشخيصي رباعي المستويات في مادة القواعد النحوية بشكل جيد ووفقًا للأحكام.

يتكون كل عنصر في الاختبار التشخيصي رباعي المستويات من أربعة مستويات. يحتوي المستوى الأول على أسئلة متعددة الخيارات مع أربعة خيارات للإجابة وإجابة واحدة صحيحة. ويطلب المستوى الثاني من المتعلمين الإشارة إلى مستوى ثقتهم في الإجابة التي اختاروها. يقدم المستوى الثالث أربعة خيارات للأسباب الكامنة وراء إجابة المتعلم. يقيس المستوى الرابع مستوى الثقة في السبب المختار، على مقياس من واحد إلى خمسة. يهدف الفصل بين مستويات الثقة للإجابات والأسباب إلى الكشف عن الاختلافات في مستوى يقين المتعلمين في الإجابة.<sup>22</sup> في المجموع، تم تطوير 25 عنصرًا للتعرف على المفاهيم الخاطئة لدى المتعلمين حول القواعد النحوية.

### الشكل 1 - مشكلة العينة

1. قرأ الآتي:
- (Tier 1) Lafadz " رَبِّهٖ " pada susunan kalimat diatas berkedudukan sebagai?
- a. Fa'il                      b. Ma'ul bih                      c. Khobar                      d. Mubtada'
- (Tier 2) Seberapa yakin Anda dengan jawaban Anda diatas?
- a. Sangat yakin                      b. Yakin                      c. Ragu-ragu                      d. Tidak yakin                      e. Sangat tidak yakin
- (Tier 3) Apa alasan Anda memilih jawaban diatas?
- a. Karena رَبِّهٖ sebagai objek pekerjaan  
b. Karena رَبِّهٖ sebagai pelengkap pekerjaan  
c. Karena رَبِّهٖ sebagai subjek pekerjaan  
d. Karena رَبِّهٖ sebagai fi'il
- (Tier 4) Seberapa yakin Anda dengan alasan Anda dalam memilih jawaban diatas?
- a. Sangat yakin                      b. Yakin                      c. Ragu-ragu                      d. Tidak yakin                      e. Sangat tidak yakin

<sup>22</sup> Irnin Agustina Dwi Astuti, Yoga Budi Bhakti, and Rendi Prasetya, "Four Tier-Magnetic Diagnostic Test (4T-MDT): Instrumen Evaluasi Medan Magnet Untuk Mengidentifikasi Miskonsepsi Siswa," *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)* 5, no. 2 (2021): 110–15, <https://doi.org/10.30599/jipfri.v5i2.1205>.

تم إعداد مفتاح الإجابة كمرجع للتحقق من عمل الطلاب، بحيث يمكن إجراء التقييم بشكل صحيح وموضوعي. يتم ترتيب مفتاح الإجابة هذا على شكل جدول يحتوي على 25 رقم سؤال، والإجابة الصحيحة لكل سؤال، والسبب الصحيح الذي يتوافق مع كل سؤال. يتم استخدام دليل الدرجات للاختبار التشخيصي رباعي المستويات كدليل إرشادي في تسجيل إجابات المتعلمين، بما في ذلك اختيار الإجابة، ومستوى الثقة، والسبب الذي تم اختياره، ومستوى الثقة في السبب. تُعطى درجة 1 إذا اختار المتعلم الإجابة أو السبب الصحيح، بينما تُعطى درجة 0 للإجابة أو السبب الخاطئ. واعتُبر مستوى الثقة مرتفعاً إذا اختار المتعلمون 4 أو 5 أو 6، ومنخفضاً إذا اختاروا 1 أو 2 أو 3. ثم تم تفسير نتائج درجات المتعلمين لتحديد فئة فهمهم لكل سؤال، أي فهم أو عدم فهم أو وجود مفاهيم خاطئة.

يُستخدم الدليل الإرشادي لتفسير الاختبار التشخيصي رباعي المستويات كأساس لتصنيف الطلاب إلى فئات الفهم، أو عدم الفهم، أو وجود مفاهيم خاطئة. وقد تم تنظيم هذا الدليل الإرشادي على شكل جدول يحتوي على نوع الاستجابة للإجابة، ومستوى الثقة في الإجابة، والسبب المختار، ومستوى الثقة في السبب، بحيث يمكن تصنيف النتائج التي تم الحصول عليها بشكل أكثر منهجية ودقة.

