

الباب الأول

المقدمة

أ. خلفية البحث

يعد التقييم جزءاً لا يتجزأ من عملية التعليم والتعلم، إذ يحصل المعلم من خلاله على معلومات حول مدى تحقيق أهداف التعليم، كما يستفيد منه في اتخاذ القرارات المتعلقة بالإجراءات اللاحقة وتحسين العملية التعليمية. وبذلك، لا يقتصر دور التقييم على منح الدرجات في نهاية العملية التعليمية فحسب، بل يعدّ أيضاً وسيلة للحصول على معلومات حول تنفيذ عملية التعليم ونتائجها. وفي تعليم اللغة العربية، يحتل التقييم مكانة مهمة؛ لأنه يمثل أحد العناصر الأساسية التي تسهم في تحقيق نجاح العملية التعليمية بصورة عامة¹.

تحتاج عملية التقييم إلى أداة يمكنها تقديم المعلومات المناسبة وفقاً للجوانب المراد معرفتها. ويعد الاختبار أحد الأدوات التي تُستخدم على نطاق واسع في تقييم العملية التعليمية. فمن خلال الاختبار، يقدم المتعلمون أو طلاب المعهد استجاباتٍ تجاه مجموعة من فقرات الأسئلة، ثم تُستخدم هذه الاستجابات مصدراً للمعلومات في عملية التقييم. ولذلك، فإن إعداد الاختبار لا يقتصر على صياغة الأسئلة اعتماداً على المادة التعليمية التي تم تدريسها فقط، بل يحتاج إلى عملية منظمة تبدأ من تحديد الأهداف، واختيار المحتوى التعليمي، وإعداد المؤشرات ومخطط الاختبار، وتطوير فقرات الأسئلة، ومراجعتها، حتى الوصول إلى تحسين الأداة الاختبارية².

¹ Bamualim, "Kedudukan Dan Tujuan Evaluasi Pembelajaran Bahasa Arab."

² Sihabuddin, "Prosedur Penyusunan Tes Berbasis Hots Pada Empat Keterampilan Berbahasa Arab."

إن اختلاف أدوات الاختبار يتيح أن تكون لكل أداة خصائص مختلفة، سواء من حيث متطلبات العمليات المعرفية أم درجة صعوبة الفقرات. ولا تقدم وثائق الأسئلة، ومفاتيح الإجابة، وسجلات استجابات طلاب المعهد المتاحة معلومات مباشرة عن توزيع المستويات المعرفية ودرجة صعوبة كل فقرة قبل إجراء التحليل. ولذلك، لا يمكن معرفة خصائص الأداة التي تم استخدامها اعتمادًا على شكل الأسئلة أو النتيجة النهائية للاختبار فقط، بل يتطلب ذلك تحليل فقرات الاختبار واستجابات طُلابِ المعهد.

ويُعد تحليل هاتين الخاصيتين ضروريًا للحصول على صورة أكثر منهجية عن أدوات الاختبار النهائي للغة العربية التي تم استخدامها. فتحليل المستوى المعرفي يوفر معلومات عن عمليات التفكير التي تتطلبها كل فقرة، في حين يمكن استخدام استجابات طلاب المعهد لتحديد درجة صعوبتها. ولا يهدف هذا البحث إلى تقويم جودة المعلمين أو مقارنة قدرات طلاب المعهد في اللغة العربية بين الفصول، وإنما يهدف إلى تقويم خصائص أدوات الاختبار بوصفها أساسًا للنظر في تطوير الاختبار النهائي للغة العربية في الفترات اللاحقة.

ومن الأطر التي يمكن استخدامها لتحديد عمليات التفكير هذه تصنيف بلوم المُنقَّح الذي قام أندرسون وكراثول بتطويره. ويميز تصنيف بلوم المُنقَّح العمليات المعرفية إلى ستة مستويات، وهي: التذكر، والفهم، والتطبيق، والتحليل، والتقويم، والإبداع. وفي هذه الدراسة، تُستخدم هذه المستويات لتحديد التصنيف المعرفي، وهي C1: التذكر، و C2: الفهم، و C3: التطبيق، و C4: التحليل، و C5: التقويم، و C6: الإبداع³.

يحتاج استخدام تصنيف بلوم المُنقَّح في تحليل الأسئلة إلى الدقة والعناية. ولا ينبغي أن يعتمد تحديد مستويات C1 إلى C6 على الأفعال الواردة في صياغة السؤال فقط.

³ Krathwohl, "A Revision Of Bloom's Taxonomy."

فقد أوضح Stanny أن الأفعال المستخدمة في الأهداف أو المهام لا تعكس بالضرورة بمفردها درجة تعقيد عملية التفكير المطلوبة. كما أشار Newton وزملاؤه إلى وجود تنوع في استخدام الأفعال لتمثيل فئات تصنيف بلوم. ولذلك، فإن تحليل المستوى المعرفي يحتاج إلى مراعاة المثير، وصياغة السؤال، والاختيارات المتاحة للإجابة، والإجابة الصحيحة، وسياق السؤال، بالإضافة إلى خطوات التفكير التي يحتاج إليها المتعلم للوصول إلى الإجابة⁴.

