صلاحية مكونات اختبار تحديد المستوى في اللغة العربية: دراسة تحليلية للمفردات، والقراءة، والنحو

The Suitability and Validity of the Items Used for the Arabic Placement Test: An Analysis of Vocabulary, Reading Comprehension, and Grammar

Kesesuaian dan Kesahan Butiran yang digunakan untuk Ujian Penempatan Bahasa Arab: Satu Kajian Analisis Perbendaharaan Kata, Kefahaman Membaca dan Tatabahasa

حسين محمد جميل علي* وان أحمد رحيمان وان حسين** أزلان سيف البهروم***

مستخلص البحث

يحاول المقال تقديم رؤية نقدية تقييمية للمشروعية والصلاحية العلمية لبنود المفردات، والقراءة، والنحو في اختبار تحديد المستوى في اللغة العربية المطبق في مركز اللغات والتنمية العلمية بالجامعة الإسلامية العالمية في ماليزيا (CELPAD). ويسعى البحث للإجابة عن الأسئلة الآتية: هل يعاني الخطاب القياسي لهذه البنود من أي ضعف أو قصور في جوانب السهولة والصعوبة والتمييز؟ وهل تستوفي تلك البنود الشروط المنهجية المطلوبة لاعتمادها أدوات لقياس ما يناط بها؟ وما أكثر العيوب البنيوية التي تعاني منها تلك البنود؟ وبناءً على هذه الأسئلة تمت دراسة تلك البنود وتحليلها وذلك اعتماداً على عينة

* دكتوراه في المناهج وطرق التدريس. محاضر بمركز اللغات والتنمية العلمية بالجامعة الإسلامية العالمية في ماليزيا، البريد الإلكتروني: hgameili@iium.edu.my

** ماجستير في تعليم اللغة العربية والدراسات الإسلامية، محاضر بمركز اللغات والتنمية العلمية بالجامعة الإسلامية العالمية في ماليزيا، البريد الإلكتروني: wrahiman@iium.edu.my

*** دكتوراه في الدراسات العربية، محاضر بمركز اللغات والتنمية العلمية بالجامعة الإسلامية العالمية في ماليزيا.

مأخوذة من اختبار تحديد المستوى الذي أُجري على الطلاب الجدد في الفصل الدراسي الثاني من العام الدراسي 2014/2013م، وتمت الدراسة وفق المعايير والصيغ الإحصائية التي رصدت تربويًا لمثل هذا النحو من الدراسات. ومن أهم النتائج التي تم التوصل إليها: 1. فاعلية معظم تلك البنود؛ 2. اعتدال معامل الصعوبة في حيث أجاب نصف الطلاب على أكثر الأسئلة؛ 3. تعاني البنود محل النظر من بعض القصور في جانب قوة بعض المشتتات في أسئلة الاختيار من متعدد للمفردات؛ 4. بنود النحو امتازت بقدر جيد من الصدق، ولكنها تحتاج إلى زيادة في عددها وفقًا للأسس المنهجية المعروفة.

الكلمات الأساسية: اللغة العربية، اختبار تحديد المستوى، بنود اختبار تحديد المستوى في اللغة العربية، مفردات، قراءة، نحو، قصور، معايير.

Abstract

This article attempts to offer a critical assesment view of the academic validity and suitability of the items of vocabulary, reading comprehension and grammar used in the Arabic Placement Test at of the Centre for Languages and Pre-Academic Development (CELPAD). The study attempts to answer the following questions: Does the testing power of these items suffer from any weakness or inadequacy in terms of easiness, difficulty and specification? Do they satisfy the scientific and methodological requirements that make them proper testing tools fulfilling the purpose for which they have been designed? What are the most structural defects of those items? In light of these questions, an examination and analysis of the abovementioned items based on a sample of an exam of Arabic Placement Test carried out with new students during Semester 2, 2013/2014. The study was conducted according to statistical criteria and schemes pedagogically designed for such kind of study. Among the important findings of the study are the following: 1. Efficacy of most of those items; 2. Moderation of the difficulty variable as half of the students answered most of the questions; 3. The items under consideration suffer from inadequacy in terms of power distribution in the multiple choice exam questions; 4. The grammar items have a good degree of validity, but needs to be increased in number in accordance with established methodological criteria in this field.

Key words: Arabic language, placement test, items of Arabic placement test, vocabulary, reading comprehension, grammar, inadequacy, criteria.

Abstrak

Artikel ini bertujuan untuk memberikan pandangan nilai yang kritikal tentang kesahihan akademik dan kesesuaian butiran perbendaharaan kata, kefahaman membaca dan tatabahasa yang digunakan dalam Ujian Penempatan Bahasa Arab di Pusat Bahasa dan Pra-Akademik Pembangunan (CELPAD). Kajian ini berusaha untuk menjawab soalan-soalan utama seperti

berikut: Adakah kekuatan ujian tentang butiran ini mengalami sebarang kelemahan atau kekurangan dari segi mudahnya, kesukaran dan spesifikasi? Adakah butiran tersebut memenuhi keperluan ilmiah dan metodologi yang menjadikannya alat ujian yang betul dalam memenuhi maksud yang telah dirancang? Apakah kelemahan yang paling ketara butiran tersebut? Berdasarkan soalan-soalan ini, pemeriksaan dan analisis butiran yang dinyatakan di atas dibuat berdasarkan sampel peperiksaan bahasa Arab untuk Penempatan Ujian dijalankan dengan pelajar-pelajar baru pada Semester 2, 2013/2014. Kajian ini dijalankan mengikut kriteria statistik dan skim pedagogi yang direka untuk jenis kajian itu. Antara dapatan penting dalam kajian ini ialah: 1. Keberkesanan kebanyakan butiran; 2. Kesederhanaan dalam pembolehubah kesukaran berdasarkan separuh daripada pelajar menjawab kebanyakan soalan; 3. Butiran yang di bawah pemerhatian mengalami kekurangan dari segi kekuatan pengagihan soalan berbentuk objektif; 4. Butiran tatabahasa mempunyai skor kesahan yang baik, tetapi perlu ditingkatkan bilangan mengikut kriteria metodologi yang diterima-pakai dalam bidang ini.