## الجدول 2 - تفسير نتائج الاختبارات التشخيصي لأربعة المستويات

المعايير	سبب الاعتقاد	السبب	إجابة الثقة	الإجابة
فهم	عالية	صحيح	عالية	صحيح
غير مفهوم	منخفضة	صحيح	منخفضة	صحيح
	منخفضة	صحيح	عالية	صحيح
	عالية	صحيح	منخفضة	صحيح
	منخفضة	خطأ	منخفضة	صحيح
	منخفضة	صحيح	منخفضة	خطأ
	منخفضة	خطأ	منخفضة	خطأ
	منخفضة	خطأ	عالية	صحيح

	عالية	صحيح	منخفضة	خطأ
المفاهيم الخاطئة	عالية	خطأ	عالية	صحيح
	عالية	خطأ	عالية	صحيح
	منخفضة	صحيح	عالية	خطأ
	عالية	صحيح	عالية	خطأ
	منخفضة	خطأ	عالية	خطأ
	عالية	خطأ	منخفضة	خطأ
	عالية	خطأ	عالية	خطأ
	عالية	خطأ	عالية	خطأ

قبل اختبار المنتج، يتم إجراء عملية تحقق من صحة المنتج لضمان صلاحية الأداة التي تم تطويرها. شارك في عملية التحقق هذه خبيران في مجالهما، وهما خبيرا المواد وخبراء التقييم. يرد ملخص لنتائج تحليل التحقق من صحة الأداة بشكل عام في الجدول 3.

### الجدول 3. ملخص نتائج التحقق من صحة أداة السؤال من قبل المدققين

المدقق	النتيجة الإجمالية	متوسط النقاط	إجمالي المئوية	النسبة	المعايير
الخبير 1	337	13,48	89,87%		صالح جداً
الخبير 2	343	13,72	91,47%		صالح جداً
الإجمالي	680	13,60	90,67%		صالح جداً

ملاحظات: مجموع الدرجات القصوى 25 سؤالاً × 15 درجة كحد أقصى = 375 لكل مدقق.

معايير صلاحية الأداة هي كما يلي:

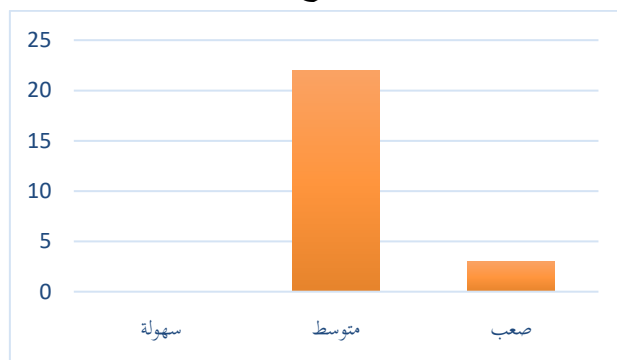
- 100% - 85,01% = صالح جداً
- 85% - 70,01% = صالحة إلى حد ما
- 70% - 50,01% = أقل صلاحية
- 50% - 0,1% = غير صالح

استنادًا إلى نتائج التلخيص، يمكن استنتاج أن أداة الأسئلة مدرجة في فئة "صالحة جدًا" بمتوسط نسبة صلاحية 90.67%. وهذا يعني أن هذه الأداة تتوافق مع المعايير المحددة، لذا فهي مناسبة للاستخدام في عملية جمع البيانات البحثية دون الحاجة إلى مراجعات كبيرة.

### ب. نتائج تجربة المنتج

تم إجراء تجارب المنتجات في هذه الدراسة على نطاق صغير. واستُخدمت نتائج التجربة لقياس الصلاحية والموثوقية ومستوى الصعوبة وقوة التمييز. استنادًا إلى نتائج تحليل تجربة الأداة على نطاق صغير، تبين أنه من بين 25 بندًا تم اختبارها، تم الإعلان عن صلاحية 22 بندًا من البنود الـ 25، بينما كانت البنود الثلاثة الأخرى غير صالحة لذا فهي بحاجة إلى القليل من التحسين. استنادًا إلى تحليل الموثوقية، تم الحصول على قيمة كرونباخ ألفا 0.860. وتتجاوز هذه القيمة قيمة الموثوقية التي تم الحصول عليها، لذا يمكن استنتاج أن أداة الأسئلة المستخدمة تتمتع بمستوى عالٍ وثابت من الموثوقية. تُظهر نتائج تحليل اختبار مستوى صعوبة الأسئلة أن هناك 3 أسئلة تندرج ضمن فئة الأسئلة الصعبة. وقد صُممت الأسئلة في هذه الفئة لقياس قدرة المتعلمين الذين يفهمون المادة بعمق. ومع ذلك، لا يزال هناك عدد قليل نسبيًا يحافظ على التوازن حتى لا يثقل كاهل الطلاب. كان هناك ما مجموعه 22 سؤالاً في الفئة المتوسطة، وكانت الأسئلة في الفئة المتوسطة قادرة على التمييز بين المتعلمين بناءً على قدراتهم بشكل أكثر فعالية. لم تكن هناك أسئلة في الفئة السهلة، مما يشير إلى أن هذه الأداة فعالة جدًا في قياس قدرات المتعلمين.