وينطبق هذا المبدأ أيضًا على أسئلة الاختيار من متعدد. إذ إن صيغة الاختيار من متعدد لا تدلّ تلقائيًا على أن السؤال يقيس قدرة التذكر فقط. ويمكن لأسئلة الاختيار من متعدد أن تتطلب عمليات تفكير أكثر تعقيدًا إذا تم إعداد المثيرات، والأسئلة، والبدائل بطريقة مناسبة⁵. وبناءً على ذلك، فإن تحديد المستوى المعرفي لا يعتمد على شكل السؤال، وإنما يعتمد على مستوى عمليات التفكير المطلوبة للإجابة عن كل فقرة من فقرات الاختبار.

وفي تعليم اللغة العربية، يمكن أن تظهر تنوعات في عمليات التفكير ضمن موضوعات مختلفة. فالسؤال الذي يطلب من طلاب المعهد التعرف على معاني المفردات التي سبق لهم تعلمها قد يتطلب عملية التذكر. أما السؤال الذي يطلب منهم تفسير المعلومات الواردة في جملة أو نص معين فقد يتطلب عملية الفهم. وإذا كان على طلاب المعهد استخدام قواعد النحو أو الصرف في سياق معين، فإن السؤال قد يتطلب عملية التطبيق. بينما عندما يُطلب منهم تحليل التراكيب، أو تمييز العلاقات بين الأجزاء، أو

⁴ Newton Et Al., "A Pragmatic Master List Of Action Verbs For Bloom's Taxonomy."

⁵ Scully, *Constructing Multiple-Choice Items To Measure Higher-Order Thinking*.

معالجة مجموعة من المعلومات للوصول إلى الإجابة، فإن متطلبات السؤال قد تتجه إلى عملية التحليل.

إن وجود أسئلة في مستوى C1 و C2 ضمن تقويم اللغة العربية لا يدل بالضرورة على أن الأداة الاختبارية غير مناسبة للطلاب. فعمليتا التذكر والفهم تظان جزءًا من أبعاد عمليات التفكير. كما أن توزيع مستويات التفكير من C1 إلى C6 لا يشترط أن يكون متساويًا، لأن تحديد استخدام هذه المستويات ينبغي أن يتناسب مع أهداف التقويم، وخصائص المادة التعليمية، والقدرات المستهدفة، وخصائص طُلابِ المَعْهَدِ. والأهم من ذلك أن تكون متطلبات عمليات التفكير في الأداة الاختبارية محددة ومنظمة ومدروسة بصورة مخططة^٦.

إضافةً إلى الخصائص المعرفية، يمكن أيضًا تحليل فقرات الاختبار من خلال درجة صعوبتها. وتوفر درجة الصعوبة معلومات حول كيفية استجابة المشاركين لكل فقرة من فقرات الاختبار. ويمكن مقارنة استجابات المشاركين في أسئلة الاختيار من متعدد بمفتاح الإجابة؛ لمعرفة عدد المشاركين الذين أجابوا إجابة صحيحة في كل فقرة. ثم تُستخدم هذه البيانات في تحليل فقرات الاختبار، بما في ذلك الحصول على معلومات حول درجة صعوبة الأسئلة^٧.

يوفر المستوى المعرفي ودرجة الصعوبة معلومات مختلفة حول كل فقرة من فقرات الاختبار. فالمستوى المعرفي يُحدّد من خلال تحليل عمليات التفكير التي يتطلبها السؤال، في حين تُحدّد درجة الصعوبة بناءً على استجابات المشاركين تجاه تلك الفقرة. ولذلك، فإن

⁶ Anderson, *A Taxonomy For Learning, Teaching, And Assessing*.

⁷ Fikriyah, "Analisis Butir Soal Ulangan Tengah Semester Mata Pelajaran Bahasa Arab Kelas Vii Semester Ganjil Smp Muhammadiyah 1 Yogyakarta Tahun Ajaran 2019/2020."

تصنيف السؤال ضمن مستويات C1 أو C2 أو C3 أو C4 أو C5 أو C6 لا يمكن استخدامه بشكل مباشر لتحديد ما إذا كان السؤال سهلاً أو متوسطاً أو صعباً. فقد يكون السؤال الذي يتطلب عملية التذكر فقط غير قابل للإجابة الصحيحة من قبل معظم المشاركين، كما أن السؤال الذي يتطلب عملية التطبيق أو التحليل لا يكون بالضرورة دائماً من أكثر الأسئلة صعوبة.

إن تحليل هاتين الخاصيتين يمكن أن يوفر معلومات أكثر شمولاً حول الأداة الاختبارية. فتحليل المستوى المعرفي يوضح نوع عملية التفكير التي يتطلبها السؤال، بينما يوضح تحليل درجة الصعوبة كيفية استجابة المشاركين لذلك السؤال. وبذلك، لا يمكن النظر إلى الأداة الاختبارية من خلال توزيع مستويات C1 إلى C6 فقط، بل يجب أيضاً النظر إليها من خلال خصائص درجة الصعوبة في كل فقرة من فقرات الاختبار. وتستخدم بيانات استجابات المشاركين في أغراض متعددة ضمن تحليل فقرات الاختبار، ومنها تحديد درجة الصعوبة، ولذلك ينبغي الاستفادة منها بما يتناسب مع أهداف التقييم المحددة⁸.