Kata kunci: Bahasa Arab, ujian penempatan, butiran ujian Bahasa Arab penempatan, perbendaharaan kata, kefahaman membaca, tatabahasa, kekurangan, kriteria.

تمهيد

يؤدي اختبار الكفاية اللغوية وتحديد المستوى وظيفة مهمّة مهمّا في العملية التعليمية التعلمية، من حيث تجويدها وتحسين مخرجاتها من خلال العديد من القرارات التي يمكن اتخاذها في ضوء نتائج هذا الاختبار؛ لذا فقد نال اهتمامًا بالغاً لدى مركز اللغات والتنمية العلمية بالجامعة الإسلامية العالمية إذ خصصت للاختبارات وحدة كاملة من أجل بناء اختبارات جيدة تقدم نتائج ومعلومات موضوعية وواقعية يمكن اعتمادها في تصنيف الطلاب، وصنع القرارات. أما بالنسبة للطالب، فهي تعد وسيلة مهمّة للتقييم الذاتي، فبناء على نتائجه في اختبار تحديد المستوى يتبين موقعه،فيتم تحديد مستواه اللغوي والدراسي، أو حتى إعفائه من دراسة اللغة العربية.

مشكلة البحث

حرص المشرفون التربويون في مركز اللغات والتنمية العلمية على أن يكون اختبار تحديد المستوى في اللغة العربية ذا كفاءة عالية، وأن يتصف بالموضوعية، والصدق، والثبات، وحسن

الإعداد والإخراج، على النحو الذي يكون فيه عونًا للمعلم والطالب على حدّ سواء؛ لأنه إذا لم يعد وفق معايير الاختبار الجيد فإنه سيلحق الأذى والضرر بالأفراد ذوي العلاقة نفسيا، وتربويًا، واجتماعيًا.[1] ويعد اختبار تحديد المستوى في اللغة العربية مصيريًا بالنسبة للمتعلمين وغيرهم من أوليائهم ومن المسؤولين التربويين، كما أنه من الاختبارات المهمة في العملية التعليمية، حيث يتوقف على نتائجه انتقال المتعلم إلى الدراسة الجامعية. وعلى الرغم من أهمية هذا الاختبار إلا أن أسئلته لم تحظَ بالاهتمام الكافي.

ويعتقد الباحثون أن اختبار تحديد المستوى في اللغة العربية لم تتوفر لبنوده حتى الآن دراسات تحليلية تقيّم مدى جودة عناصره، من حيث الصعوبة والسهولة والتمييز؛ ولهذا السبب سعى هذا البحث لفحص عينة من أسئلة القراءة، والمفردات، والنحو للاختبار المعني.

أهمية البحث

يمكن إيجاز أهمية هذا البحث فيما يأتي:

1. يساعد هذا البحث في التعرف على فعالية الأنواع المختلفة من الأسئلة في اختبارات تحديد المستوى في اللغة العربية.

2. يحاول الباحثون إلقاء الضوء على مدى توافق بنود اختبارات القراءة التي تعدها لجان وحدة الاختبار والقياس بمركز اللغات والتنمية العلمية في الجامعة الإسلامية العالمية بماليزيا مع المعايير التربوية لبناء احتبارات الكفاية الجيدة.

3. يساعد هذا البحث في الكشف عن مدى كفاءة لجان إعداد الاختبار في بناء اختبار تحديد المستوى في اللغة العربية، والوقوف على نقاط القوة لديهم والتركيز عليها، ونقاط الضعف ومعالجتها من خلال الدورات التدريبية، واللقاءات العلمية المتخصصة.

4. إعداد نواة لبنك أسئلة اختبار تحديد المستوى يسهل الانتقاء منه في إعداد الورقة

---

[1] مراد، صلاح أحمد وسليمان، أمين علي، **الاختبارات والمقاييس في العلوم النفسية والتربوية: خطوات إعدادها وخصائصها** (القاهرة: دار الكتاب الحديث، 2002).

الامتحانية.

5- توضيح أثر عدد المشتتات وتغير موقع المشتت القوي في اختبار الاختيار من متعدد على التقديرات المختلفة لنظرية استجابة الفقرة.

6. إطلاع أصحاب القرار على واقع الاختبارات الحالية، ومدى حاجتها إلى إعادة النظر فيها، بغرض تطويرها وتحسينها.

7. كما يستمدّ هذا البحث أهميته من كونه – على حد علم الباحثين – الدراسة الأولى من نوعها التي تهدف إلى تقييم بنود اختبار القراءة في اختبار تحديد المستوى في اللغة العربية بالجامعة الإسلامية العالمية بماليزيا.

8. وتتمثل أهمية البحث كذلك في السعي للإسهام في التراكم المعرفي في هذا المجال، حيث يمكن أن تحث نتائج الدراسة الحالية همم الباحثين للقيام ببحوث مماثلة.

أسئلة البحث

هناك مجموعة من الأسئلة الأساسية قادت البحث، ويمن إجمالها فيما يأتي:

1. هل تعاني بنود اختبار تحديد المستوى في اللغة العربية في الجامعة الإسلامية من ضعف في معاملات الصعوبة، والسهولة، والتمييز؟

2. ما فاعلية البدائل في أسئلة الاختيار من متعدد من حيث التمييز؟

3. هل صيغت تلك العناصر وفقًا لمواصفات لغوية قياسية محددة؟

4. هل صيغت تلك العناصر وفق التصنيفات التربوية لمستويات المعرفة؟

تلخص هذه الأسئلة المنهجية صلب المشكلة التي يحاول هذا البحث أن يضعها أمام معايير القياس؛ للتحليل،والتقييم،والتفسير.

فرضيات البحث

للإجابة عن أسئلة البحث فقد وضع الباحثون الفرضيات الآتية:

1. تتوافر في بنود اختبار تحديد المستوى في اللغة العربية معايير الاختبار الجيد.

2. تتميز بنود اختبار تحديد المستوى في اللغة العربية بمعامل معتدل الصعوبة.

3. تتميز قدرة كلّ سؤال من أسئلة اختبار تحديد المستوى في اللغة العربية بالتمييز بين الطالب القوي، والطالب الضعيف في اللغة العربية.