### الشكل 2 - خلاصة نتائج تحليل مستوى الصعوبة



استنادًا إلى نتائج تحليل القدرة على التمايز، هناك 3 عناصر تقع في الفئة الممتازة. وهذا يدل على أن الأسئلة الثلاثة فعالة جدًا في التمييز بين مستويات قدرات الطلاب. بالإضافة إلى ذلك، هناك

10 عناصر صُنِّفت في فئة جيد، وهو ما يُظهر أيضًا فعالية عالية في التمييز بين قدرات الطلاب. وفي الوقت نفسه، كان هناك 12 عنصرًا في فئة مقبول، والتي لا يزال من الممكن استخدامها للتمييز بين قدرات المشاركين على الرغم من أن فعاليتها أقل. لم يتم العثور على أي عناصر في فئتي ضعيف أو ضعيف جدًا، لذلك بشكل عام، أظهرت هذه الأداة قوة تمييز كافية.

### الشكل 3 - خلاصة نتائج تحليل القدرة التفاضلية



### ت. نتائج الاختبار على نطاق واسع

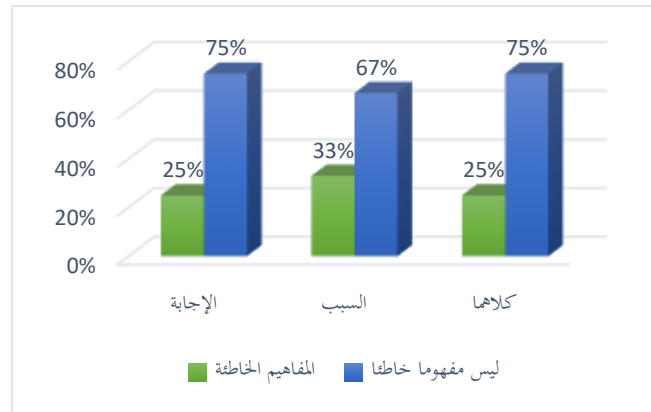
يمكن تحديد المفاهيم الخاطئة للمتعلمين من خلال تحليل قيم CDQ (حاصل تمييز الثقة). تشير قيمة CDQ السالبة إلى أن المتعلمين غير قادرين على التمييز بين الفهم الصحيح وغير الصحيح، مما يعني أن لديهم مفاهيم خاطئة بينما تشير الأسئلة التي تحمل قيمة CDQ إيجابية إلى الفهم الصحيح. تم الحصول على درجات CDQ من خلال تحليل خيارات الإجابة والأسباب التي تم اختيارها والعلاقة بين الاثنين على 25 عنصرًا قام بها 48 متعلمًا في التجربة واسعة النطاق. وكانت قيم CDQ التي تم الحصول عليها من خلال نتائج تحليل البيانات في نطاق  $-0.49$  إلى  $2.21$ .

عانى 25% أو 12 متعلمًا من المفاهيم الخاطئة لأنهم كانوا واثقين جدًا من الإجابات التي اعتبروها صحيحة. وفي الوقت نفسه، لم يواجه 75% أو 36 متعلمًا مفاهيم خاطئة. حدثت المفاهيم الخاطئة في اختيار السبب لدى 33% أو 16 متعلمًا، مما يشير إلى أنه على الرغم من إجاباتهم بشكل صحيح، إلا أنهم لم يتمكنوا من شرح السبب وراء الإجابة.

عانى ما يصل إلى 25% أو 12 متعلمًا من سوء الفهم في اختيار الإجابات والأسباب لاعتقادهم بصحة اختياراتهم. وقد استند هذا الاعتقاد إلى فهمهم الأولي والمنطق الشخصي الذي استخدموه في الإجابة. في حين أن 75% أو 36 متعلمًا لم يتعرضوا لمفاهيم خاطئة، لكنهم انقسموا

إلى فئتين: أولئك الذين فهموا المفهوم حقًا وأولئك الذين لم يفهموا المفهوم على الرغم من أنهم لم يظهروا مفاهيم خاطئة. تشير هذه النتيجة إلى وجود تفاوت في مستوى فهم المتعلمين في الإجابة عن الأسئلة وإبداء الأسباب.