تُعَدُّ دراسة الخصائص المعرفية ودرجة الصعوبة أمراً مهماً في سياق تعليم اللغة العربية في معهد الجامعة بجامعة سونان غونونغ جاتي الإسلامية الحكومية باندونغ. ويعد تعليم اللغة العربية أحد الأنشطة الأكاديمية التي تُنفَّذ في المعهد. وتشير المعلومات الصادرة عن جامعة سونان غونونغ جاتي الإسلامية الحكومية باندونغ إلى أن المعهد يستخدم اختبار تحديد

⁸ Fikriyah, "Analisis Butir Soal Ulangan Tengah Semester Mata Pelajaran Bahasa Arab Kelas Vii Semester Ganjil Smp Muhammadiyah 1 Yogyakarta Tahun Ajaran 2019/2020."

المستوى (*placement test*) بوصفه جزءًا من عملية تحديد فصول الطلاب الجدد في

المعهد⁹. كما يُنفَّذ تعليم اللغة العربية بصورة منتظمة باعتباره جزءًا من أنشطة المعهد¹⁰

ويرتبط استخدام اختبار تحديد المستوى بتنوع الخلفيات التعليمية الأولية لدى طلاب المعهد في تعلم اللغة العربية. ويُستخدم تحديد الفصل بناءً على نتائج اختبار تحديد المستوى لتقسيم طلاب المعهد قبل مشاركتهم في الأنشطة التعليمية. ومع ذلك، فإن نتائج اختبار تحديد المستوى في هذه الدراسة لا تُستخدم بوصفها مقياسًا للقدرة النهائية لطلاب المعهد، ولا تُستخدم أساسًا للمقارنة بين قدرات الفصول المختلفة. وإنما يُنظر إلى موقع الفصل على أنه جزء من سياق تحديد مصدر بيانات البحث.

وفي تعليم اللغة العربية في معهد الجامعة، يعد كتاب العربية للناشئين أحد المصادر التعليمية المستخدمة في عملية التعلم. ويُستخدم هذا الكتاب بوصفه أحد مصادر تعليم اللغة العربية للمتعلمين غير الناطقين بها، كما يتضمن موضوعات ومحتويات لغوية تدعم عملية التعليم¹¹. وفي هذه الدراسة، يُوضَع هذا الكتاب في إطار سياق مصدر المادة التعليمية فقط، وليس بوصفه موضوعًا لتحليل مدى توافق محتواه مع أسئلة الاختبار النهائي.

بناءً على وثائق البحث، يُنفَّذ تعليم اللغة العربية في معهد الجامعة في ثلاث عشرة شعبةً دراسية، تتكون من عشر شعب للطالبات وثلاث شعب للطلاب. ويُدرّس كل شعبة معلّم مختلف، كما أُسند إعداد أداة الاختبار النهائي للغة العربية إلى المعلّم في كل شعبة.

⁹ Uin Sunan Gunung Djati Bandung, "Pendaftaran Santri Ma'had Al-Jami'ah Telah Dibuka."

¹⁰ Uin Sunan Gunung Djati Bandung, "Yuk Gunakan Kesempatan Emas Untuk Gabung Bersama Ma'had Al-Jami'ah: Ini Cara Daftar Dan Fasilitasnya."

¹¹ Roviin, "Analisis Buku Teks Al 'Arabiyah Li Al Nasyi'in Karya Mahmud Ismail Shini, Dkk."

ونتيجةً لذلك، تختلف أدوات الاختبار المستخدمة بين الشعب الدراسية. وقد جاءت جميع اختبارات النهاية التي تمثل موضوع هذه الدراسة في صورة أسئلة اختيار من متعدد، وتم تنفيذها من خلال تطبيق Google Forms.

ويؤدي اختلاف أدوات الاختبار إلى إمكانية ظهور تنوع في خصائص كل أداة، سواء من حيث توزيع عمليات التفكير أو من حيث درجة صعوبة فقرات الاختبار. ولذلك، تحتاج كل أداة إلى التحليل وفقاً لخصائصها الخاصة. ولا تهدف هذه الدراسة إلى تقييم جودة المعلمين الذين قاموا بإعداد الأسئلة، ولا إلى تحديد أيّ الفصول يمتلك مستوى أعلى في القدرة على اللغة العربية. وإنما يتركز التحليل على الأداة المستخدمة، وذلك من خلال دراسة متطلبات عمليات التفكير في كل فقرة، ودرجة الصعوبة بناءً على استجابات طلاب المعهد الذين أجابوا عن تلك الأداة. وتُعدّ البيانات المتاحة من خلال Google Forms مهمة؛ لأنها لا توفر معلومات حول اختيارات إجابات المشاركين فقط، بل تتيح أيضاً معرفة عدد الإجابات الصحيحة في كل فقرة. وتُستخدم استجابات طلاب المعهد لتحديد درجة صعوبة كل فقرة.