حدود البحث

يقتصر هذا البحث على عناصر المفردات والقراءة والنحو في اختبار تحديد المستوى في اللغة العربية الذي طبق بمركز اللغات والتنمية العلمية بالجامعة الإسلامية العالمية في ماليزيا في الفصل الدراسي الثاني من العام الدراسي 2014/2013م.

منهج البحث

اعتمد البحث الحالي على المنهج الوصفي التحليلي الذي يعتمد على الأسلوبين الكمي والكيفي في تحليله لأسئلة اختبار تحديد المستوى في اللغة العربية، إذ إنه يمكن على هذا النحو تحليل وتقويم هذا الاختبار، وبالتالي التوصل إلى بعض الآليات والمقترحات التي يمكن أن تؤدي إلى تحسينه وتطويره.

مصطلحات البحث

المصطلحات الواردة تمثل التعريف الإجرائي لها في هذا البحث:

– **اختبار تحديد المستوى:** وهو اختبار يركز على قياس مدى استعداد الطالب لتعلم لغة أجنبية ما، ما رصيده من مهارات تلك اللغة؟ وذلك بهدف تحديد المستوى الذي يناسبه[1].

– **معامل الصعوبة** Item Difficulty index: يفيد معامل الصعوبة في إيضاح

1- انظر: Hughes, Arthur, *Testing for Language Teachers*, London Cambridge, p. 10-11. والتنقاري، صالح، "الاختبارات اللغوية"، **المجلة العربية للدراسات اللغوية**، معهد الخرطوم الدولي للغة العربية، العدد 25، ديسمبر 2006، ص2.

مدى سهولة أو صعوبة سؤال ما في الاختبار، وهو عبارة عن النسبة المئوية من الطلاب الذين أجابوا عن السؤال إجابة صحيحة ويحسب بتطبيق المعادلة الآتية:

$$\text{معامل الصعوبة} = \frac{\text{عدد الإجابة الصحيحة على الفقرة}}{\text{عدد الطلاب}} \times 100\ \%$$

ووفقاً لويلسون فإن صعوبة البند هي العنصر الأكثر أهمية في تحليل البند ومع ذلك، ليس هذا هو السبيل الوحيد لتقييم عناصر الاختبار.[1]

– **معامل التمييز** Item discrimination index: يقصد بمعامل التمييز قدرة فقرة الاختبار، أو السؤال على التمييز بين الطلبة الممتازين والضعاف. فالامتحان الجيد هو الذي يميز بين الطلاب. ويحسب بعد تقسيم مجموعة الطلاب إلى مجموعتين عليا ودنيا. وطريقة حسابه كما يلي:

معامل التمييز = (عدد الاجابات الصحيحة في المجموعة العليا – عدد الاجابات في المجموعة العليا) ÷ عدد طلاب إحدى المجموعتين.

– **المشتّت** Distractor: هو الخيار المموه في بنود اختبار الاختيار من متعدد.

– **المشتت القوي:** يقصد به البديل غير الصحيح الذي اختاره العدد الأكبر من المفحوصين بعد اختيارهم للبديل الصحيح.

– **موقع المشتت القوي:** يقصد به موقع المشتت القوي سواء أكان قريبًا بجانب البديل الصحيح، أم بعيدًا ببديل أو بديلين من موقع البند الصحيح.

العينة الطبقية Striated Sample: نستخدم هذا النوع من العينات عندما يكون هناك تباين (عدم تجانس) واضح في مجتمع الدراسة، بحيث يمكن تقسيم مجتمع الدراسة إلى مجموعات أو طبقات بناءً على هذا التباين.

2 Wilson, M., *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum, 2005).

الدراسات السابقة

لم تتوفر دراسة تحليلية ـ على حدّ علم الباحثين ـ لبنود (فقرات) اختبار تحديد المستوى في اللغة العربية بالجامعة الإسلامية العالمية في ماليزيا. وأكثر الدراسات ذات الصلة التي أجريت في مجال تحليل الاختبارات كانت في أغلبها موجهة للاختبارات التحصيلية في معظم المؤسسات التعليمية التي اطلع عليها الباحثون.

وفيما يلي عرض لبعض الدراسات السابقة التي أجريت في مجال هذه الدراسة :

1. هدفت دراسة زايد صالح بني عطا وإبراهيم محمد الرباعي[1] إلى التحقق من أثر عدد بدائل الفقرة، وموقع المموه القوي في فقرات اختبار الاختيار من متعدد على خصائص الاختبار، ومعالم الفقرات، وقدرة الأفراد. ولتحقيق الغرض من الدراسة تمّ بناء اختبار تحصيلي من نوع الاختيار من متعدد. وقد كشفت نتائج التحليل أنه لا يوجد فروق ذات دلالة إحصائية بين متوسطات معالم الصعوبة للفقرات تعزى لعدد بدائل الفقرة وموقع المموه القوي، وكذلك أظهرت النتائج أنه لا توجد فروق ذات دلالة إحصائية بين متوسطات معالم التخمين للفقرات تعزى لمتغير الدراسة، في حين أظهرت النتائج فروقًا ذات دلالة إحصائية بين متوسطات معالم التمييز ذات صلة لموقع المموه القوي، وقد جاءت تلك الفروق لصالح النموذج الثاني (5- بدائل موقع المموه القوي). وبينت نتائج التحليل كذلك أنه ليست هنالك فروق دالة إحصائيةً بين متوسطات معلمة القدرة للأفراد تعزى لعدد بدائل الفقرة وموقع المموه القوي. ولقد أوصى الباحثان بإجراءات تتناول أثر عدد البدائل وموقع المموه القوي على دقة التقديرات لمعالم الفقرات وقدرة الأفراد من خلال الخطأ المعياري للتقدير.

---

[1] بني عطا، والرباعي، "أثر عدد البدائل وتغير موقع المموه القوي في فقرات اختبار الاختيار من متعد على معالم الفقرات وقدرة الفرد ودالة المعلومات"، **المجلة الأردنية في العلوم التربوية**، 2012م.