#### الشكل 4- نتائج تحليل المفاهيم الخاطئة لدى الطلاب



استنادًا إلى الشكل 2، يمكن ملاحظة أن أكثر أشكال الفهم الخاطئ السائدة التي يعاني منها الطلاب تحدث على مستوى التعليل، بنسبة 33%. هذا يشير إلى أنه على الرغم من أن معظم المتعلمين قادرين على اختيار الإجابة الصحيحة، إلا أنهم غير قادرين على التوصل إلى السبب الصحيح، مما يشير إلى فهم غير صحيح للمفهوم الكامن وراء الإجابة.

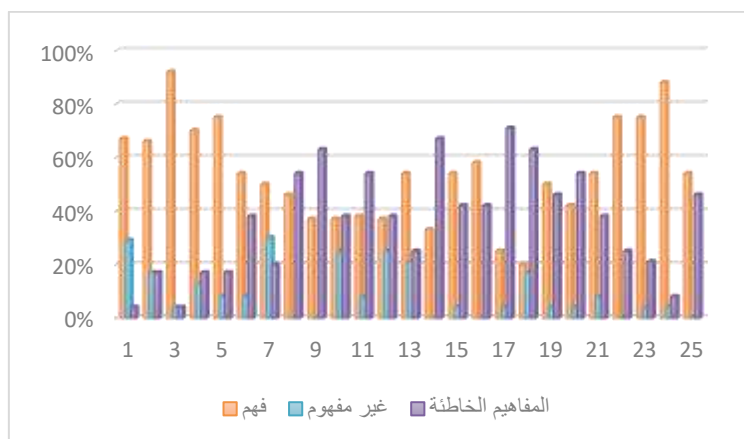
بالإضافة إلى ذلك، وُجدت أيضًا مفاهيم خاطئة على مستوى الإجابات والأسباب معًا بنسبة 25%. يشير هذا إلى أن المتعلمين لم يختاروا الإجابة الخاطئة فحسب، بل لم يفهموا أيضًا المفهوم الأساسي الذي تقوم عليه المسألة، وبالتالي قدموا أسبابًا غير مناسبة. يعكس هذا النوع من المفاهيم الخاطئة خطأ مفاهيمي أعمق.

كما بلغت نسبة المفاهيم الخاطئة على مستوى الإجابة فقط 25% من المشاركين. في هذه الفئة، أعطى المتعلمون إجابة خاطئة، لكن مع ذلك أظهر التعليل مؤشرًا على الفهم الذي كان يسير في الاتجاه الصحيح، وإن لم يكن دقيقًا تمامًا.

بشكل عام، تُظهر نتائج هذا التحليل أن المفاهيم الخاطئة لا تتضح دائمًا من الإجابات الخاطئة وحدها، بل يمكن أن تحدث أيضًا بسبب التفكير الخاطئ حتى لو كانت الإجابة صحيحة. لذلك، من المهم بالنسبة للمعلمين ليس فقط تقييم الإجابات النهائية للمتعلمين، ولكن أيضًا تحليل تفكيرهم

وتفسيراتهم. يسمح هذا النهج بتحديد أكثر عمقًا للمفاهيم الخاطئة ويساعد في تصميم تدخلات تعليمية أكثر فعالية.<sup>23</sup>

الشكل 5: النسبة المئوية للمتعلمين الذين يعانون من المفاهيم الخاطئة حول كل بند من بنود الأسئلة



واستنادًا إلى الشكل 3، هناك عشرة عناصر تدرج ضمن فئة الفهم الخاطئ المنخفض، وهي الأسئلة أرقام 1 و 2 و 3 و 4 و 5 و 5 و 7 و 13 و 22 و 23 و 24، بينما يندرج أحد عشر عنصرًا ضمن فئة الفهم الخاطئ المتوسط، وهي الأسئلة أرقام 6 و 8 و 10 و 10 و 11 و 12 و 15 و 16 و 19 و 20 و 21 و 25، بينما تدرج العناصر الأربعة الأخرى، وهي الأسئلة أرقام 9 و 14 و 17 و 18، ضمن فئة الفهم الخاطئ المرتفع.

من حيث الفهم، صنف 23 سؤالاً في فئة المستوى المنخفض لعدم الفهم، بما في ذلك الأسئلة

رقم

1،2،3،3،4،4،5،6،6،8،8،9،9،10،10،11،11،12،13،13،14،14،15،15،16

،16،17،18،18،19،20،21،22،23،24،

7، على أنه مستوى متوسط من عدم الفهم.

بالنسبة لفئة الفهم، السؤالان رقم 17 و 18 ضمن فئة مستوى الفهم المنخفض. في حين تم

إدراج 15 سؤالاً في فئة مستوى الفهم المتوسط، وهي الأسئلة أرقام

<sup>23</sup> Nila Hidayah et al., "Pengembangan Tes Four Tier Multiple Choice ( 4TMC ) Untuk Mendeteksi Miskonsepsi Peserta Didik," *Chemistry in Education* 11, no. 2 (2022): 86–94.