وقد أُجريت دراسات حول المستوى المعرفي لأسئلة اللغة العربية في سياقات مختلفة. فعلى سبيل المثال، قامت فوزية وزملاؤها بتحليل الاختبار النهائي للفصل الدراسي في اللغة العربية من جوانب متعددة، منها المستوى المعرفي ودرجة الصعوبة. ومن بين أربعين فقرة تم تحليلها، وُجد سؤالان في مستوى C1، وسؤالان في مستوى C2، وستة وثلاثون سؤالاً في مستوى C3، بينما لم توجد أسئلة في مستويات C4 إلى C6. وأظهرت هذه الدراسة أن أداة اختبار اللغة العربية قد تركز في مستوى معرفي معين، كما يمكن تحليلها من خلال خصائص الفقرات الأخرى.

كما أوضحت دراسة أوتامي أن المستوى المعرفي ودرجة الصعوبة يمكن تحليلهما معًا في دراسة واحدة حول أداة اختبار اللغة العربية. وقد استخدمت هذه الدراسة أسئلة اختبار TOAFL وأوراق إجابات المشاركين لتحليل الصدق، ودرجة الصعوبة، والمستوى المعرفي وفقًا لتصنيف بلوم. وأظهرت النتائج أن درجة الصعوبة كانت بشكل عام في المستوى المتوسط، وأن التوزيع المعرفي اشتمل على المستويين C1 و C4 وفقًا للنسب التي ذكرتها الدراسة^{١٢}.

في حين ركزت دراسات أخرى بشكل أكبر على الخصائص المعرفية. فقد استخدم قدري تصنيف بلوم المنقّح في تحليل أسئلة امتحان UAMBN للغة العربية، ووجد أن توزيع عناصر اللغة، ومهارات اللغة، والفئات المعرفية لم يُظهر توازنًا كافيًا^{١٣}. كما وجدت سلسبيل أن من بين ٢٨ فقرة من أسئلة التقويم النهائي للغة العربية التي تم تحليلها، كانت هناك فقرة واحدة فقط في مستوى C4، ولم توجد أسئلة في مستويي C5 و C6^{١٤}. كما توصلت زلفى وزملاؤها إلى وجود تنوع في توزيع مستويات MOTS و LOTS و HOTS في أداة اختبار اللغة العربية التي تم تحليلها^{١٥}.

في المقابل، وجدت دراسة أسرور نتائج مختلفة في أسئلة الأولمبياد الوطني للغة العربية. فمن بين ١٠٠ فقرة من أسئلة الاختيار من متعدد التي تم تحليلها، كانت نسبة ٥٣٪ منها ضمن مجموعة C1-C3، بينما كانت نسبة ٤٧٪ ضمن مجموعة C4

تحليل بنود أسئلة اختبارات اللغة العربية لغير الناطقين بها على ضوء نظرية تصنيف بلوم: دراسة وصفية تحليلية " Utami, 12
"في الجامعة الإسلامية الحكومية فونوروكو"

13 Qodri Et Al., "The Suitability Of Arabic Language Question Item In The Perspective Of Bloom Taxonomy Revision L Tanasub Bunudi As'ilah Al-Lughah Al-'Arabiyyah Fi Dhau Tashnif Bloom Al-Mu'addal."

14 Salsabila, "المدرسة الثانوية الحكومية (Hots) دراسة التقويم النهائي لتعلم اللغة العربية القائم على مهارات التفكير العليا"، ١١ باندونج

15 Zulfa Et Al., "Analisis Butir Soal Bahasa Arab Perspektif Anderson Dan Krathwohl."

C6 وفقًا لتصنيف الدراسة¹⁶. ويشير اختلاف نتائج الدراسات السابقة إلى أن خصائص المستوى المعرفي لا يمكن تحديدها اعتمادًا على شكل السؤال أو المادة الدراسية فقط، بل تحتاج إلى تحليل مباشر للأداة المستخدمة.

وتوضح هذه الدراسات أن استخدام تصنيف بلوم في تحليل أدوات اختبار اللغة العربية ليس أمرًا جديدًا. كما أن تحليل درجة الصعوبة قد استخدم أيضًا في دراسات أدوات اختبار اللغة العربية، سواء بشكل مستقل أو مع خصائص أخرى لل فقرات. ولذلك، فإن موقع هذه الدراسة لا يقوم على الادعاء بأن الجمع بين المستوى المعرفي ودرجة الصعوبة لم يسبق دراسته.

استنادًا إلى مراجعة تسع دراسات سابقة، يتضح أن استخدام تصنيف بلوم المُنقَّح وتحليل درجة الصعوبة ليسا أمرين جديدين في بحوث أدوات اختبار اللغة العربية. وبالمقارنة مع تلك الدراسات، يركّز هذا البحث بصورة خاصة على تقويم عدة أدوات مختلفة للاختبار النهائي للغة العربية داخل مؤسسة واحدة، وذلك من خلال اعتبار كل فقرة وحدةً للتحليل، ثم مقارنة المستوى المعرفي ودرجة الصعوبة في الفقرة نفسها؛ بهدف صياغة الآثار المترتبة على تطوير أداة الاختبار النهائي للغة العربية.

وبناءً على العرض السابق، أُجريت هذه الدراسة بعنوان "تقويم المستوى المعرفي ودرجة صعوبة فقرات أسئلة الاختبار النهائي للغة العربية في معهد الجامعة بجامعة سونان غونونغ جاتي الإسلامية الحكومية باندونغ".