2. أجرى شيزيوكا وتاكيوشي وياشيما[1] بحثاً حول بدائل الاختبار وعددها، وبينت دراستهم أن الاختبار ذا المشتتات الثلاثة يؤدي الغرض الذي يحقق الاختبار ذو الأربعة مشتتات، وأشار هؤلاء الباحثون إلى أن زيادة عدد مشتتات البنود تقلل من وقت الاختبار، وتزيد من فرصة عامل التخمين. ويرى الباحثون غير ذلك ؛ إذ إن زيادة المشتات القوية تضعف التخمين خاصة إذا كان عدد المشتتات أربعة إذ تتساوى نسب التخمين بنسبة خمس وعشرين بالمائة.

3. هدفت دراسة السليطي وتايه[2] فسعت إلى تحليل وتقويم أسئلة اختبارات الشهادة الثانوية العامة في مادة اللغة العربية بدولة قطر خلال الأعوام الدراسية من 1997 حتى 2001م، ذلك للتعرف على مدى تمشي أنماط الأسئلة المتضمنة في تلك الاختبارات مع تصنيف بلوم للمستويات المعرفية، وتحديد المستويات التي تم التركيز عليها، والمستويات التي أغفلتها تلك الاختبارات، وكذلك تحديد مستويات الفهم التي تقيسها تلك الاختبارات، كما تهدف إلى تحديد مدى تضمنها للمستويات الوجدانية، ودرجة تنوع الأسئلة في تلك الاختبارات ما بين المقالية والموضوعية. ومن أهم النتائج التي توصلت إليها: أن هناك تشابهاً بين القسمين العلمي والأدبي من حيث التركيز على الأسئلة المقالية أكثر من الأسئلة الموضوعية، مما يحرم الطالب مما تحققه الأسئلة الموضوعية من مزايا وأهداف. وكذلك عدم التوازن في توزيع الأسئلة على المستويات المعرفية المختلفة، حيث تركز الأسئلة بنسبة كبيرة على مستوى التذكر أكثر من باقي المستويات؛ الأمر الذي ينعكس سلباً على الاهتمام بالمستويات المعرفية العليا التي يؤكد عليها التربويون لما تحققه من أثر في قدرات ومهارات المتعلم، إضافة إلى عدم الدقة في صياغة

---

[1] Shizuka,Takeuchima & Yoshizawa, *A comparison of Three-and Four option English Tests for University Entrance functioning distracters in multiple-choice questions*, 9, 2-11.

[2] السليطي، حمدة حسن، وتايه، خضر عبد الله، **دراسة تحليلية تقويمية لأسئلة اختبارات الشهادة الثانوية العامة لمقررات اللغة العربية** (القسمان: العلمي والأدبي) من عام 1997-2001م بدولة قطر، وزارة التربية والتعليم – دولة قطر، 2001م.

بعض الأسئلة في الاختبارات كما أن بعضها مركب مما يؤدي إلى تشتت ذهن الطالب، وإرباك فهمه، مما يتسبب في عدم القدرة على الإجابة بصورة صحيحة وكاملة عن جميع فروع السؤال، وقد يتسبب كذلك في نسيان الإجابة عن فرع أو أكثر في السؤال، وخاصة وأنه في موقف امتحان مصيري، وقد يرجع ذلك إلى عدم تدريب القائمين على وضع الأسئلة الامتحانية، على كيفية صياغة الأسئلة، فيتم وضع الأسئلة بصورة عشوائية وغير مخطط لها.

4. كما أجرت فاطمة المطاوعة[1] دراسة هدفت إلى تحليل وتقويم أسئلة القراءة بالمرحلة الابتدائية في ضوء تصنيف بلوم للأهداف المعرفية، وفي إطار مهارات الفهم القرائي بمستوياتها المختلفة، وتوصلت إلى أن أسئلة اختبارات القراءة انحصرت في مستويات التذكر، والفهم، والتطبيق، أما بقية المستويات فكانت بنسب متدنية.

5. أما فقوسة[2] فقد أشارت نتائج دراسته إلى وجدود فوارق ذات دلالة إحصائية بين متوسطات معاملات الصعوبة للبنود الاختبارية ذات مشتتات الثلاثة. أما بالنسبة لمعاملات التمييز فقد أظهرت نتائج دراسته وجود فروق ذات دلالة إحصائية بين متوسطات معاملات تمييز البنود ذوات الخمسة مشتتات والبنود ذوات الأربعة مشتتات، وكانت لصالح البنود ذوات الخمسة مشتتات.

6. كما أجرى مولتون (Moulton)[3] دراسة من بين أهدافها التعرف على مدى توزيع أسئلة عينة من الاختبارات على الأهداف المعرفية، بحيث تم تحليل ثمانية وسبعين اختباراً في مقررات مختلفة. وتوصلت الدراسة إلى أن الأسئلة تركزت على الأهداف

---

[1] المطاوعة، فاطمة عبدالرحمن، "أسئلة القراءة وامتحاناتها في المرحلة الابتدائية بدولة قطر"، **مجلة مركز البحوث التربوية**، العدد الثامن عشر، السنة التاسعة، مركز البحوث التربوية بجامعة قطر، الدوحة، يوليو 2000م.

[2] فقوسة، محمد، "أثر عدد البدائل في اختبار الاختيار من متعدد على تقديرات القدرة"، **المجلة الأردنية في العلوم التربوية**، 1994، ص 309-335.

[3] Moulton, M., "Effects of level review and sequence of inserted questions of test processing". *Journal of Educational Psychology*, 2003, 73, pp. 427- 436

المعرفية الدنيا، كما وضحت عدم وجود فروق أساسية في توزيع هذه الأسئلة تعزى إلى التخصص، أو الجنس، أو الخبرة.

يتضح من العرض السابق للدراسات العربية والأجنبية أن هناك اهتماماً كبيراً بتحليل الأسئلة الاختبارية لما تحتله من مكانة، وما تمثله من أهمية، وقد تشابهت جميعها في اتخاذها من تصنيف بلوم المعرفي أساساً للدراسة.