6,7,8,8,9,10,10,11,11,12,12,13,13,14,15,15,16,19,20,21,25

وأظهرت ستة أسئلة مستوى فهم مرتفع، وهي الأسئلة أرقام 1,2,3,4,5,22,23,24.

تُظهر نتائج تحليل الإختبار التشخيصي لأربعة المستويات أن مستوى المفاهيم الخاطئة لدى المتعلمين يتفاوت في كل مؤشر. فقد كان أقل المفاهيم الخاطئة في مؤشر تحليل موضع اللام في الجملة 10.5% أو 5 متعلمين، بينما كانت أعلى نسبة في مؤشر تحليل الأخطاء في الجملة 49% أو 24 متعلمًا.

تُظهر المؤشرات الأخرى أيضًا مستوى عالٍ إلى حد ما من المفاهيم الخاطئة وهي تحديد الإعراب 35% أو 17 متعلمًا، وفهم بنية الجملة 47% أو 23 متعلمًا، وتحليل أنواع الجمل (40% أو 20 متعلمًا)، وتحديد أمثلة الإعراب المناسبة 25% أو 12 متعلمًا. تشير هذه النتائج إلى أن المفاهيم الخاطئة هي الأكثر انتشارًا في الجوانب المتعلقة ببنية الجملة وتطبيق الإعراب، مما يتطلب المزيد من الاهتمام في عملية التعلم.

بشكل عام، تُظهر هذه النتائج أن الفهم العميق للمفاهيم مهم جدًا في عملية التعلم. يحتاج المعلمون إلى التركيز بشكل أكبر على ضمان فهم المتعلمين للمادة بشكل واضح. إذا لم يتم تصحيح المفاهيم الخاطئة على الفور، فقد يتعارض ذلك مع فهم المتعلمين للمواد المترابطة التالية.<sup>24</sup> أظهرت نتائج المقابلات التي أجريت مع المعلمين أن الإختبار التشخيصي رباعي المستويات تلقى ردوداً إيجابية. يعتبر المعلمون هذا الإختبار مفيداً كأداة تقييم لأنه قادر على تحديد الطلاب الذين يفهمون المادة الدراسية والذين لا يفهمونها، وأولئك الذين يعانون من مفاهيم خاطئة. وتساعد هذه المعلومات المعلمين في تحديد خطوات التعلم المناسبة والفعالة، وبالتالي يكون لها تأثير إيجابي على الطلاب وعملية التعلم ككل.

وفي الوقت نفسه، تم إجراء مقابلات مع المتعلمين بعد أن تم تصنيف نتائج تحليل إجاباتهم إلى الفهم وعدم الفهم والمفاهيم الخاطئة. تم إجراء المزيد من المقابلات مع المتعلمين الذين تم تحديدهم على أن لديهم مفاهيم خاطئة للتأكد من السبب، سواء كان السبب في الحقيقة أخطاء مفاهيمية أو عوامل

<sup>24</sup> Sri Rahmadani Pulu and Abd. Haji Amahoru, "Analisis Miskonsepsi Mahasiswa Pada Pembelajaran IPA Menggunakan Tes Diagnostik Multiple Choice Berbantuan CRI (Certainty of Response Index)," *Jurnal Pendidikan Mipa* 13, no. 2 (2023): 478–86, <https://doi.org/10.37630/jpm.v13i2.1039>.

أخرى مثل عدم التركيز أو أخطاء فنية. كما استكشفت المقابلة أيضاً المعرفة الأساسية للمتعلمين، بما في ذلك مصادر تعلمهم، لفهم أصل المفهوم الخاطئ. وتوفر النتائج أساساً للمعلمين لتصميم استراتيجيات تعلم أكثر ملاءمة، إما من خلال مناهج سياقية أو أساليب أكثر ملاءمة.

## الخلاصة

تهدف هذه الدراسة إلى تطوير أداة الإختبار التشخيصي لأربعة المستويات للتعرف على المفاهيم الخاطئة لدى الطلاب في فهم مادة القواعد النحوية في مدرسة المعلمين المعلمين دينانيار جومبانج. تتكون الأداة من أربعة مستويات: اختيار الإجابة، والاعتقاد في الإجابة، وإعطاء الأسباب، والاعتقاد في الأسباب. وبهذا النهج، تكون الأداة قادرة على الكشف عن الفهم المفاهيمي وعقلية الطلاب بشكل أكثر عمقاً.