¹⁶ Asror Et Al., "Cognitive Level Mapping Of Indonesian Arabic Olympiad Questions."

ب. أسئلة البحث

بناءً على خلفية البحث السابقة، يمكن صياغة مشكلات البحث على النحو الآتي:

١. كيف يكون توزيع المستوى المعرفي في أداة أسئلة الاختبار النهائي للغة العربية في معهد الجامعة بجامعة سونان غونونغ جاتي الإسلامية الحكومية باندونغ؟
٢. ما درجة صعوبة فقرات أسئلة الاختبار النهائي للغة العربية بناءً على استجابات طلاب المعهد في كل فئة من فئات المستوى المعرفي؟
٣. ما دلالات نتائج تقويم المستوى المعرفي ودرجة الصعوبة في تطوير أداة الاختبار النهائي للغة العربية في معهد الجامعة بجامعة سونان غونونغ جاتي الإسلامية الحكومية باندونغ؟

ج. أهداف البحث

- بناءً على أسئلة البحث السابقة، يهدف هذا البحث إلى ما يأتي:
١. وصف توزيع المستوى المعرفي في أداة أسئلة الاختبار النهائي للغة العربية في معهد الجامعة بجامعة سونان غونونغ جاتي الإسلامية الحكومية باندونغ.
 ٢. وصف درجة صعوبة فقرات أسئلة الاختبار النهائي للغة العربية بناءً على استجابات طلاب المعهد في كل فئة من فئات المستوى المعرفي.
 ٣. بيان دلالات نتائج تقويم المستوى المعرفي ودرجة الصعوبة في تطوير أداة الاختبار النهائي للغة العربية في معهد الجامعة بجامعة سونان غونونغ جاتي الإسلامية الحكومية باندونغ.

د. فوائد نتائج البحث

١. الفوائد النظرية

من الناحية النظرية، يُتوقع أن يسهم هذا البحث في إضافة فهم تجريبي وتحليلي لخصائص أدوات تقييم تعليم اللغة العربية، ولا سيما من حيث المستوى المعرفي ودرجة صعوبة فقرات الاختبار في سياق الاختبار النهائي للغة العربية في معهد الجامعة بجامعة سونان غونونغ جاتي الإسلامية الحكومية باندونغ. كما يقدم هذا البحث تصورًا عن تطبيق تصنيف بلوم المُنقَّح في تحديد متطلبات العمليات المعرفية في كل فقرة، والاستفادة من استجابات طلاب المعهد للحصول على معلومات حول درجة صعوبة الأسئلة.

ويُتوقع أيضًا أن يعزز هذا البحث الفهم القائل إن تحديد المستوى المعرفي للسؤال لا يكفي أن يعتمد على الفعل الإجرائي فقط، بل ينبغي أن يراعي البناء الكامل للفقرة وعملية التفكير اللازمة للوصول إلى الإجابة. بالإضافة إلى ذلك، يوضح هذا البحث أن المستوى المعرفي ودرجة الصعوبة يمثلان خاصيتين مختلفتين، إلا أنه يمكن استخدامهما معًا في تقييم أداة الاختبار. فالمستوى المعرفي يقدم معلومات عن متطلبات عملية التفكير في السؤال، بينما تقدم درجة الصعوبة معلومات تجريبية مستندة إلى استجابات طلاب المعهد عن كل فقرة.

٢. الفوائد التطبيقية

بالنسبة لمعهد الجامعة، يُتوقع أن توفر نتائج هذا البحث معلومات حول خصائص أداة الاختبار النهائي للغة العربية بناءً على توزيع المستوى المعرفي ودرجة صعوبة الأسئلة.

ويمكن استخدام هذه المعلومات بوصفها أحد المصادر للتقويم في تطوير نظام إعداد أدوات الاختبار النهائي في المستقبل.

وبالنسبة لمعلمي اللغة العربية، يُتوقع أن تكون نتائج هذا البحث مادةً للتأمل في إعداد فقرات الاختبار ومراجعتها وتطويرها. ويمكن استخدام المعلومات المتعلقة بتوزيع مستويات C1 إلى C6 للنظر في تنوع متطلبات عمليات التفكير، بينما يمكن الاستفادة من معلومات درجة الصعوبة في مراجعة الفقرات المستخدمة في الاختبار.

وبالنسبة للباحثين في المستقبل، يُتوقع أن تكون نتائج هذا البحث أحد المراجع في الدراسات المتعلقة بتقويم أدوات اختبار اللغة العربية، وخاصة الدراسات التي تربط بين تحليل محتوى المستوى المعرفي وتحليل استجابات المشاركين حول خصائص الفقرات.

هـ. الإطار الفكري

ينطلق هذا البحث من فكرة أن الاختبار النهائي للغة العربية يعد أحد أشكال التقويم الذي يستخدم أداة الاختبار للحصول على معلومات حول عملية التعليم. ولا تشتمل كل أداة اختبارية على المادة التي يتم اختبارها فقط، بل تمتلك أيضًا خصائص معينة. وفي هذا البحث، تتمثل الخصائص التي تحظى بالاهتمام في المستوى المعرفي ودرجة صعوبة فقرات الاختبار.