7. ولصالح التنقاري دراستان متعلقتان بالموضوع، ولذلك فهما جديرتان بالوقوف عندهما، الأولى بعنوان "أساليب إعداد أسئلة فهم المقروء"[1] هدفت إلى فحص أسئلة فهم المقروء لمعرفة العناصر التي يتوجب تضمينها في فهم المقروء، وتوصل فيها إلى أن تلك العناصر تتمثل في الفهم الحرفي، وإعادة التنظيم، والاستناج، والتقويم. كما بينت هذه الدراسة أن أفضل بنود الاختيار لقياس فهم المقروء هي: الاختيار من متعدد، والتتمة، والأسئلة القصيرة. وقد أخذت عينة بحثه من مركز اللغات في الجامعة الإسلامية العالمية في ماليزيا. ومع ذلك يظلّ الاختلاف قائماً بين بحث التنقاري، والبحث الحالي؛ لأن الأول ركز على الاختبارات التحصيلية،بينما البحث الحالي هدفه التركيز على اختبار تحديد المستوى. أمَّا الدراسة الثانية فهي بعنوان "قضايا في اختبار تحديد المستوى"،[2] وهي دراسة نظرية عن مواصفات الاختبار الجيّد مع التركيز على اختبار تحديد المستوى، ولم يكن من أهدافها فحص عناصر المفردات، والقراءة، والنحو في اختبار تحديد المستوى، وهذا ما سيقوم به البحث الحالي.

ويتفق البحث الحالي مع الدراسات السابقة من حيث الهدف العام وهو تحليل أسئلة الامتحانات، إلا أنه يختلف عنها من حيث نوع الاختبار والمستهدفين به. كما يتفق معها في بعض الإجراءات العملية ومنها إعداد أدوات الدراسة (استمارة

---

[1] التنقاري، "أساليب إعداد أسئلة فهم المقروء"، **المجلة العربية للدراسات اللغوية**، معهد الخرطوم الدولي، العددان 29-30، ديسمبر 2011م. التنقاري، صالح، "قضايا في اختبار تحديد المستوى"، ورقة بحثية غير منشورة، قدمت في **مؤتمر اللغة العربية وهوية الأمة**، الجامعة الأردنية .15-17 أكتوبر، 2012م.

التقويم/التحليل) والتأكد من صدقها وثباتها.

وتُعَدّ هذه الدراسة استكمالاً للجهود والمبادرات السابقة إن وجدت، من خلال تحليل وتقويم بنود اختبار تحديد المستوى في اللغة العربية للتعرف على جوانب القصور والقوة فيها، ومحاولة تقديم بعض المقترحات التي يمكن أن تؤدي إلى تحسين تلك البنود، لتصبح أكثر فعالية في تحقيق الأهداف المرجوة منها.

عينة البحث

اشتمل مجتمع عينة البحث على ماتتين وسبعة وخمسين ممتحنًا (252) ذوو معرفة لغوية سابقة باللغة العربية. وبلغ عدد الطلاب الأدبيين في مجتمع العينة مائة وسبعة عشر ممتحنًا (112)، بينما بلغ عدد الطلاب العلميين مائة وأربعين ممتحنًا (140). والجدول الآتي يبين البيانات الأولية لمجتمع عينة البحث:

الجالسون من القسم الأدبي	الجالسون من القسم العلمي	جملة مجتمع العينة
112 طالباً وطالبة	140 طالبًا وطالبة	252
النسبة المئوية = 44.5%	النسبة المئوية = 55.5 %	%100

جدول (1): بيانات مجتمع البحث

رأى الباحثون اختيار واحدًا وأربعين (41) ممتحنًا علميا وفق نظام العينة الطبقية، وذلك ليمثلوا المجموعات العلمية في مجتمع الدراسة. إن هدف اختيار العينة طبقيا هو لضمان التمثيل المرغوب فيه للجماعات الفرعية. وبذلك حدد المجموع الكلي لعينة البحث بمائة وثلاثة وخمسين (153) ممتحنًا بنسبة تصل إلى 60.7%.

وتمثلت خطوات اختيار العينة الطبقية فيما يأتي:

أ. تحديد المجتمع الأصلي للعينة وتعريفه.

ب. تحديد حجم العينة المرغوب فيها.

جـ. تمّ تصنيف برامج الطلاب العلميين، وتحديد نسب الجالسين من كل برنامج مأخوذة من العدد الكلي لمجتمع عينة البحث.

د. تم اختيار الطلاب العلميين عشوائيا وفق النسبة الطبقية التي حددت من قبل بحيث يتناسب عدد كل فئة في العينة مع عددها في المجتمع الأصلي للدراسة، كما استبعدت نسب البرامج التي يقل عدد الطلاب فيها عن عشرة طلاب، والجدول الآتي يوضح ذلك:

البرنامج	الأحياء	الهندسة الإعمارية	العلوم الصحية	الهندسة	الصيدلة	تكنولوجيا المعلومات	الفيزيا ء	الطب
العدد	33	29	26	24	17	6	4	1
الوزن	25.6	22.5	20.2	18.6	13.2	0	0	0
العدد المطلوب	11	9	8	8	5	0	0	0

جدول (2): اختيار الطلاب العلميين للعينة وفق نسبهم

أما بالنسبة للطلاب الأدبيين، فقد تمّ اعتمادهم جميعًا بلا استثناء ضمن عينة البحث، ويعود ذلك – في رأي الباحثين – إلى الآتي :

1. الطلاب الأدبيون حاولوا الإجابة عن كل بنود أسئلة الاختبار المقصود بالبحث، بينما اقتصر الطلاب العلميون على إجابة القسمين الأوليين في الغالب الأعم.

2. الطلاب الأدبيون يلزمهم دراسة اللغة العربية في الجامعة الإسلامية حتى المستويات العليا؛ ولذا لم يتركوا سؤالاً يقيس مقدرتهم وكفايتهم اللغوية إلا وحاولوا الإجابة عنه، ربما بغرض الإعفاء، أو بهدف الوقوف على تلك المقدرة لأول مرة. وتشكل عينة الدراسة ما نسبته 60.7% من مجتمع الدراسة.