وقد خضعت الأداة التي تم تطويرها لعملية التحقق من صحة الأداة من قبل خبراء بنتائج ممتازة، وتتمتع بموثوقية عالية تبلغ 0.860، متجاوزة جدول (0.482)  $r$ . من بين 25 سؤالاً، صُنّف 22 سؤالاً على أنها متوسطة الصعوبة و3 أسئلة صعبة، وتراوحت قوة التفريق بين مناسب وممتاز. أظهرت النتائج أن 25% من المتعلمين عانوا من المفاهيم الخاطئة في الإجابات، و33% منهم عانوا من الأسباب، و25% منهم عانوا من الاثنين معاً. وكانت أعلى نسبة سوء فهم في مؤشر تحليل أخطاء الجملة (49%)، بينما كانت النسبة الأقل في تحديد موضع اللام في الجملة (10.5%).

## المراجع

- Amin, Nur Fadilah. "Konsep Dasar Evaluasi Pembelajaran Bahasa Arab." *Matluba: Journal of Arabic Language and Education* 1, no. 1 (2023): 65–76. <https://doi.org/10.57215/matluba.v1i1.320>.
- Aripiani, Shabrina Khairunnisa, Achmad Samsudin, Ida Kaniawati, Hera Novia, Adam Hadiana Aminudin, Asep Dedy Sutrisno, and Bayram Coştu. "Diagnostic Instruments of Four-Tier Test Work and Energy (FORTUNE) to Identify The Level of Students' Conceptions." *Tadris: Jurnal Keguruan Dan Ilmu Tarbiyah* 8, no. 1 (2023): 19–32. <https://doi.org/10.24042/tadris.v8i1.13524>.
- Astuti, Irnin Agustina Dwi, Yoga Budi Bhakti, and Rendi Prasetya. "Four Tier-Magnetic Diagnostic Test (4T-MDT): Instrumen Evaluasi Medan Magnet Untuk Mengidentifikasi Miskonsepsi Siswa." *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)* 5, no. 2 (2021): 110–15. <https://doi.org/10.30599/jipfri.v5i2.1205>.
- Azizah, Fina Nur, and Afif Kholisun Nashoih. "Development of Functional Arabic Grammar

- (Nahwu) Textbooks Based on Local Wisdom for Arabic Language Students.” *Arabiyatuna: Jurnal Bahasa Arab* 8, no. 2 (2024): 703–30.
- Azizah, S. N., H. Akhsan, M. Muslim, and M. Ariska. “Analysis of College Students Misconceptions in Astronomy Using Four-Tier Test.” *Journal of Physics: Conference Series* 2165, no. 1 (2022). <https://doi.org/10.1088/1742-6596/2165/1/012004>.
- Caleon, Imelda S., and R. Subramaniam. “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, no. 3 (2010): 313–37. <https://doi.org/10.1007/s11165-009-9122-4>.
- Diani, R., J. Alfin, Y. M. Anggraeni, M. Mustari, and D. Fujiani. “Four-Tier Diagnostic Test with Certainty of Response Index on the Concepts of Fluid.” *Journal of Physics: Conference Series* 1155, no. 1 (2019). <https://doi.org/10.1088/1742-6596/1155/1/012078>.
- Fariyani, Qisthi. “Pengembangan Four-Tier Diagnostic Test Untuk Mengungkap Miskonsepsi Fisika Siswa Sma Kelas X.” *Journal of Innovative Science Education* 4, no. 2 (2015): 41–49.
- Febriyana, Sylvi Aidiya, Winny Liliawati, and Ida Kaniawati. “Identifikasi Miskonsepsi Dan Penyebabnya Pada Materi Gelombang Stasioner Kelas Xi.” *Jurnal Fisika Dan Pendidikan Fisika* 5, no. 2 (2020): 42–51.
- Hidayah, Nila, Sigit Priatmoko, Sri Wardani, and Sri Nurhayati. “Pengembangan Tes Four Tier Multiple Choice ( 4TMC ) Untuk Mendeteksi Miskonsepsi Peserta Didik.” *Chemistry in Education* 11, no. 2 (2022): 86–94.
- Javidanmehr, Zahra, and Mohammad Reza Anani Sarab. “Cognitive Diagnostic Assessment: Issues and Considerations.” *International Journal of Language Testing* 7, no. 2 (2017): 73–98.
- Jubaedah, Dedah Siti, Ida Kaniawati, Iyon Suyana, Achmad Samsudin, and Endi Suhendi. “Pengembangan Tes Diagnostik Berformat Four-Tier Untuk Mengidentifikasi Miskonsepsi Siswa Pada Topik Usaha Dan Energi,” SNF2017-RND-35-SNF2017-RND-40. Pendidikan Fisika dan Fisika FMIPA UNJ, 2017. <https://doi.org/10.21009/03.SNF2017.01.RND.06>.
- Jumadi, Jumadi, Moh Irma Sukarelawan, and Heru Kuswanto. “An Investigation of Item Bias in the Four-Tier Diagnostic Test Using Rasch Model.” *International Journal of Evaluation and Research in Education* 12, no. 2 (2023): 622–29. <https://doi.org/10.11591/ijere.v12i2.22845>.
- Jumini, Sri, Banar Dwi Retyanto, and Vivi Noviyanti. “Identifikasi Miskonsepsi Fisika Menggunakan Three-Tier Diagnostic Test Pada Pokok Bahasan Kinematika Gerak.” *SPEKTRA: Jurnal Kajian Pendidikan Sains* 3, no. 2 (2017): 196. <https://doi.org/10.32699/spektra.v3i2.38>.
- Maison, Maison, Dwi Agus Kurniawan, and Retno Sari Widowati. “The Quality of Four-Tier Diagnostic Test Misconception Instrument for Parabolic Motion.” *Jurnal Pendidikan Dan Pengajaran* 54, no. 2 (2021): 359. <https://doi.org/10.23887/jpp.v54i2.35261>.
- Mathoriyah, Lailatul, and Afif Kholisun Nashoih. “Taṭwīr Adawāt Taqwīm Ta’līm Al Balāḡah ‘Abra Minaṣṣah Wordwall Fī Dauī Madkhal Al-Ikhtibār at-Tasykhīsyī Li Arba’Ati Mustawayāt.” *Journal of Arabic Education and Literature* 8, no. 2 (2024): 227–56.
- Meila, Nafisa, Fakhrudin, and Islami Nur. “Pengembangan Tes Diagnostik Four Tier Untuk Menganalisis Konsepsi Peserta Didik SMP Pada Materi Listrik Dinamis Berdasarkan Rasch Model.” *Jurnal Pendidikan Dan Pengajaran* 2, no. 3 (2024): 331–37.
- Mohammadi, Mojtaba, Maryam Zarrabi, and Jaber Kamali. “Formative Assessment Feedback to Enhance the Writing Performance of Iranian IELTS Candidates: Blending Teacher and