ترتبط الخصائص المعرفية بعمليات التفكير التي يحتاج إليها طُلابُ المَعَهَدِ للوصول إلى الإجابة الصحيحة. ويتم تحليل هذه العمليات باستخدام تصنيف بلوم المُنَفَّح لأندرسون وكراثوول، والذي يشتمل على C1: التذكر، و C2: الفهم، و C3: التطبيق، و C4: التحليل، و C5: التقويم، و C6: الإبداع. ويتم تحديد هذه الفئات من خلال تحليل محتوى

كل فقرة من فقرات الاختبار، مع مراعاة المثير، وصياغة السؤال، والبدايل، والإجابة الصحيحة، والسياق، وعمليات التفكير المطلوبة للوصول إلى الإجابة.

وبعد حصول كل فقرة على تصنيف المستوى المعرفي، يتم حساب عدد الأسئلة في كل مستوى، بحيث يتم الحصول على توزيع مستويات C1 إلى C6 في كل أداة اختبارية. ويُستخدم هذا التوزيع للإجابة عن مشكلة البحث الأولى، كما يقدم تصورًا حول تنوع عمليات التفكير التي تتطلبها أدوات الاختبار النهائي للغة العربية.

وفي المرحلة التالية، تُستخدم بيانات استجابات طلاب المعهد المسجلة من خلال Google Forms. حيث تتم مقارنة إجابات المشاركين في كل فقرة مع مفتاح الإجابة، للحصول على عدد المشاركين الذين أجابوا إجابة صحيحة. ثم تُستخدم هذه البيانات لتحديد درجة صعوبة كل فقرة وفقًا لإجراءات تحليل درجة الصعوبة المعتمدة في البحث.

ويُجرى تحليل درجة الصعوبة لكل فصل بشكل منفصل، لأن فصول الطالبات (A)، F، I وفصل الطلاب (B) تستخدم أدوات اختبار نهائي مختلفة. ولذلك، لا تُستخدم نتائج درجة الصعوبة للحكم بأن أحد الفصول يمتلك قدرة أعلى أو أقل في اللغة العربية مقارنة بالفصول الأخرى. وإنما تُفسّر درجة الصعوبة بوصفها خاصية من خصائص الفقرة في سياق مجموعة المشاركين الذين أدوا الاختبار.

بعد ذلك، تتم مقارنة نتائج تحليل المستوى المعرفي ودرجة الصعوبة. وتتيح هذه المقارنة للباحث معرفة كيفية ظهور خصائص الصعوبة في الأسئلة التي تنتمي إلى مستويات C1 أو C2 أو C3 أو C4 أو C5 أو C6. ولا يهدف هذا التحليل إلى إثبات أن

المستوى المعرفي يحدد درجة الصعوبة، وإنما يهدف إلى تقديم صورة أكثر شمولاً حول خصائص كل أداة اختبارية.

بعد ذلك، يتم تفسير نتائج تقويم توزيع المستوى المعرفي ودرجة الصعوبة بالاعتماد على النظريات والدراسات السابقة. وتُستخدم هذه النتائج أساساً لصياغة الدلالات المتعلقة بتطوير أدوات الاختبار النهائي للغة العربية، مثل تخطيط توزيع المستويات المعرفية من خلال جدول المواصفات، وإعادة مراجعة الفقرات بناءً على نتائج استخدامها، وتعديل الأسئلة، وتطوير الأدوات الاختبارية بصورة أكثر تنظيمًا وتخطيطًا.

و. الدراسات السابقة

أُجريت العديد من الدراسات حول تقويم أدوات اختبار اللغة العربية من قبل عدد من الباحثين مع اختلاف مجالات التركيز وموضوعات البحث. فقد استخدمت بعض الدراسات تصنيف بلوم لتحليل المستوى المعرفي للأسئلة، بينما جمعت بعض الدراسات الأخرى بينه وبين تحليل خصائص الفقرات، مثل درجة الصعوبة، والصدق، والثبات، في حين ركزت دراسات أخرى بشكل خاص على عمليات التفكير العليا. وتُستخدم هذه الدراسات السابقة لتحديد موقع هذا البحث بين الدراسات الموجودة.

قامت أوتامي من خلال رسالتها العلمية حول تحليل أسئلة اختبار اللغة العربية لغير الناطقين بها في الجامعة الإسلامية الحكومية فونوروكو بدراسة الصدق، ودرجة الصعوبة، والمستوى المعرفي وفقًا لتصنيف بلوم، باستخدام أسئلة الاختبار وأوراق إجابات المشاركين. وأظهرت نتائج الدراسة أن مستوى الصدق كان مرتفعًا، وأن درجة الصعوبة كانت بشكل عام في المستوى المتوسط، بينما كان توزيع المستوى المعرفي يتكون من C1 بنسبة ٥٠٪.

و C4 بنسبة ٥٠٪. وترتبط هذه الدراسة بالبحث الحالي من حيث استخدام أداة اختبار اللغة العربية واستجابات المشاركين، وكذلك الاهتمام بالمستوى المعرفي ودرجة الصعوبة. إلا أن الاختلاف يكمن في موضوع الدراسة ونطاقها، حيث يحلل هذا البحث عدة أدوات للاختبار النهائي أعدّها معلّمون مختلفون، ولا يهدف إلى اختبار الصدق^{١٧}.