المعالجات الإحصائية

للإجابة عن أسئلة البحث قام الباحثون بتحليل أسئلة اختبار تحديد المستوى وفقا للأداة المعدة لذلك الغرض، وتمّ تفريغ البيانات على استمارة التحليل، وبعدها تم إجراء

المعالجات الإحصائية الآتية:

أ. حساب معامل الصعوبة لبنود الاختبار:

الهدف من هذه الخطوة هو الإجابة عن السؤال الأول لأسئلة البحث وذلك بتحديد مستوى صعوبة الأسئلة بغرض استبعاد، أو تعديل، ما يظهر من أسئلة سهلة أو صعبة أو غامضة، وقد استخدمت المعادلة التالية لحساب معامل لصعوبة[1].

$$\text{معامل الصعوبة} = \frac{\text{عدد الإجابة الصحيحة على الفقرة}}{\text{عدد الطلاب}} \times 100\ \%$$

وبتطبيق المعادلة السابقة، حصل الباحثون على معاملات الصعوبة لكل بند من بنود الاختبار. ووجد الباحثون أن البند الذي يصل معامله إلى أكثر من 85% يعد بنداً شديد السهولة، وأن البند الذي يقل معامله عن 30% يكون شديد الصعوبة. وبشكل عام يعتمد معامل الصعوبة المطلوب على الغرض من الاختبار، وفي الاختبارات العادية فإن أفضل معامل صعوبة للسؤال هو 50% وما حولها.

ويوضح الجدول (3) معاملات سهولة وصعوبة وتمييز بنود اختبار تحديد المستوى في اللغة العربية:

مجال القياس	السؤال	مجموع الإجابات الصحيحة	مجموع الذين حاولوا الإجابة	معامل الصعوبة %	مستوى الصعوبة والسهولة
المفردات	1	116	153	75.8	معتدل السهولة
	2	138	153	90.1	سهل جدا
	3	122	153	79.7	معتدل السهولة

[1] عبد الرحمن، سعد، **القياس النفسي: النظرية والتطبيق** (القاهرة: دار الفكر العربي، ط3، 1998)، ص205.

4	138	153	90.1	سهل جدا	
5	51	153	33.3	معتدل الصعوبة	
6	50	153	32.6	معتدل الصعوبة	
7	123	153	80.3	معتدل السهولة	
8	102	153	66.6	معتدل الصعوبة	
9	77	153	50.3	معتدل الصعوبة	
10	93	153	60.7	معتدل السهولة	
11	50	153	32.6	معتدل الصعوبة	
12	36	153	23.5	صعب جدا	
13	6	153	3.9	صعب جدا	
14	48	153	31.3	معتدل الصعوبة	
15	21	153	13.7	صعب جدا	
16	51	153	33.3	معتدل الصعوبة	
17	95	153	62.0	معتدل السهولة	
18	91	153	59.4	معتدل الصعوبة	
19	41	153	26.7	صعب جدا	
20	9	153	5.8	صعب جدا	
21	8	153	5.2	صعب جدا	
22	11	153	7.1	صعب جدا	
23	10	153	6.5	صعب جدا	
24	52	153	33.9	معتدل الصعوبة	
25	5	153	3.2	صعب جدا	
26	32	153	20.9	صعب جدا	
27	97	153	63.3	معتدل السهولة	
28	99	153	64.7	معتدل السهولة	
29	109	153	71.2	معتدل السهولة	
30	61	153	39.8	معتدل الصعوبة	

القراءة	31	141	153	92.1	سهل جدا
	32	92	153	60.1	معتدل السهولة
	33	89	153	58.1	معتدل الصعوبة
	34	100	153	65.3	معتدل السهولة
	35	60	153	39.2	معتدل الصعوبة
	36	29	153	18.9	صعب جدا
	37	32	153	20.9	صعب جدا
	38	33	153	21.5	صعب جدا
	39	45	153	29.4	صعب جدا
	40	20	153	13.0	صعب جدا
	41	54	153	35.2	معتدل الصعوبة
	42	70	153	45.7	معتدل الصعوبة
	43	13	153	8.4	صعب جدا
	44	7	153	4.5	صعب جدا
	45	4	153	2.6	صعب جدا
	46	18	153	11.7	صعب جدا
	47	11	153	7.1	صعب جدا
	48	22	153	14.3	صعب جدا
	49	12	153	7.8	صعب جدا
	50	5	153	3.2	صعب جدا
	51	4	153	2.6	صعب جدا
	52	1	153	0.6	صعب جدا
	53	2	153	1.3	صعب جدا
	54	1	153	0.6	صعب جدا
	55	1	153	0.6	صعب جدا

النحو	56	2	153	1.3	صعب جدا
	57	6	153	3.9	صعب جدا
	58	14	153	9.1	صعب جدا
	59	12	153	7.8	صعب جدا
	60	10	153	6.5	صعب جدا
	61	22	153	14.3	صعب جدا
	62	0	153	0	صعب جدا
	63	0	153	0	صعب جدا
	64	1	153	0.6	صعب جدا
	65	0	153	0	صعب جدا
	66	1	153	0.6	صعب جدا
	67	0	153	0	صعب جدا
	68	2	153	1.3	صعب جدا
	69	1	153	0.6	صعب جدا

جدول (3): معامل الصعوبة والسهولة والتمييز ومستوى الأسئلة[1]

وقد استخدم الباحثون في تفسير تحليل البنود بالجدول الآتي:

معايير معامل الصعوبة (درجة الصعوبة)	
مدى القيم	مستوى الصعوبة
من صفر إلى 30	صعب جدا (ضعيف)
من 31 إلى 59	معتدل الصعوبة
من 60 إلى 84	معتدل السهولة
من 85 إلى 100	سهل جدا

جدول (4): معايير معامل الصعوبة

[1] جدول من تصميم الباحثين بني اعتمادا على كتب الإحصاء التربوي وخبرة الباحثين.

وبناء على ذلك يتضح من الجدول رقم (3) أن أربعة وعشرين (24) سؤالاً من تسعة وستين (69) تعد أسئلة مقبولة لوقوعها بين الاعتدال والسهولة المعتدلة. كما يتضح أن الأسئلة التي ظهرت بمظهر الصعوبة، يعود ذلك – في رأي الباحثين – إلى أن الامتحان يضع في اعتباره حدود النجاح للمستويات (Cut Points) والتي تراعي بدورها الكفاية اللغوية لكل مستوى. ويتضح من الجدول أيضاً أن الأسئلة مناسبة إلى حد كبير فكانت إما معتدلة السهولة، أو معتدلة الصعوبة حتى السؤال الحادي والثلاثين وهذا كافٍ لقياس المستوى الأول من مستويات الكفاية اللغوية.