- Automated Writing Evaluation.” *International Journal of Language Testing* 13, no. 1 (2023): 206–24. <https://doi.org/10.22034/ijlt.2022.364072.1201>.
- Mukhlisa, Nurul. “Miskonsepsi Pada Peserta Didik.” *SPEED Journal: Journal of Special Education* 4, no. 2 (2021): 66–76. <https://doi.org/10.31537/speed.v4i2.403>.
- Nashoih, Afif Kholisun. “HOTS-Based Qiro’ah Textbook Development to Enhance Arabic Learners’ Extensive Reading Skills.” *DINAMIKA: Jurnal Kajian Pendidikan Dan Keislaman* 09, no. 02 (2024): 24–42. <https://doi.org/10.32764/dinamika.v9i2.4751>.
- Nashoih, Afif Kholisun, and M Faridl Darmawan. “Pengembangan Bahan Ajar Nahwu Berbasis Kontrastif Untuk Mengatasi Interferensi Bahasa Indonesia Terhadap Bahasa Arab.” *Arabiyatuna: Jurnal Bahasa Arab* 3, no. 2 (2019): 335–50. <https://doi.org/10.29240/jba.v3i2.1008>.
- Nasyidiah, Funky Iqlima, Parsaoran Siahaan, and Dedi Sasmita. “Pengembangan Instrumen Four-Tier Diagnostic Test Untuk Mendeteksi Miskonsepsi Siswa Kelas X Pada Materi Impuls.” *WaPFI (Wahana Pendidikan Fisika)* 5, no. 2 (2020): 31–40. <https://doi.org/10.17509/wapfi.v5i2.27156>.
- Natalia, Defrina, Jeffry Handhika, and Farida Huriawati. “Pengembangan Instrumen Tes Diagnosis Konsep IPA Fisika.” *Momentum: Physics Education Journal* 1, no. 2 (2017): 103. <https://doi.org/10.21067/mpej.v1i2.1966>.
- Nurbayan, Dedeng Rosyidin, Yayan Nurbayan, and Kurnia Nurul Falah. “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, no. 2 (2020): 274–90. <https://doi.org/10.18860/ijazarabi.v3i2.8293>.
- Nuryanto, Ahmad Rifa’i, and Abbas Sofwan Matlail Fajar. “The Development of Nahwu Teaching Materials Based on Mind Maps Using the ADDIE Model” 1, no. 1 (2024): 35–47. <https://creativecommons.org/licenses/by-sa/4.0/>.
- Qomari, Nur, Khairul Mizan, Fatimatuz Zahro, Bakri Mohammed Bakheet Ahmed, and Khusnul Khotimah. “Evaluation of Concept Maps Strategy in Learning Al-Qowa’id Al-Nahwiyyah 1 Taqvim Istiratijiyah Kharait Al-Mafahim Fi Ta’lim Al-Qawa’id Al-Nahwiyyah.” *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab* 16, no. 1 (2024): 125. <https://doi.org/10.24042/albayan.v16i1.19286>.
- Rawh, Pujia, Achmad Samsudin, and Muhamad Gina Nugraha. “Pengembangan Four-Tier Diagnostic Test Untuk Mengidentifikasi Profil Konsepsi Siswa Pada Materi Alat-Alat Optik.” *WaPFI (Wahana Pendidikan Fisika)* 5, no. 1 (2020): 84–89. <https://doi.org/10.17509/wapfi.v5i1.22888>.
- Ritonga, Pangoloan Soleman, and Arif Yasthophi. “Pengembangan Instrumen Test Diagnostik Multiple Choice Four Tier Pada Materi Ikatan Kimia.” *Konfigurasi : Jurnal Pendidikan Kimia Dan Terapan* 3, no. 1 (2019): 23–31. <https://doi.org/10.24014/konfigurasi.v3i1.6797>.
- Roghdah, Sanaa Jauza, Muhammad Zammi, and Julia Mardhiya. “Pengembangan Four-Tier Multiple Choice Diagnostic Test Untuk Mengetahui Tingkat Pemahaman Konsep Peserta Didik Pada Materi Termokimia.” *Phenomenon: Jurnal Pendidikan MIPA* 11, no. 1 (2021): 57–74. <https://doi.org/10.21580/phen.2021.11.1.8573>.
- Sarsarabi, Sharareh Sadat, and Zeinab Sazegar. “Investigating Different Kinds of Stems in Multiple-Choice Tests: Interruptive Vs. Cumulative.” *International Journal of Language Testing* 13, no. 2 (2023): 170–87. <https://doi.org/10.22034/IJLT.2023.392953.1250>.
- Sri Rahmadani Pulu, and Abd. Haji Amahoru. “Analisis Miskonsepsi Mahasiswa Pada Pembelajaran IPA Menggunakan Tes Diagnostik Multiple Choice Berbantuan CRI