وقام فوزية وشهاب الدين وسوفيان من خلال بحثهم بعنوان تحليل جودة اختبار اللغة العربية القائم على مهارات التفكير العليا (HOTS) بتحليل الاختبار النهائي للفصل الدراسي في اللغة العربية من جوانب متعددة، تشمل الصدق، والثبات، وقواعد أسئلة الاختيار من متعدد، ودرجة الصعوبة، والقدرة على التمييز، وفعالية المشتتات، بالإضافة إلى المستوى المعرفي. ومن بين ٤٠ فقرة تم تحليلها، وُجدت فقرتان في مستوى C1، وفقرتان في مستوى C2، و ٣٦ فقرة في مستوى C3، بينما لم توجد فقرات في مستويات C4 إلى C6. وأظهرت هذه الدراسة أن تحليل المستوى المعرفي ودرجة الصعوبة يمكن أن يكونا جزءًا من تقييم الأداة الاختبارية. ويختلف البحث الحالي عنها في أنه لا يهدف إلى إجراء تقييم سيكومتري شامل، وإنما يركز على المستوى المعرفي ودرجة الصعوبة بوصفهما خاصيتين أساسيتين للتحليل^{١٨}.

وقام قدرى وعسري وبحر الدين بتحليل أسئلة امتحان UAMBN للغة العربية باستخدام منظور تصنيف بلوم المنقّح. وقد استخدمت هذه الدراسة الوثائق والمقابلات، وتوصلت إلى أن الأسئلة ترتبط بالكفاءات التي تمثل مرجعًا لها، إلا أن توزيع عناصر اللغة، ومهارات اللغة، والمستوى المعرفي لم يُظهر توازنًا كافيًا. وترتبط هذه الدراسة بالبحث الحالي

تحليل بنود أسئلة اختبارات اللغة العربية لغير الناطقين بها على ضوء نظرية تصنيف بلوم: دراسة وصفية تحليلية “ Utami, 17
”في الجامعة الإسلامية الحكومية فونوروكو

18 Syihabuddin Et Al., “Wikalah Ibt.”

من جانب تحليل المستوى المعرفي، بينما لا يقوم هذا البحث بتحليل مدى توافق الأسئلة مع وثائق الكفاءة، بل يضيف تحليل درجة الصعوبة بناءً على استجابات طُلابِ المَعَهْدِ^{١٩} وقامت محيسوه وسورية بتحليل أسئلة الاختبار المدرسي للغة العربية وفقاً لتصنيف بلوم، بالإضافة إلى دراسة مدى توافقها مع مخطط الاختبار، وصدقها، وثباتها. وأظهرت نتائج الدراسة أن نسبة التوافق مع مخطط الاختبار بلغت ٩٢٪، كما ظهر وجود ميل نحو فئات معرفية معينة في الأداة الاختبارية التي تم تحليلها. وقد استُخدمت بيانات إجابات الطلاب في تلك الدراسة لدعم تحليل خصائص الأداة الاختبارية. أما البحث الحالي فيختلف عنها من حيث التركيز، إذ تُستخدم بيانات إجابات طلاب المعهد بشكل خاص للحصول على درجة الصعوبة، ثم تتم مقارنتها مع التصنيف المعرفي لكل فقرة^{٢٠}.

وقامت سلسبيلا من خلال رسالتها بعنوان دراسة التقويم النهائي لتعلم اللغة العربية القائم على مهارات التفكير العليا (HOTS) في المدرسة الثانوية الحكومية ١١ باندونغ بدراسة التقويم النهائي للغة العربية مع التركيز على مهارات التفكير العليا. وأظهرت نتائج الدراسة أنه من بين ٢٨ فقرة تم تحليلها، كانت هناك فقرة واحدة فقط في مستوى C4، بينما لم توجد فقرات في مستويي C5 و C6 وترتبط هذه الدراسة بالبحث الحالي من حيث تحليل عمليات التفكير، إلا أن البحث الحالي لا يقتصر على فئات التفكير العليا

¹⁹ Qodri Et Al., "The Suitability Of Arabic Language Question Item In The Perspective Of Bloom Taxonomy Revision L Tanasub Bunudi As'ilah Al-Lughah Al-'Arabiyyah Fi Dhau Tashnif Bloom Al-Mu'addal."

²⁰ Makhisoh And Suroyya, *Analisa Butir Soal Ujian Madrasah Bahasa Arab Menurut Taksonomi Bloom*.

فقط، بل يحلل جميع المستويات من C1 إلى C6 ، ثم يقارن نتائج التصنيف مع درجة صعوبة كل فقرة^{٢١}.

وقامت زلفى وزنيورة وبايتي من خلال بحثها بعنوان تحليل فقرات أسئلة اللغة العربية من منظور أندرسون وكراثوول باستخدام تحليل المحتوى لأسئلة الاختبار النهائي للغة العربية. وأظهرت نتائج البحث أن ٢٥ سؤالاً كانت ضمن فئة LOTS ، و ٢٣ سؤالاً ضمن فئة MOTS ، وسؤالين ضمن فئة HOTS وفقاً للتصنيف المستخدم في الدراسة. وترتبط هذه الدراسة بالبحث الحالي من الناحية المنهجية؛ لأنها تستخدم تحليل متطلبات التفكير في الأسئلة، إلا أن البحث الحالي يحافظ على التصنيف المباشر من C1 إلى C6 ، ويضيف تحليل درجة الصعوبة بناءً على استجابات طلاب المعهد^{٢٢}.