وبالنسبة للسؤال الثاني من أسئلة البحث وهو الذي يخص فاعلية بدائل الاختيار من متعدد، فقد وجد الباحثون أن عناصر الاختيار لم تتجاوز خمسة في مجال المفردات النحوية. قام الباحثون بتحليل تلك البنود؛ لمعرفة معامل التمييز (discrimination index) وقوة البديل، أما معامل صعوبة هذه البنود فقد سبق ذكره في جدول رقم (3). واستفاد الباحثون من جدول المعدلات المثالية للتمييز الخاصة ببنود الاختيار من متعدد الذي طوره التربوي اللورد Lord F.M [1] على النحو الآتي:

شكل العناصر	معامل التمييز
العنصر الذي يتكون من خمسة خيارات	70
العنصر الذي يتكون من أربعة خيارات	74
العنصر الذي يتكون من ثلاثة خيارات	77
العنصر الذي يتكون من خيارين (صحيح – خطأ)	85

جدول (5): المعدلات المثالية للتمييز الخاصة

[1] F.M., Lord, "The Relationship of the Reliability of Multiple-Choice Test to the Distribution of Item Difficulties," *Psychometrika*, 1952, pp. 8, 181-194.

فاعلية البدائل: في الأسئلة ذات الاختيار من متعدد تكون الخيارات بدائل محتملة للإجابة عنها، ويكون هنالك بديل واحد هو الإجابة الصحيحة، ويفترض أن البدائل الأخرى تمثل إجابات محتملة، وأنها تجذب بعض الطلاب بمعنى أنه يتم اختيار أي بديل من البدائل الخاطئة من قبل طالب واحد على الأقل أو بنسبة لا تقلّ عن 5% من الطلاب على أن يكون غالبيتهم من الفئة الدنيا من الطلاب. أما البدائل التي لا تجذب أحداً منهم أو القليل جدا منهم، فهي بدائل غير فعّالة، ويفضل استبدال غيرها بها. كما تحذف البدائل التي تجذب من طلاب المجموعة العليا أكثر مما تجذبه من طلاب المجموعة الدنيا. وحسبت فعالية البدائل من العلاقة الآتية:

$$\text{معامل الصعوبة} = \frac{\text{عدد الذين اختاروا البديل من الفئة العليا} - \text{عدد الذيناختاروا البديل من الفئة الدنيا}}{\text{عدد طلاب إحدى المجموعتين}} \times 100\ \%$$

وعليه كانت البيانات على النحو الآتي:

السؤال	الخيار	عدد الذين اختاروا الخيار الصحيح من المجموعة العليا	عدد الذين اختاروا الخيار الصحيح من المجموعة الدنيا	النسبة المئوية للبديل %
26	أ	9	10	
	ب	12	8	12.9
	ج	10	9	
	د	0	1	
	صفر	0	3	
27	أ	1	5	
	ب	1	6	

	ج	22	11	35.4
	د	6	6	
	صفر	1	3	
28	أ	0	11	
	ب	31	5	83.8
	ج	0	5	
	د	0	7	
	صفر	0	3	
29	أ	29	17	38.7
	ب	0	2	
	ج	1	9	
	د	1	0	
	صفر	0	3	
30	أ	0	1	
	ب	0	16	
	ج	15	7	
	د	16	4	38.7
	صفر	0	3	
الجملة		31	31	

تمثل الأرقام التي حولها دائرة البديل الصحيح، كما تمثل النسب المئوية في العمود الأخير درجة فعالية ذلك البديل. وقد كانت أكبر قيمة للبديل القوي تساوي (83.8%) للعنصر (28)، وأقلّ قيمة للبديل القوي تساوي (12.9%) للبند (26). وبناء

على جدول المعدلات المثالية رقم (5) فإن البند (28) تجاوز الاعتدال، أما البنود الأخرى فكانت أقل معدلات وفق ذلك الجدول بل لم تقارب تلك المعدلات بنسب كبيرة.

وفيما يتعلق بسؤال البحث الثالث، الخاص بصياغة تلك البنود وفقًا لمواصفات لغوية قياسية محددة، فإن الباحثين يرون أن هناك مواصفات على درجة لغوية قياسية واضحة الهدف. وتتضح دقة تلك المواصفات فيما يأتي:

1- الأهمية النسبية للأنواع المختلفة من الأسئلة، وذلك على النحو الآتي:

الرقم	نوع الأسئلة	الدرجة المخصصة	النسبة المئوية لدرجات السؤال إلى الدرجة الكلية للاختبار
1	أسئلة الترتيب	2.5	% 2.5
2	أسئلة التكميل	9.5	%9.5
3	الاختيار من متعدد	5	%5
4	أسئلة الفهم والاستيعاب	25	%25
5	أسئلة تحليل الأخطاء	6	%6
6	أسئلة التتمة	2	%2
7	أسئلة المقال	40	%40
8	أسئلة الشفوي	10	%10
	مجموع درجات الاختبار	100	%100

جدول (6): الأهمية النسبية للأنواع المختلفة من الأسئلة

يتضح من الجدول (6) أن أعلى نسبة أهمية كانت للأسئلة المقالية 40% في حين كانت أدنى نسبة أهمية لأسئلة التتمة 2%، وأسئلة الاختيار من متعدد 5%، وتحليل الأخطاء بنسبة 6%، والشكل يوضح التفاوت في الأهمية النسبية للأنواع المختلفة من الأسئلة.