- (Certainty of Response Index).” *Jurnal Pendidikan Mipa* 13, no. 2 (2023): 478–86. <https://doi.org/10.37630/jpm.v13i2.1039>.
- Sugiyono. *Metode Penelitian Kuantitatif, Kualitatif, Dan R&D*. 19th ed. Vol. 7. Bandung: CV. Alfabeta, 2013.
- Tiruchttampalam, Shanthi, and Alistair Ross. “Validation of a Screening Tool for Dyslexia in English among Arabic-Speaking University Students in the Gulf Area.” *International Journal of Arabic-English Studies* 24, no. 2 (2024): 231–48. <https://doi.org/10.33806/ijaes.v24i2.662>.
- Torang Siregar. “Stages of Research and Development Model Research and Development (R&D).” *DIROSAT: Journal of Education, Social Sciences & Humanities* 1, no. 4 (2023): 142–58. <https://doi.org/10.58355/dirosat.v1i4.48>.
- Yasthophi, Arif, and Pangoloan Soleman Ritonga. “Pengembangan Instrumen Test Diagnostik Multiple Choice Four Tier Pada Materi Ikatan Kimia.” *Konfigurasi : Jurnal Pendidikan Kimia Dan Terapan* 3, no. 1 (February 2019): 23. <https://doi.org/10.24014/konfigurasi.v3i1.6797>.
- Yuberti, Yuberti, Yani Suryani, and Indah Kurniawati. “Four Tier Diagnostic Test With Certainty of Response Index to Identify Misconception in Physics.” *Indonesian Journal of Science and Mathematics Education* 3, no. 2 (2020): 245–53. <https://doi.org/10.24042/ijsme.v3i2.6061>.
- Yulianawati, Dewi, Asih Wahyuningsih, and Nur Aisa Pebriana. “Pengembangan Instrumen 4TSDT (Four Tier – Science Diagnostic Test) Untuk Mengidentifikasi Level Konsepsi Calon Guru Sekolah Dasar.” *Jurnal Basicedu* 6, no. 6 (2022): 9483–90. <https://doi.org/10.31004/basicedu.v6i6.4117>.
- Zaleha, Zaleha, Achmad Samsudin, and Muhamad Gina Nugraha. “Pengembangan Instrumen Tes Diagnostik VCCI Bentuk Four-Tier Test Pada Konsep Getaran.” *Jurnal Pendidikan Fisika Dan Keilmuan (JPFK)* 3, no. 1 (2017): 36. <https://doi.org/10.25273/jpfb.v3i1.980>.