وقامت ألفيانا بدراسة أسئلة الاختبار النهائي للغة العربية وفقاً لتصنيف بلوم، بالإضافة إلى إجراء اختبار الصدق والثبات. وأظهرت الدراسة أن من بين ٤٥ سؤالاً تم تحليلها، كان ٤٤ سؤالاً مستوفياً لمعايير الصدق المستخدمة، كما بلغت قيمة ثبات الأداة ٠,٥٥٦. وتشترك هذه الدراسة مع البحث الحالي من حيث موضوع البحث الذي يتمثل في الاختبار النهائي للغة العربية، بينما يختلف عنها في أن البحث الحالي لا يجري اختبار الصدق والثبات، وإنما يركز على توزيع المستوى المعرفي ودرجة صعوبة الفقرات^{٢٣}.

وقام مغفرة وأمر الله وواحد مورني وأنور بتحليل تدريبات الأسئلة في كتاب *Mahir fi Lughoh Arabiyah* وفقاً للمستويات الستة لتصنيف بلوم. وأظهرت

في المدرسة الثانوية الحكومية (Hots) دراسة التقويم النهائي لتعلم اللغة العربية القائم على مهارات التفكير العليا“، Salsabila, ١١ باندونج

٢٢ Zulfa Et Al., “Analisis Butir Soal Bahasa Arab Perspektif Anderson Dan Krathwohl.”

٢٣ Alfiana, “تحليل بنود الأسئلة في الامتحان النهائي للغة العربية على ضوء تصنيف بلوم لدى طلاب الفصل السابع بالمدرسة “، المتوسطة الإسلامية الحكومية ١ موجوكرتو

نتائج الدراسة أن التدريبات تتضمن أنواعًا مختلفة من متطلبات التفكير، بدءًا من معرفة المفردات والقواعد اللغوية، وصولًا إلى أنشطة التطبيق وتحليل النصوص العربية. ويمكن الاختلاف بين هذه الدراسة والبحث الحالي في موضوع البحث؛ إذ استخدمت تلك الدراسة تدريبات الكتاب التعليمي، بينما يستخدم البحث الحالي أداة الاختبار النهائي التي تم تنفيذها فعليًا من قبل طُلابِ المَعْهَدِ، مما يتيح توفر بيانات الاستجابات لتحليل درجة الصعوبة^{٢٤}.

وقام أسرور ورحمة وشكر ومالياتوسيكاه بتحليل ١٠٠ سؤال من أسئلة الاختيار من متعدد في أولمبياد اللغة العربية الوطني باستخدام تصنيف بلوم المنقّح. وأظهرت نتائج الدراسة أن ٥٣٪ من الأسئلة كانت ضمن مجموعة C1-C3، بينما كانت ٤٧٪ ضمن مجموعة C4-C6. وترتبط هذه الدراسة بالبحث الحالي من الناحية المنهجية؛ لأنها تستخدم تحليل المحتوى وتصنيف عمليات التفكير الست، إلا أن موضوعها يختلف، حيث يتمثل في أدوات اختبار المسابقة، ولا تستخدم بيانات استجابات المشاركين كما هو الحال في هذا البحث^{٢٥}.

وبناءً على الدراسات التسع السابقة، يمكن معرفة أن تحليل المستوى المعرفي قد استخدم في دراسة مختلف أدوات اختبار اللغة العربية. كما أن تحليل درجة الصعوبة ليس موضوعًا جديدًا، فقد جمعت بعض الدراسات السابقة بين درجة الصعوبة وتحليل خصائص أخرى للأداة الاختبارية. ولذلك، لا يقوم تميز هذا البحث على مجرد استخدام تصنيف بلوم أو تحليل درجة الصعوبة فقط.

²⁴ Maghfiroh Et Al., "Analisis Latihan Soal Bahasa Arab Pada Buku Ajar Mahir Fi Lughoh Arabiyah Perspektif Taksonomi Bloom."

²⁵ Asror Et Al., "Cognitive Level Mapping Of Indonesian Arabic Olympiad Questions."

ويتمثل موقع هذا البحث في تقييم عدة أدوات مختلفة للاختبار النهائي للغة العربية أعدّها معلّمون مختلفون في سياق مؤسسة تعليمية واحدة، مع تصنيف كل فقرة بصورة مباشرة ضمن مستويات C1 إلى C6 بناءً على متطلبات عمليات التفكير، وتحليل درجة الصعوبة بناءً على استجابات طلاب المعهد في كل أداة اختبارية، ثم مقارنة نتائج التحليلين بوصفهما أساساً لصياغة الدلالات المتعلقة بتطوير الأداة الاختبارية.

ومن خلال هذا الموقع، يُتوقع أن يقدم البحث تصوّراً أكثر شمولاً حول خصائص أدوات الاختبار النهائي للغة العربية المستخدمة في معهد الجامعة بجامعة سونان غونونغ جاتي الإسلامية الحكومية باندونغ.