2- مراعاة الأسئلة لمستويات المعرفة:

يعد تصنيف بلوم (Bloom's Taxonomy) في المجال المعرفي بمستوياته المختلفة من أكثر أشكال التصنيف شيوعاً واستخداماً في مجال تحليل وتقويم أسئلة الاختبارات، ويتضمن هذا التصنيف ستة مستويات للمعرفة، مرتبة ترتيباً هرميا، يعتمد كل مستوى منها على مدى إتقان المستويات السابقة له، وهذه المستويات هي التذكر، الفهم، التطبيق، التحليل، التركيب، التقويم. وقد راعى اختبار تحديد المستوى في اللغة العربية في توزيع بنوده تلك على المستويات المعرفية. والجدول الآتي يوضح ذلك:

المستوى	توزيع العلامات حسب المستويات المعرفية					
	تذكر	فهم	تطبيق	تحليل	تركيب	تقييم
الأول	☑					
الثاني	☑					
الثالث		☑				
الرابع			☑			
الخامس				☑	☑	
السادس					☑	☑

جدول رقم (7): توزيع العلامات لكل مستوى في ضوء المستويات الدراسية

خاتمة: نتائج وتوصيات

توصل البحث بعدد من النتائج والتوصيات والمقترحات:

أهم النتائج:

1. فاعلية معظم تلك البنود واعتدال معامل الصعوبة حيث أجاب نصف الطلاب على أكثرها.

2. تعاني بعض أسئلة الاختيار من متعدد قصوراً في جانب قوة بعض المشتتات.

3. امتازت بنود النحو بقدر جيد من الصدق، ولكنها تحتاج إلى زيادة في عددها وفقًا للأسس المنهجية المعروفة.

التوصيات:

1. يوصي الباحثون بتجريب الفقرات الاختبارية التي تتميز بمعدل صعوبة معتدل مرة أخرى، فإن حافظت على ثباتها يمكن اعتمادها أداة قياس مجربة في الامتحانات المستقبلية.

2. يوصي الباحثون باستبعاد الفقرات شديدة الصعوبة واستبدالها بفقرات أخرى.

3. يوصي الباحثون باستخدام بنود الاختيار من متعدد في أسئلة فهم المقروء؛ بوصفها أكثر موضوعية من الأسئلة المباشرة على أن ينتدب لتنفيذها أساتذة من ذوي الخبرة لوضعها.

المقترحات:

1. يقترح الباحثون إنشاء نواة لبنك الأسئلة يحوي الفقرات التي تم فحصها والإبقاء عليها.

2. يقترح الباحثون إنشاء لجنة خاصة داخل وحدة الاختبار تعنى بتحليل الاختبارات في كلّ فصل دراسي.

3- تنظيم برامج تدريبية وورش عمل للمعلمين والمعلمات؛ لتنمية قدراتهم في كيفية إعداد الاختبارات، وصياغتها، وتحليلها.

المراجع:

References:

Abdul Rahman, Saôd, al-QiyÉs al-Nafsê al-NaÐariyyah wa al-TaÏbêq (Cairo: DÉr al-Fikr al-ÑArabi, 3rd edition, 1998).

Al-Mutawi'ah, Fatimah Abdul Rahman, "Asôilat al-QirÉôah wa ImtiÍÉnÉtihÉ fê al-MarÍalat al-IbtidÉôiyyah bi Dawlat Qatar", Majallat Markaz al-BuÍËth al-Tarbawiyyah, issue no. 8, year 9, Markaz al-BuÍËth al-Tarbawiyyah – Qatar University, Doha, July 2000

Al-Sulayti, Hamdah Hasan and Tayih, Khadir Abdullah, DirÉsah TaÍlÊliyyah Taqwêmiyyah li AsÒilat IkhtibÉrÉt al-ShahÉdah al-ThÉnawiyyah al-ÑÓmmah li MuqarrarÉt al-Lughah al-ÑArabiyyah (Doha-Qatar: WizÉrat al-Tarbiyah wa al-TaÑlÊm, 2001).

Bani ÑAta and ar-RibaÒi, "Athar ÑAdad al-BadÉÒil wa Taghayyur MawqiÑ al-Mumawwah al-Qawê fÊ FaqarÉt IkhtibÉr al-Ikhtiyér min MutaÑaddin ÑalÉ MaÑÉlim al-FaqarÉt wa Qudrat al-Fard wa Dallat al-MaÑlËmat", al-Majallah al-Urduniyyah fÊ al-ÑUlum al-Tarbawiyyah, 2012).

Eltingari, Salih, "al-IkhtibÉrÉt al-Lughawiyyah", al-Majallah al-ÑArabiyyah li al-DirÉsÉt al-Lughawiyyah, Ma'had al-KhurÏËm al-DawlÊ li al-Lughah al-ÑArabiyyah, issue no. 25, Desember 2006).

Eltingari, Salih, "AsÉlib IÑdÉd AsÒilat Fahm al-MaqrËÒ", al-Majallat al-ÑArabiyyah li al-DirÉsÉt al-Lughawiyyah, Ma'had al-KhurÏËm al-DuwalÊ, issue no. 29 & 30, Desember 2011.

Eltingari, Salih, "QaÌÉyÉ fÊ IkhtibÉr TaÍdÊd al-MustawÉ", unpublished paper presented in International Conference on al-Lughah al-ÑArabiyyah wa Huwiyyat al-Ummah, University of Jordan, 15-17 October 2012.

F.M., Lord, "The Relationship of the Reliability of Multiple-Choice Test to the Distribution of Item Difficulties," Psychometrik, 1952, 18.

Faqqusah, Muhammad, "Athar Adad al-badail fi Ikhtibar al-Ikhtiyar min Muta'addid ala taqdirat al-Qudrah", Majallat al-Urduniyyah fi al-ÑUlËm al-Tarbawiyyah, 1994.

Hughes, Arthur, Testing for Language Teachers, London Cambridge, no date).

Moulton, M., "Effects of level review and sequence of inserted questions of test processing", Journal of Educational Psychology, 2003, 73.

Murad, Salah Ahmad and Sulayman, Amin Ali, al-IkhtibÉrÉt wa al-MaqÉyÊs fÊ al-ÑUlËm al-Nafsiyyah wa al-Tarbawiyyah: KhuÏuwÉt IÑdÉdihÉ wa KhaÎÉÒiÎihÉ (Cairo Dar al-Kitab al-Hadith, 2002).

Shizuka, Takeuchima, and Yoshizawa, A comparison of Three-and Four option English Tests for University Entrance functioning distracters in multiple-choice questions, 9.

Wilson, M., Constructing measures: An item response modeling approach (Mahwah-NJ: Lawrence Erlbaum, 2005).